Towards a practice of reflexivity

Verwoerd, Lisa

2022

document version
Publisher's PDF, also known as Version of record

Towards a practice of reflexivity

On the normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency

Lisa Verwoerd

Towards a practice of reflexivity

On the normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency

Lisa Verwoerd

Cover design and illustrations by Lisa Verwoerd and Dirk Verwoerd

Printed by ProefschriftMaken | www.proefschriftmaken.nl

ISBN/EAN: 978-94-6423-712-2

Global Academic Press

VRIJE UNIVERSITEIT

Towards a practice of reflexivity

On the normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan

de Vrije Universiteit Amsterdam,

op gezag van de rector magnificus

prof.dr. J.J.G. Geurts,

in het openbaar te verdedigen

ten overstaan van de promotiecommissie

van de Faculteit der Bètawetenschappen

op vrijdag 18 maart 2022 om 13.45 uur

in een bijeenkomst van de universiteit,

De Boelelaan 1105

door

Lisa Verwoerd

geboren te Amersfoort

promotor:	dr. B.J. Regeer
copromotoren:	dr. P. Klaassen dr. A.M.C. Loeber dr. E.M. Kunseler

Members of the thesis committee:

- prof.dr. J.E.W. Broerse
- prof.dr.ir. J.T. Mommaas
- prof.dr. E.H.W.J. Cuppen
- dr. B.C. van Mierlo
- prof.dr. M.A. van der Steen

Content

Account

Chapters 4, 7 and 8 are based on co-authored articles that have been published in peer-reviewed journals. For Chapter 5, revisions are currently in progress. Chapter 6 has been resubmitted.

Chapter 4 – Verwoerd, L., Klaassen, P. & Regeer, B.J. (2021). *How to normalise reflexive evaluation? Navigating between legitimacy and integrity.* Evaluation. 27(2), 229-250.

Chapter 5 – Verwoerd, L., Brouwers, H.J.H., Kunseler, E.-M., Regeer, B.J., de Hoop, E. *Negotiating space for co-production.* Science & Public Policy. (revisions in progress).

Chapter 6 – Brouwers, H.J.H., Verwoerd, L.[1], Loeber, A.M.C., Regeer, B.J., Klaassen, P. *Impact assessment in context: exploring the interplay between political dimensions of knowledge co-production and actors' understandings of impact.* Environmental Science & Policy. (resubmitted)

Chapter 7 – Klaassen, P., Verwoerd, L., Kupper, F., & Regeer, B. J. (2021). *Reflexive Monitoring in Action as a methodology for learning and enacting Responsible Research and Innovation.* In E. Yaghmaei & I. Van de Poel (Eds.), Assessment of Responsible Innovation: methods and practices (pp. 222–243). Routledge.

Chapter 8 – Verwoerd, L., Klaassen, P., Van Veen, S. C., De Wildt-Liesveld, R., & Regeer, B. J. (2020). *Combining the roles of evaluator and facilitator: Assessing societal impacts of transdisciplinary research while building capacities to improve its quality*. Environmental Science & Policy, 103, 32-40.

[1] Shared first authorship

1 | INTRODUCTION

"Obviously, we don't need to discuss the appropriateness of reflexive evaluation today, for we know by now that this is a valid method."

(dr.prof. Bas Arts, PBL's chief scientist,[2] during the final internal presentation of the Inter-Administrative Programme Vibrant Rural Areas reflexive evaluation, February 11th 2021)

With this quote, the chief scientist of the PBL Netherlands Environmental Policy Assessment Agency (Dutch: *Planbureau voor de Leefomgeving*, henceforth PBL) circumvented the discussion on the validity and acceptability of knowledge co-production in policy evaluation within the organisation. Upon reading it, one might forget there was a time when knowledge co-production in policy evaluation at the PBL was anything less than straightforward. When in 2014 a large-scale evaluation of the Natuurpact policy programme commenced and the idea was coined that nature policy's complex and intractable character would benefit from a deliberative and reflexive research approach aimed at knowledge co-production, the conditions conducive to such an approach were hardly in place. Well-maintained organisational standards of independency, objectivity and autonomy as well as the custom of emphasising accountability and compliance in the policy context worked to keep the domains of science and policy strictly apart, and made the implementation of knowledge co-production in policy evaluation like rowing upstream.

Knowledge co-production in policy evaluation, also referred to as reflexive evaluation, or '*lerende evaluatie'* in Dutch, has since been identified by the PBL as an important expansion of its research repertoire to enhance its reflexive practice. It is considered a crucial approach to sustain the organisation's credibility and legitimacy within the Dutch science-policy arena in light of wicked policy issues. While hardly the organisation's first undertaking of knowledge co-production, the Natuurpact research project was its first endeavour of knowledge co-production in *evaluation* of its scale and duration. Over the past five years, I've had the opportunity to study how the process of normalisation of knowledge co-production in policy evaluation takes shape in the organisation. Not as replacement for other standing approaches to policy research, but as valid approach in its own right and embedded within the organisation's wider research repertoire. I have accompanied policy researchers at the PBL on their journey and observed them navigating (social) structures rooted in traditions of technocracy and rationality in their search for a more deliberative and

[2] The chief scientist advises the Executive Board on the organisation's scientific quality assurance and scientific integrity and is an important driver of methodological innovation.

reflexive research practice better equipped at producing socially robust knowledge for today's complex environmental and sustainability problems. Over time, I witnessed how capacity was built and space negotiated to allow the emergence and normalisation of knowledge co-production in policy evaluation, exemplified by the initiation of various other policy studies marked by deliberation and knowledge co-production in the organisation. Nevertheless, this journey towards a practice of reflexivity remains on-going and full of challenges and dilemmas as technocratic traditions continue to be privileged by policy researchers and policy actors alike, with the risk of tokenising participation and retaining the status quo for science-policy interfaces.

1.1 Why is the normalisation of knowledge co-production in policy evaluation an important object of study?

Overall, the emerging practice of knowledge co-production at the PBL may be appreciated in the light of wider calls across science, policy and society for reflexive modes of knowledge production (reflexive research, in short). Such modes are perceived as better equipped to address the wicked and intractable character of environmental and sustainability issues, such as climate change, global food security and loss of biodiversity. One of reflexive research's basic tenets is the appreciation of the limitations of science in providing ultimate truths for dealing with wicked problems (Jasanoff, 2003; Sarewitz, 2004). Reflexive research is issue-driven and aimed at developing actionable knowledge that is socially robust to inform decision-making (e.g., Arnott et al., 2020b; Funtowicz & Ravetz, 1993; Nowotny et al., 2001). In light of this, a move away is visible from technocratic knowledge interactions in which scientific knowledge has primacy, towards more inclusive and deliberative interactions between science, policy and society. With reflexive research, academic and non-academic actors collaborate in processes of knowledge co-production marked by mutual learning and knowledge integration (e.g., Pohl et al., 2010; Scholz & Steiner, 2015a).

Knowledge co-production's popularity has been accompanied by a lot of scholarly attention for its theoretical and epistemic underpinnings, assessing its societal impacts and providing conceptual and methodological suggestions to advance its quality. A large part of the literature has been concerned with the observed discrepancy between its theoretical ambitions and the actual outcomes that are achieved in practice – which tend to be of a rather less transformative nature than is theorised in literature – and with formulating suggestions on how to remedy this discrepancy (Felt et al., 2016; Flinders et al., 2016; Jagannathan et al., 2020; Scherhaufer, 2014). The argument has been made that, in practice, knowledge co-production processes appear

to closely resemble their technocratic counterparts they are presumed to diverge from, embodying classical ideas on how science and policy should (not) interact (Reinecke, 2015; Turnhout et al., 2013; Van der Hel, 2016). As opposed to its transformative and empowering ideals, the inclusion of policy and societal actors in the research process is observed to have a mostly instrumental character, as a consequence of which participation has become (unwittingly) tokenistic and technocratised (Braun & Könninger, 2018; Chilvers, 2008). Scholars have argued that, rather than the transformation of science-policy relations that was promised – e.g., the '*new social contract'* of science (Gibbons, 1999) – the status quo of traditional interactions is sustained as researchers are perceived to do *'more of the same under a different name'* (Van der Hel, 2016:173). It appears that in their pursuit of knowledge co-production's theoretical ideals in real-life settings, researchers run into barriers that compromise its process and outcomes.

Scholars have pointed out such barriers in the political, social and institutional contexts that surround knowledge co-production to explain the observed discrepancy between theory and real-life settings (Braun & Könninger, 2018). It has been argued that the way researchers are able to navigate these barriers determines the transformative potential of knowledge co-production processes (Wise et al., 2014). In light of this, institutional change and the institutionalisation of reflexive research is called for (e.g., Flinders et al., 2016; Kueffer et al., 2012; Schneidewind & Augenstein, 2012; Yarime et al., 2012; Zweekhorst et al., 2002). Yet, so far, there has been relatively little empirical or theoretical attention for the process by which institutionalisation might be achieved and what it might look like (Arnott et al., 2020a). It is to this particular matter this thesis seeks to contribute.

1.2 Focus, scope and aim of this thesis

In this thesis, I explore how a practice of knowledge co-production in policy evaluation becomes *normalised*. With normalisation, I refer to a process by which practices are implemented, embedded and integrated into the daily life of policy researchers (May & Finch, 2009). I explicitly do not mean that knowledge co-production becomes the new normal, thereby replacing all other approaches to knowledge production. Rather, I refer to a situation where knowledge co-production has become *just as normal* as other approaches, and to a degree of ease or self-evidence with which knowledge co-production is selected as appropriate approach and executed.

The focus of the research presented in this thesis is on the work that policy researchers do to support normalisation, and how their work interacts with the contexts in which knowledge co-production is introduced. I adopt a practice

perspective to access the practical challenges and dilemmas policy researchers encounter, the actions they undertake to address these (or not) and the interactions with the organisational and policy contexts that not just form the backdrop against which knowledge co-production occurs, but also determine the confines of what is conventional and deemed acceptable for science-policy interactions.

The central aim of this thesis is to contribute to the advancement of reflexive research in science-policy systems by deepening understanding of the process of normalisation of a practice of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency. I have selected the PBL as paradigmatic case (Flyvbjerg, 2006): a boundary organisation operating on the Dutch environmental science-policy interface. Boundary organisations may be of particular interest to study the interactions between policy researchers and their contexts, as they must continuously ensure adherence to the latest scientific standards while simultaneously guard alignment to developments in their social and political contexts to guarantee societal and political legitimacy and credibility (Pesch et al., 2012; Pielke, 2007). The PBL organisation has the increasingly explicit ambition to practise reflexive modes of research, given the uncertain, intractable and multi-actor, multi-level governance character of the environmental policy issues the organisation attends to, and their implications for PBL's role and position. It follows that PBL presents an opportune case to study in detail how the process of normalisation of knowledge co-production takes shape.

In the following chapters I first elaborate on this thesis' central theoretical concepts and frameworks (**Chapter 2**), followed by the research design including the general research approach, research sub-questions and studies on which the research in this thesis builds (**Chapter 3**). **Chapter 4 to 8** present the studies' findings, which are used to answer the research sub- and main question in **Chapter 9**. **Chapter 9** also includes a reflection on the research approach, and a general discussion and propositions for ways forward for both research and practice.

2 | THEORETICAL BACKGROUND

In the previous chapter I introduced the increasing calls for opening up science to policy and society in light of environmental wicked problems, and the call for reflexive modes of research this gave fruition to. I also touched upon the challenges with putting the theoretical ideals of new modes of knowledge production to practice in real-world settings. In this chapter I elaborate on these calls and novel modes of research as well as the challenges in more detail, and propose a way to study these challenges and how these are addressed by policy researchers.

2.1 Shifts in the relationship between science and policy

Modern day society is faced with complex, urgent and wicked sustainability problems, by some referred to as the grand challenges of our time (Felt et al., 2012; Gibbert et al., 2021). They are characterised by their progressive systemic character and complexity due to the interwovenness of social-economic and ecological aspects, the myriad perspectives and opinions resulting in the contestation of facts and values, and an incomplete – or even contradictory – knowledge base (Rittel & Webber, 1973). Developing potential solutions to such problems present an unprecedented challenge for both policy and science alike.

As regards public policy, it is argued that regular policy processes characterised by traditional bureaucratic routines are unsuitable to address wicked problems (Hovik & Hanssen, 2015). With the changing context of public administration, 'more than before, solutions for pressing problems cannot be found within the boundaries of sovereign polities' (Hajer, 2003:175), which has led administrators and policymakers to reconsider their approaches to policy design and execution. An emerging trend is perceived in the transformation from government to governance: from hierarchical control to more pluricentric configurations where power is dispersed over multiple (governmental) levels and multiple actors, including citizens and industry (Hajer, 2003; Lo, 2018). For this 'energetic society' characterised by active citizenship and social entrepreneurship, Van der Steen et al. (2015) have conceptualised four perceived roles of government: the *lawful, performing, networking* and *participatory* government. Each role differs as regards the degree of governmental (top-down) management of policy goals and the levels of involvement and participation with society. While the lawful and performing government pertain to classic government models such as Traditional Public Administration (TPA) and New Public Management (NPM) traditions, the networking and participatory government – also referred to as network or collaborative governance (e.g., Evers et al., 2020; Klijn et al., 2010) – may be understood as to fall under the emerging paradigm of New Public Governance (NPG) (Nederhand et al., 2019a). NPG draws from the scholarly body of work on co-production and assumes citizens and other societal actors take active part in

developing policy and delivering public services; stakeholder engagement with policy is generally considered essential for dealing with wicked problems (Cuppen, 2012). In parallel to the rise of NPG, scholars have made the case that wicked problems demand a reconsideration of the dominant institutions and routinised ways of thinking and doing, as these are in many ways part of the problem (Beck et al., 1994), and require societal innovation and transformation towards more sustainable societal systems (Voss et al., 2006). It is argued that the governance of wicked problems should therefore also include a 'reflexive perspective', meaning that things that are habitually done and taken for granted are challenged for their self-evidence (A. Loeber et al., 2007). *Reflexive* governance strives to do just so, as a mode of governance that encourages actors to (re)consider their underlying assumptions, institutions and practices to promote sustainable development (Hendriks & Grin, 2007). **Box 2.1** provides some background information on these developments as regards the Dutch science-policy system in particular.

Box 2.1 The Dutch science-policy system: legacy of the Poldermodel

Science-policy systems (or policy advisory systems; Halligan, 1995) comprise complex constellations of interdependent scientific, public and societal actors and organisations that interact and compete to produce knowledge to support policy-decision making (Van den Berg, 2017). These constellations of actors and their relationships are unique to each country (and within each country, may differ from policy sector to policy sector) (Craft & Howlett, 2013) as regards which actors are included, the role of the public and the authoritative position of knowledge institutes (Hermann et al., 2015; Kunseler, 2017). Country-specific political cultures (Halffman, 2005; Kunseler, 2017) that reflect the dominant norms for the appropriate relationships between science, policy and society, may explain these differences.

In the Netherlands, the political culture is characterised by elements of New Public Management (NPM) bureaucracy, technocratic discourses, as well as corporatist traditions and a deliberative, consensus-seeking approach in which societal interest groups have the opportunity to lobby their interests to the state (in Dutch also referred to as the 'Poldermodel', characteristic to European continental countries including Austria, Germany, Belgium and the Netherlands) (Van den Berg, 2017). In the Netherlands, this consensus-seeking approach – focused on deliberation and mutual learning – may be traced back to the 1980s, when government, labour unions and employer collectives adopted this model to derive to win-win outcomes for economic and social reform. The Poldermodel has been appraised for its importance for remedying high unemployment rates in the '80s and the growth of the Dutch economy in the '90s (Keune, 2016; Visser & Hemerijck, 1997). While engaging the public in policy processes has become fashionable in Europe since the late 1990's (Hagendijk & Irwin, 2006), in the Netherlands there appears a somewhat longer historical basis policy actors may draw upon in their pursuit of the New Public Governance (NPG) paradigm and reflexive governance.

The increasingly intractable and systemic character of environmental and sustainability issues, accompanied by the above trends of network and reflexive governance, have had their repercussions on how the relationship between science and policy is understood and the subsequent position of knowledge institutes as regards decision-making processes. This relationship has long been perceived as a linear one, in which the worlds of science and policy are believed to be strictly separated. This perspective draws from 18th-century Enlightenment ideals and traditions of technocracy and rationality, as well as notions of value-free and impartial science, and is rooted in epistemologies such as positivism and empiricism (Kunseler & Vasileiadou, 2016; Lentsch & Weingart, 2011). Also referred to as normal science (Funtowicz & Ravetz, 1993) or Mode-1 knowledge production (Gibbons et al., 1994), researchers in this view are associated with producing objective and independent facts free from social values. Supposedly, these facts are subsequently transferred to policy actors who use them for public decision-making and taking action. Public welfare is thereby linearly advanced, or so is thought (Dahler-Larsen, 2012; Jasanoff, 2011). Science in this understanding allegedly 'speaks truth to power' – a notion that draws short in the face of wicked problems, as perceptions of truth turn out to be pluralistic and value-laden, and as power has become dispersed over pluricentric networks of governance. The linear understanding has been criticised by scholars in the field of Science, Technology and Society (STS) who argue that the distinction between science and society in real-world settings is at best artificial (Latour, 1987), and that knowledge production and social order co-evolve in intertwined manners. From this follows that knowledge is at the same time '*a product of social order and […] constitutive of forms of social life'* (Jasanoff, 2004:274, in Van Der Hel, 2016).

Accompanied by this shift in how the science-policy relationship is understood is another development that has affected the position of scientific knowledge and expertise in society. A characteristic of wicked problems is that they are marked by knowledge controversies. While controversy is part and parcel of all scientific advancement, some controversies are persistent. They tend not to be just about disputes over scientific facts, but are interlaced with contestations over the political and societal implications of that knowledge (Turnhout et al., 2020). In other words, these controversies are for a large part also about values and interests. Under the header of 'post-truth' society or politics, it has become increasingly common for (populist) politicians (but also citizens, industry and other societal actors) to disregard evidence and critical thinking to inform action, in favour of interpreting and mobilising facts in line with personal opinions, interests and beliefs (Beck & Mahony, 2018; Turnhout et al., 2020). The Trump presidency has become widely associated with post-truth politics, and its negligence of climate change is commonly perceived as an

exemplar case (Kunseler, 2017). More recently, similar developments may be viewed internationally as regards the public controversy and rejection of scientific facts about the Covid-19 pandemic by politicians and citizen actors alike. In a post-truth society the credibility and legitimacy of scientists and scientific institutions are no longer self-evident, and their epistemic and political authority are in dispute. Paradoxically, it seems that the more urgently scientific advice is called for, the more its authority is questioned by policy actors, stakeholders and citizens (also referred to as the 'paradox of scientific authority'; Bijker et al., 2009).

2.2 New modes of knowledge production

The awareness of the interwovenness of science and policy and the increase in public controversy, and subsequent demise of scientific authority have given rise to new ideas on the process of knowledge production and the roles of researchers, and policy and societal actors during this process. Such orientations appreciate the limits of science in providing the ultimate answers for dealing with wicked problems, and seek to open up the process of knowledge production and decision-making to other forms of knowledge, such as experiential and practical knowledge (Sarewitz, 2004). In light of this, a broader range of actors is included in democratic and deliberative processes to co-produce socially robust knowledge geared towards societal transformation and sustainable development (e.g., Bunders et al., 2010; Pohl et al., 2010; Scholz & Steiner, 2015a)

This democratic and deliberative outlook on knowledge production forms the base of diverse scientific traditions and approaches that have found uptake in various fields, including sustainability science, environmental policy studies and policy science. In their suggested paradigm of post-normal science, for instance, Funtowicz & Ravetz (1993) argue that normal science is inadequate for producing valid knowledge for issues characterised by high levels of uncertainty and high decision stakes. They argue that the quality of knowledge should be assured by engaging those actors who have a stake in the decision in dialogue and mutual learning processes. These ideas have been furthered in Gibbons' and colleagues conceptualisation of Mode-2 knowledge production (as opposed to Mode-1) (1994). The case is made for a new social contract between science and society, through transparent and participatory interactions that guarantee that scientific knowledge is socially robust, meaning that it must be scientifically robust while 'also sensitive to a much wider range of social implications' (Gibbons, 1999:C82). Somewhat more recently, Responsible Research and Innovation (RRI) has found traction in European policy cycles and may be perceived as a concept aimed at embedding a more democratic and responsible mode of knowledge production to better align research and innovation processes and outcomes with

societal values and needs (Klaassen et al., 2018; Owens et al., 2004; Stilgoe et al., 2013).

In fields including policy science and public administration, new modes of knowledge production have gained similar resonance. In the face of high levels of uncertainty and complexity of wicked problems, policy makers increasingly seek for ways to design 'robust' policies, i.e., policies that deliver despite turbulent and changing conditions (Van der Steen & Van Twist, 2018). Ideas to enhance such robustness may for instance be found in Deliberative Policy Analysis (DPA), an approach proposed as an answer to the mismatch between technocratic and positivist types of policy analyses and the increasingly networked styles of governance (Hajer & Wagenaar, 2003). Its advocates argue policy analysis should be practice-oriented, interpretive and deliberative in order to have political authority and legitimacy (Li & Wagenaar, 2019).

Reflexive research

Each of these various traditions has its own particularities and focus, and has found uptake in differing – although associated – fields. At the same time, they share a number of interrelated theoretical and practical elements that I wish to emphasise for the purpose of this thesis. To start, these traditions are *issue-driven* and aimed at producing *actionable knowledge* (e.g., Arnott et al., 2020b; Funtowicz & Ravetz, 1993) that is relevant, legitimate and credible – or, socially robust – to inform decision-making to address this issue (Cash et al., 2002; Nowotny et al., 2001). Second, a *changed relationship between knowledge producers and knowledge users* is encompassed in these traditions. The role of researchers is not to produce value-free facts that can be unilaterally used by policymakers to inform decision-making, nor to develop definitive solutions to end political discussion. Rather, researchers facilitate policy and societal actors' capacity for deliberation and collective learning (Hajer & Wagenaar, 2003). Their role shifts from pure scientists or issue advocates, to facilitators of mutual learning processes, knowledge brokers and change agents (Pielke, 2007; Pohl et al., 2010; Turnhout et al., 2013). At the same time, policymakers are no longer perceived as passive recipients of facts, but as legitimate knowledge holders and co-investigators who play an active part in the knowledge production process (Tengö et al., 2014; West et al., 2019). Finally, and strongly related to the previous notions, these traditions share that they seek to *advance reflexivity* both of science and scientists themselves, as of the policy and the actors who constitute the policy field under scrutiny.

Reflexivity, as it was originally introduced in the 1990s, refers to how modern society inevitably impacts itself due to its own continuous development. The modernisation of modern society has come to impact itself by producing (unintentional) negative side-effects of otherwise legitimate activities, via ambiguous, uncertain and complex routes

that have proven impossible to predict in full (Beck et al., 1994; 2003). Think, for instance, on the use of fossil fuels and its importance for powering the industrial revolution and the subsequent societal advancements, including reduced global poverty and enhanced food security. At the same time, the extensive use of fossil fuels has induced possibly irreversible climate change (paradoxically threatening the acquired welfare and food security, amongst other societal and environmental aspects), demanding us to reconsider how we produce and use energy on a global scale. This was never the intention when the production of energy via fossil fuels was initiated, and demonstrates how '*the environment is turning back on society'* (Beers & Van Mierlo, 2017:244). Reflexivity, as such, is considered an inherent feature of wicked problems and the nature of any modern society. Voss et al. (2006) refer to this as first-order reflexivity.

Second-order reflexivity captures what Beck and colleagues understand as pro-actively dealing with the detrimental effects of reflexive modernisation and refers to someone's awareness of the reflexivity of his or her own situation in a society marked by first-order reflexivity (Voss et al., 2006). As Beers & Van Mierlo (2017) explain, second-order reflexivity goes beyond mere reflection on society, and also involves understanding the implications of those reflections for one's own behaviour and actions: a so to speak '*reflection-on-reflection'* (Beck et al., 2003:16, in Beers & Van Mierlo, 2017). In this light, reflexivity also becomes understood as a (human) capacity and as a virtue of experts who are aware of their position, assumptions and understanding of the world, how these affect their problem framing and interpretation of produced knowledge, and the subsequent political and societal implications of their work (Fischer, 2009; Jasanoff, 2003). Second-order reflexivity as capacity is also apportioned to policy programmes or interventions aimed at societal transformation to indicate their ability to interact with and affect the societal and political environment within which it operates (Van Mierlo et al., 2010). Reflexivity as a capacity arguably may be more or less prevalent and can be advanced to support societal transformation and sustainable development (Beers and Van Mierlo, 2017) by approaches to knowledge production that encourage self-reflection and that are learning-oriented (Voss et al., 2006).

For the remainder of this thesis, I use the term reflexivity to refer to second-order reflexivity. Furthermore, I understand modes of knowledge production that feature the above mentioned characteristics as *reflexive research*. As this thesis aims at providing insights into how reflexive research may be advanced in science-policy systems, I turn to how reflexive research may be put to practice in such systems next.

Advancing reflexivity through knowledge co-production

Reflexivity may be pursued through a growing range of methods or methodologies which aim at more structural and deliberative science-policy interactions. Approaches such as transdisciplinary research (Hirsch Hadorn et al., 2008; Lang et al., 2012) and knowledge co-production (Miller & Wyborn, 2020; Norström et al., 2020) have found wide traction in international environmental and sustainability science-policy arenas (West et al., 2019). Characterised by opening up the knowledge production process to those actors who have a stake in the issue under study, stakeholders participate and take an active role during all phases of the research to integrate different ways of knowing and co-produce knowledge that is deemed actionable and legitimate (Mach et al., 2020).

To gain insight into how reflexive research may be advanced, the research I present in this thesis revolves around the coming into being of a practice of knowledge co-production. I understand knowledge co-production to comprise four theoretical-ideal key-features: *stakeholder diversity and inclusion, reflection and mutual learning, transparency and openness for knowledge integration* and an *emerging and responsive design*.

Stakeholder diversity and inclusion

Including a diversity of stakeholders with policy and science is considered a critical aspect for dealing with wicked problems (Cuppen, 2012a; Leventon et al., 2016). In literature, three rationales for stakeholder participation are generally distinguished (Fiorino, 1990; Scherhaufer, 2014; Stirling, 2008). The first is the instrumental rationale, which views the engagement of stakeholders during knowledge production as necessary for establishing impact. Its dominant reasoning is that by providing stakeholders insight and opportunity to provide input into the development of scientific findings, credibility, legitimacy and support for these findings are increased. The second rational is the normative, which holds that individuals have a right to participate and to have control over decisions that affect their lives. Engaging them with knowledge production processes is the right thing to do and has value in and of itself, regardless of the outcomes. Empowerment (of marginalised groups) is often a goal of the normative rationale. Thirdly, the substantive rationale assumes that stakeholder participation will advance the quality of the knowledge produced via the integration of various knowledges, values and ideas, leading to more socially robust knowledge. Stakeholders provide substantial and elemental input to advance the relevance and practicability of scientific findings for informing decision-making. Recently, Schmidt et al. (2020) have added a fourth rationale, namely that of social learning. This rationale holds that involving stakeholders stimulates processes of

social learning required to better understand and solve the complex issue at hand. Stakeholder inclusion is directed at bringing different stakeholders together to establish a network, trust and a shared understanding.

Reflection and mutual learning

The co-evolvement of the understanding of science and policy actors of a social-ecological issue to jointly develop actionable knowledge is considered as one of knowledge co-production's main aims. In other words, facilitating mutual learning processes between researchers from different discipline and non-scientific actors can be considered as one of its central features (Roux et al., 2017). Reflection and mutual learning are considered important proponents for the understanding of one another's problem framing (also understood as social learning), which is widely believed to contribute to new insights and understandings (Cuppen, 2012a) and collective action (Walter et al., 2007). Also, knowledge co-production processes may inform experiential learning and learning-by-doing by encouraging recurrent reflection on how things are going (Lang et al., 2012).

Transparency and openness for knowledge integration

Knowledge integration is widely perceived as crucial means for co-producing action-oriented solutions to complex problems (Hoffmann et al., 2017). It is understood to comprise both interdisciplinary and transdisciplinary (also involving non-scientific actors) knowledge integration. In order to integrate knowledge, there is argued for the involvement of all relevant actors (scientific and non-scientific) during all phases of the research process. Knowledge integration may be established by jointly defining the problem the research searches to address and by co-developing and conducting the research (Mauser et al., 2013). Through shared interpretation of the findings and drawing joint conclusions, different perspectives and ideas may be incorporated to ideally develop knowledge that is perceived of as relevant, legitimate and actionable through the eyes of all those involved.

Responsive and emergent design

Finally, processes of knowledge co-production feature a responsive and emergent design (Regeer & Bunders, 2009). Initially, its process (and, sometimes, also its goals) is more globally described, focusing on planning only the subsequent phase in more detail. As insights and experience increase – on the basis of observation and reflection – the following phase gains purpose and can subsequently be planned. Designing the process this way allows for research to be responsive to newly acquired knowledge as well as to contingencies due to unexpected (external) developments that affect the goal of the knowledge production process.

Knowledge co-production in policy evaluation

These ideas on knowledge co-production and its features have found wide uptake in academic, commercial and policy settings alike. As touched upon previously, they resonate particularly in fields of policy science and public administration, illustrated by traditions such as DPA. However, as regards policy research approaches, common practice in most societies of the Global North is to predominantly focus on accountability and compliance, and on impact assessment based on pre-defined outcome criteria (Chouinard, 2013; Van Twist et al., 2015). This practice is rooted in NPM ideals and holds a technocratic and linear view on knowledge production and science-policy interactions. It is argued that, in the face of wicked problems which require reflexive governance and adaptive capacity, evaluation for accountability and compliance becomes meaningless (Van Twist et al., 2015). From the 1990s onwards, there has been called for participant-oriented and collaborative approaches to policy evaluation aimed at advancing policy learning in network settings that are fraught with uncertainty and complexity (Borrás & Højlund, 2015; Guba & Lincoln, 1989; Patton, 2010; Shulha et al., 2015). Scholars have additionally argued for policy evaluation approaches to include a system perspective to stimulate system transformation (Borrás & Laatsit, 2019; Moore et al., 2019) as well as a reflexive view, to advance the reflexivity of policy makers (and their policies) as regards their relationship to these systems (Arkesteijn et al., 2015; Botha et al., 2016; Regeer et al., 2009). Indeed, policy evaluation is considered increasingly paramount for informing reflexive learning in support of reflexive governance (Sanderson, 2002). Knowledge co-production and its features may thus also be recognised in new forms of policy evaluation. However, scholars have found that knowledge co-production in policy evaluation is faced with difficulties with putting its theoretical ideals to practice, as the predominant structures of science-policy systems are not always conducive to these ideals. In the following section, I further explore these difficulties.

2.3 Understanding how to move from theory to practice

Challenges with practising knowledge co-production in policy evaluation

This thesis centres around exploring how theoretical ideals are put into practice and understanding how knowledge co-production is normalised in real-world settings. With the wide uptake of knowledge co-production in international science-policy arenas to address complex contemporary sustainability and environmental problems, the scholarly body of literature has proliferated with contributions aimed at theoretical and epistemological advancement (Regeer & Bunders, 2003; Scholz & Steiner, 2015a), assessing co-production's long term, contingent and sometimes unintended effects, and demonstrating and explaining co-production's societal impact (Walter et al., 2007;

Wiek et al., 2014). A significant part of the literature is dedicated to providing conceptual and methodological suggestions to further the quality and impact of co-production processes (Bergmann et al., 2021; Jahn & Keil, 2015; Lang et al., 2012; Norström et al., 2020). In this vein, scholars have recently begun to point out a discrepancy between co-production's theorised outcomes of actionable knowledge and societal transformation and the outcomes that are achieved in practice (which tend to be of a less transformative nature) (Flinders et al., 2016; Jagannathan et al., 2020). It seems that in reality, actors who seek to enact knowledge co-production run into various challenges that make it difficult to put co-production's theoretical ideals to practice.

For instance, there is found that the instrumental rationale for stakeholder participation often gains prominence, a situation which is criticised for eliciting tokenistic stakeholder participation and undermining true democratisation of science (Wynne, 2006). Institutions may advertise the democratic character of their knowledge production process and the subsequent social robustness of their findings, while in reality they are not very interested in being responsive to alternative perspectives. Here, negative researcher bias towards the claims of stakeholders may form a profound barrier to open dialogue, as researchers tend to be in positions of power as regards how issues are framed in relation to stakeholders and citizens (Cuppen et al., 2009). Some scholars even report outright failures of co-production processes during which the status quo for marginalised groups was reinforced rather than transformed, as was the intention (Felt et al., 2016; Turnhout et al., 2020). Similarly, inter- and transdisciplinary knowledge integration arguably remains poorly understood and expertise on how to establish knowledge integration is found to be lacking (Bammer et al., 2020). Consequently, co-production processes run the risk of being mostly guided by prevailing power dynamics and serving the interests and beliefs of those of power, and less by democratic interactions for mutual learning and collaborative action.

Explanations for the difficulties with attaining knowledge co-production's proposed outcomes are found in societal, political, cultural and institutional structures that characterise the science-policy systems of societies in the Global North, and which privilege more traditional modes of knowledge production and respective science-policy interactions (Felt et al., 2007; Flinders et al., 2016; Scholz & Steiner, 2015b). For instance, as regards policy evaluation, accountability and compliance have become institutionalised to an extent that the golden standards of evaluation are marked by objectivity and impartiality. At face value, such standards appear to debar more responsive and deliberative evaluation approaches. This has severe implications for their implementation, as from the perspective of accountability any form of interaction

may be understood as to undermine the credibility of its researchers (Chouinard, 2013). In the face of such unconducive standards for what determines expertise, researchers are said to run the risk of defaulting into classical modes of knowledge production (Van der Hel, 2016), despite their ambition for co-production. **Box 2.2** provides some background information on the emergence of deliberative discourses on expertise in the Netherlands.

Box 2.2 Technocratic and deliberative discourses on expertise in the Netherlands

The trends discussed in Section 2.1 – the move from New Public Management (NPM) to New Public Governance (NPG) and reflexive governance, and the demise of scientific authority – have also affected how in the Netherlands science-policy interfaces are perceived and, consequently, how the expertise of knowledge institutes is valued. While under technocratic discourses their expertise remains largely unquestioned, more deliberative discourses on expertise have become more prominent in the past decennia (Van den Berg, 2016; Kunseler, 2017). This is for instance visible in the aforementioned rise of Deliberative Policy Analysis (DPA), an approach of Dutch origin geared towards modes of policy science that align to the increasingly networked styles of governance (Hajer & Wagenaar, 2003). Somewhat more recently, similar calls have been made for policy evaluation approaches that are public value-driven (Dutch: opgavegericht evalueren), meaning that the evaluation approach is tailored to fit to produce relevant knowledge of the policy issue and its characteristics (e.g., its levels of complexity and prevailing governance style) at hand (Van der Steen et al., 2018). Such developments are also witnessed in other countries, including Austria, Germany and the United States of America (Chouinard, 2013; Hermann et al., 2015; Reinecke, 2015), where science-policy interfaces are increasingly characterised by deliberation. Such interfaces are defined as '*social processes which encompass relations between scientists and other actors in the policy process, and which allow for exchanges, co-evolution, and joint construction of knowledge with the aim of enriching decision-making'* (Van den Hove, 2007:807).

Within the Dutch science-policy system, there are currently three national public knowledge institutes that have a prominent position on the science-policy interface and draw the boundaries for political negotiations between policy actors and societal interest groups, and as such function as 'linesman of politics'. These are the so-called planning agencies (Dutch: *planbureaus*) (Halffman & Hoppe, 2005). These agencies – the CPB Netherlands Bureau for Economic Policy Analysis (*Centraal Planbureau*), the SCP Netherlands Institute for Social Research (*Sociaal en Culturaal Planbureau*) and the PBL Netherlands Environmental Policy Assessment Agency (*Planbureau voor de Leefomgeving*) – are policy research agencies that conduct independent and scientific policy evaluations, outlook or foresight studies, and methodological and conceptual studies, to provide the national government with knowledge about the country's current and future state, and the role of governmental policy therein. By neutral and unpartisan assessment of (potential) policy outcomes, they are portrayed as powerful institutes capable of disciplining policymakers into rational policy-making (Halffman, 2009; Kunseler, 2017).

In the face of the above mentioned trends, the planning agencies are confronted with demands for more reflexive research. Nevertheless, they seem to remain hesitant towards deliberative modes of knowledge production, mostly sticking to traditional role distributions between science and policy

actors and holding scientific knowledge as primary source for scientific and societal advancement (Kunseler, 2017). Similar hesitation has been observed in German public knowledge institutes (Heinrichs, 2015) and is attributed to the concern that acknowledgement of the plural nature of knowledge and expertise will undermine the credibility and authority of policy researchers (Bijker et al., 2009).

In this thesis, one particular planning agency is of interest: the PBL. More so than appears the case for the other agencies, the PBL has made explicit its ambition to innovate towards a more reflexive research practice to address the increasingly complex, multi-actor and multi-level character of the contemporary environmental policy issues the agency seeks to study. While this innovation is considered elemental to maintain the organisation's credibility and legitimacy in the Dutch science-policy system, it poses particular challenges for the organisation, as the methodologies and epistemologies, and technocratic discourse on which the organisational research practice is predominantly rooted, at face value appear incompatible with the more deliberative and participatory approaches embodied in reflexive research. This makes the PBL an especially opportune case to study processes of normalisation. In **Section 3.3** I further expound on my decision to select the PBL as paradigmatic case.

The challenges with practising co-production may be understood as the result of complex interactions between researchers who aspire co-production and the contextual structures – such as rules, norms, beliefs and customs – unconducive to co-production they encounter in various settings. The ways researchers navigate these structures arguably affect the transformative potential of knowledge co-production processes (Wise et al., 2014). To advance the understanding of the challenges with practising co-production and how policy researchers navigate these, scholars have made the case for mobilising social theory and adopting a practice-based approach to implementing processes for knowledge co-production (West et al., 2019), which is discussed in the following section.

Practice theory to advance understanding

From the 1970s onward, ideas put forward by practice theorists such as Bourdieu (1977), Giddens, (1984) and Schatzki et al. (2001) have gained resonance to analyse social and organisational processes, ranging from phenomena such as governance, science, consumption, sports and language (Nicolini, 2012). While the landscape of theories of practice is considerable, with each variant having its own distinct features, these theories share a number of tenets that together provide a theoretical lens through which to consider the interactions between policy researchers and the (social) contexts in which they seek to practice co-production (Nicolini, 2012; Spaargaren et al., 2019). To start, part of the appeal of practice theories is that they propose to offer a remedy to dissolve enduring dichotomies – such as actor-system – that remain unsolved by other approaches to study social phenomena. Rather than setting agency

apart from contextual structures that exist in some reified, objective and sterile 'out there', these theories generally share a reflexive perspective on agency and structures. These are both perceived as being in flux and interconnected, and as continuously shaping and being shaped by one another (Arts et al., 2013). Such interactions occur in the dynamics of everyday practices, at the interface where human actors and structures inevitably meet. Adopting a practice view allows to appreciate policy researchers' challenges with implementing co-production and the actions they take to address these from a reflexive perspective on their interactions with the structures they encounter.

Second, practice theories hold that the study of social order should start with practices: practice comes first. It is argued that, to study human behaviour, a move away is required from research that focuses on isolated individuals, their intentions, motives and personal values, towards taking social practices as central units of analysis. What then, comprises a social practice? While many definitions are provided, the most cited (according to Spaargaren et al., 2019) is provided by Reckwitz (2002:249).

> *'A 'practice' … is a routinised type of behaviour which consists of several elements, interconnected to one another: forms of bodily activities, forms of mental activities, 'things' and their use, a background knowledge in the form of understanding, know-how, states of emotion and motivational knowledge.'*

This implies that the focus of inquiry should not be on individual researchers' personal opinions or outlooks on the roles of policy researchers and policy actors during knowledge production processes, but rather should be the doings and sayings of these actor groups to co-produce knowledge and the challenges and dilemmas they encounter while doing so.

Thirdly, practice theories seek to pursue '*the middle ground'* (Spaargaren et al., 2019:4) between subjectivist (i.e., foregrounding agency, personal values and human actions as the determinants that make up society) and objectivist (i.e., foregrounding systemic physical, social, political, juridical and biological structures as determinants for human action, and thereby society) understandings of social order. While practical activities are given precedence over rational decision making and explicitly subjective elements like personal motives, and while social practices are considered as routinised and automatic, this does not automatically mean that all doings and sayings come about non-discursively. For example, when new ideas enter into a practice they might disturb it, for instance when radical innovations or external developments like natural disasters deprive (parts of) the practice of meaning. Subsequently, practitioners may (temporarily) shift into a more discursive, reflexive and conscious mode of decision-

making and acting, until a new status quo is attained – for the time being. As such, practices are continuously open to contestation, and therefore are by definition changeable. From this it follows that the implementation of co-production may be understood as a disturbance of research practice, prompted by the shift in how science-policy relations are understood and the emergence of post-truth discourses that diminish the political authority of science.

Finally, and immediately related to the previous point, human actors' discursive and reflexive mode of doing is defined by the very practices they stem from. That is to say, human actors' capability to transform practices is bounded by the skills, competences, the sense of what is right or wrong and what is considered (un)acceptable and (in)appropriate, that are rooted in their respective practice. Scholars have referred to this phenomenon as 'bounded agency' or 'bounded creativity', underscoring that while agency structures practices, practices also structure agency. This implies that policy researchers who seek to practice co-production are bounded by institutionalised standards and protocols for policy research, as well as by societal and political expectations on the process and outcomes of their policy research.

As regards knowledge co-production, there is one practice theory developed in the field of implementation science that offers particular potential for understanding the (lack of) normalisation of knowledge co-production: Normalisation Process Theory (May & Finch, 2009). In the following section, I expand on this specific theory and its merits.

2.4 Introduction to Normalisation Process Theory

Some 50 years ago, scholars started studying why some innovations succeed to become routinised and embedded into standard practice, while others do not. Scholars in education innovation conceptualised the implementation of such innovations as a process that occurs within a wider context that has a significant effect on whether the innovation finds uptake (Fullan & Pomfret, 1977; Havelock, 1970, 1971; Huberman & Miles, 1984). May & Finch (2009) have built on this work to develop Normalisation Process Theory (NPT). NPT is a theory that focuses on the work that is done by involved actors to normalise a novel practice in contexts marked by complexity and emergence, developed to enhance understanding of how some practices become routinely embedded in everyday life. It has been developed in the field of implementation science and originally has been directed at providing a theoretical and practical lens to understand, guide and evaluate the normalisation of complex health interventions (Murray et al., 2010).

Since its origin, the theory has found uptake in numerous studies and has proven its value for understanding the complex dynamics of implementing and institutionalising new healthcare technologies or innovations, including e-health and telehealth, in complex social systems (May et al., 2018; McEvoy et al., 2014). As the authors emphasise, NPT is relevant not just within the field of healthcare but provides a theoretical and practical framework for investigating the normalisation of any material practice in its social setting (May & Finch, 2009).

May & Finch (2009) favour the term normalisation over concepts such as institutionalisation and stabilisation, as both are perceived to focus on a 'final stage' of a process of implementation and adoption. The authors centralise *normalisation* as primary concept to emphasise the on-going, dynamic and emergent process by which practices are *implemented*, become routinely *embedded* and sustained (*integrated*) in everyday life. As I emphasised in **Chapter 1**, with normalisation I refer to a situation where a knowledge co-production has become *just as normal* as other policy research approaches, not to a situation where knowledge co-production has become the new normal and has subsumed all other approaches. I have chosen to follow May & Finch (2009) in adopting the term normalisation as it captures the complex and on-going dynamics between policy researchers and the (social) structures they encounter in different settings as they seek to practice co-production.

NPT's four core mechanisms

NPT is a practice-based theory in the sense that it focuses on the actions of actors to implement, embed and integrate – normalise – a novel practice. The actions of actors to implement a practice centre around four non-linear generative mechanisms that are interdependent and may occur simultaneously: *sense-making, engagement, enactment* and *appraisal*. I attend to each of these mechanisms briefly (drawing on e.g., May, 2015; May & Finch, 2009; Murray et al., 2010) and operationalise them from the perspective of normalising knowledge co-production.

Sense-making work

Sense-making work is directed at establishing *coherence* and encompasses actors' actions to develop a shared understanding and invest meaning into knowledge co-production. Shared understanding and meaning allow actors to act in concert in its enactment, as they share a basic view on its necessity and merit. Ways to establish coherence include differentiating co-production from other (established) knowledge production practices and specifying its particular features. Notably, for establishing coherence the meaning that is ascribed to co-production should not be externally defined or normatively imposed via top-down procedures, but rather becomes internalised within actors as they learn, share and experience its process.

Engagement work

Engagement work aims to develop *cognitive participation*: the engagement, commitment and buy-in to co-production by actors beyond its original instigators. Engagement is relevant for co-production to find traction and for it to become initiated, as implementation depends upon actors acting in concert and organising themselves, and (material) support. It constitutes work that is directed at developing a 'community of practice' that involves actors that are either directly involved with the knowledge co-production process or whose involvement is of a more symbolic nature, in the sense that it provides endorsement and support for co-production. A condition for engagement is legitimisation: that co-production is perceived as legitimate by a large enough community is essential for it to have viability within, amongst or instead of existing knowledge production practices. Consequently, whether engagement is developed is closely linked to the norms and conventions that reside within the contexts of implementation.

Enactment work

Enactment work consists of the *collective action* actors undertake to enact knowledge co-production.[3] Such work comprises the operationalisation of co-production features in order for it to become workable within the intended context. It is thus concerned with matters of compatibility and the impact a co-production processes may have on established divisions of tasks, resources, responsibility and power amongst involved actors. Enactment work also involves the knowledge that is necessary to enact co-production as it was intended, and the (extensive) training that may be required before they may implement it.

Appraisal work

Finally, appraisal work is intended to encourage *reflexive monitoring*[4] of co-production processes and includes all actions to judge its added value and effectiveness over the course of its implementation. Such actions may take shape in the form of formal or informal appraisal and evaluation. It helps actors understand co-production's

[3] Enactment work, or collective action, was originally referred to as Normalisation Process Model (NPM). It consists of four sub-elements (contextual integration, relational integration, interactional workability and skill set workability) and aims at explaining factors that affect actors' collective behaviour to implement a practice. As it did not enhance understanding of why actors engaged or supported a new practice (or not) and how they valued it, NPM later became subsumed in NPT. For more details on NPM see May (2006).
[4] Not to be confused with Reflexive Monitoring in Action (van Mierlo et al., 2010), an approach for a type of monitoring that seeks to enhance the reflexivity of projects and programmes (and its initiators) that are directed at system innovation.

(dis)advantages and develops their understanding of its effects. In doing so, appraisal work especially may inform the other mechanisms as it feeds into the meaning actors ascribe to co-production, its perceived legitimacy and actors' ideas on how it should be enacted to stay true to its original purpose.

These four core mechanisms of NPT focus on the work actors do – or, the agency that they have – and constitutes what they contribute to enact knowledge co-production. This work, however, does not tell the whole story of how co-production may become normalised, as the work actors do is governed by the dynamic structures of the context in which it is introduced, as well as by co-production's features, and how well these align to these contextual structures. I attend to these next.

Elasticity, plasticity and readiness for change

Contextual elasticity

Specific settings or contexts constitute the dynamic environment in which practices are introduced and implemented. Such contexts comprise institutionalised normative and relational rules and conventions that shape the agency actors have for implementing a co-production, as they determine the social norms (institutionalised rules that govern actors' behaviour) and social roles – (institutionalised identities and relationships which frame science-policy interactions) (May, 2013). Additionally, agency is determined by the material and cognitive resources that transpire within the contexts of implementation. This includes, for instance, access to infrastructures, information and knowledge that are required to enact co-production (ibid). Importantly, alike most practice theories, in NPT these contextual structures are not perceived as static but as continuously in flux. May et al. (2016) say that contextual structures have *elasticity*: they can be stretched or compressed to make space for co-production features to allow them to fit. The higher the degree of elasticity, the less work (via the aforementioned mechanisms) actors have to do to enact knowledge co-production.

Plasticity of practice components

Practices comprise various components or features that actors are required to mobilise and make operational in order to enact it. For instance, as discussed earlier, co-production's key feature include stakeholder participation, knowledge integration and an responsive and emerging design. These features require operationalisation to be put to practice. May et al. (2016) argue how practices and their components have *plasticity*: they are malleable and can be modified to fit the contexts in which they are introduced. The higher the degree of plasticity, the more discretion actors have with operationalising co-production features to align to contextual structures. A certain degree of modification is generally required when new practices are introduced in

order for them to have workability and fit with prevailing social roles and norms. Here, however, scholars forewarn compromising the practice's original intent by modifying the practice beyond the '*zone of drastic mutation'* (Roitman & Mayer, 1982:3).

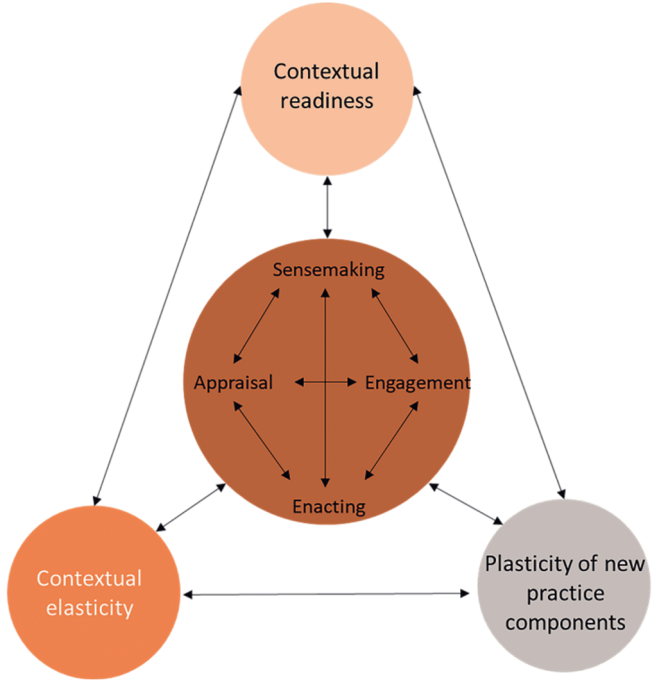

Figure 2.1 Schematic overview of NPT's core mechanisms and concepts to advance understanding of how knowledge co-production is practiced and becomes normalised (or not) in its context of implementation (after May & Finch, 2009; May, 2013 and May et al., 2016).

Contextual readiness for change

Finally, whether co-production may succeed to become standard practice depends upon the potential for its normalisation residing within the contexts of implementation. May (2013) says that this potential is determined by contextual readiness: the intention of individual actors to take part in co-production processes and their motivation to advance its normalisation, as well as the collective commitment of a community of actors to implement and advance co-production, despite possibly incongruent contextual norms and identities. This community shares the belief that the normalisation of co-production is possible and necessary. Such contextual readiness may be more or less prevalent prior to the introduction of knowledge co-production, and may for instance be influenced by wider discourses on knowledge production processes and perspectives on science-policy interfaces.

3 | RESEARCH DESIGN

With this thesis I aim to contribute insights that can help the advancement of reflexive research in incumbent science-policy systems of the global North. To this end, what happens when reflexive research theory meets real-world situations was studied by an in-depth and situated exploration of the process through which an emerging knowledge co-production practice in policy evaluation is normalised at the PBL Netherlands Environmental Policy Assessment Agency. This research aim translates into the following main research question:

> *How does the process of normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency take shape?*

In this chapter I describe the research design of this thesis. The main research approach and the decision for the PBL as a paradigmatic case are motivated, and the research sub-questions and methods to explore these questions are discussed. At the end of this chapter the validity of this research is considered.

3.1 Practice-based approach

The research I present in this thesis centres around the interactions between policy researchers who aspire knowledge co-production in policy evaluation and the (social) structures they encounter in the contexts of implementation that affect their practice. In **Chapter 2** I made the case for a practice-based perspective to advance the understanding of these interactions. From the tenets that I introduced follow a couple of presumptions that guide the research on which this thesis is based:

i. To start, to access the practice under study, practice comes first in my focus of analysis. This means that rather than inquiring after the personal opinions or ideals of individual policy researchers, I focus on their doings and sayings to implement knowledge co-production and to address the challenges they encounter in light of this.

ii. Secondly, I presume that agency and structures are interconnected, and continuously shape and are shaped by one another. The actions of policy researchers to implement co-production affect the contextual (social) structures they encounter and vice versa: they co-evolve.

iii. Third, I presume practices may be disturbed by internal or external developments, in response to which actors may shift from a routinised mode of decision-making and acting into a discursive one. Practices are thus changeable and in continuous evolution. I understand knowledge co-production as an innovation of PBL's organisational research practice and as a response of policy researchers to external developments - i.e., the increasingly uncertain,

intractable, multi-actor and multi-level governance character of contemporary environmental and sustainability policy issues and their implications for PBL's role and position in the science-policy system – that have disturbed it.

iv. Finally, I presume that policy researchers who seek to adopt knowledge co-production in their research projects are bounded by institutionalised organisational standards and protocols for policy research quality, as well as by societal and political expectations on the process and outcomes of their policy research that transpire in the wider contexts in which their projects occur. At the same time, by bumping up against these boundaries, I expect they also stretch these and thereby create space for innovation.

Practice-oriented research is concerned with studying phenomena in their real-world setting. To do so, a naturalistic and interpretive approach is generally adopted, that includes methods which allow for 'praxeologising' (Spaargaren et al., 2019:2): obtaining a thick empirical understanding of the practice under study. Practice theorists argue that only by actively participating in the practice of inquiry may researchers develop a true practical understanding and access insider knowledge. Participant observation is a proven method to gain such knowledge (Schmidt, 2016). As opposed to the view in the social sciences that sees researchers as detached observers of social phenomena, which traditionally is more dominant, practice-oriented research approaches comprise interpretive and collaborative work to access the challenges and dilemmas studied. As such, researchers may adopt a role of researcher-practitioner (Schön, 1983) and learn alongside other practitioners with and from their situated experiences. They thereby obtain a thick, in-depth understanding of the practice under study aimed at generating knowledge that simultaneously holds merit for the practice of interest, as for foundational or theory-building purposes (Candy, 2006).

Researcher-practitioners become deeply engrained in the practice that they study; distance to this practice cannot be maintained, nor would be functional as it would obstruct the insider view that is searched for. Nevertheless, too much familiarisation with the practice and its constituents may compromise analytical distance. This does not necessarily presents a problem for the position of the researcher-practitioner, nor necessarily compromises her ability to conduct her research with the appropriate distance. But, in order for this to go well, it is vital she is able to adopt a self-reflective stance towards her own biases, identity and assumptions, and their potential implications for how these might affect how research is interpreted (Zuiderent-Jerak, 2007). Yanow (2000) suggests that such a reflective stance is supported by the *'juxtaposition of the analyst's "estrangement" from the analytic situation and her growing familiarity with that situation'* and the balancing act *'between "stranger-ness and "insider-*

ness"' (p:7). This balance is supported by moving in and out of the practice, for example, through working in research teams in which members (some of whom are not involved with the practice under study) challenge each other's interpretations and normative standpoints, and through exercising (extended) peer review and installing (external) review committees.

3.2 Empirical setting: the PBL Netherlands Environmental Policy Assessment Agency

As discussed previously (**Box 2.2**), the Dutch science-policy system currently includes three national public knowledge institutes with authority which reports get prominent attention by media and politics alike and are mostly accepted as an unquestioned representation of the state of affairs (Halffman, 2009). The function of these planning agencies is defined in the Protocol for the Policy Assessment Agencies (Dutch: *Aanwijzingen voor de Planbureaus*) (Staatscourant, 2012) which includes three core guiding principles: the agencies are to produce knowledge that holds relevance for strategic policymaking, that is based on current scientific standards and that is produced independently from everyday policy concerns. The rhetoric that underlies these guiding principles and their institutionalisation via the formal Protocol are illustrative of the technocratic-positivist paradigm under which the planning agencies have been established (Kunseler, 2017).

In this thesis, one particular planning agency is of interest: the PBL, which I propose as paradigmatic case (Flyvbjerg, 2006). The PBL is an organisation that performs a role in linking science and policy decision-making: it is a boundary organisation operating on the Dutch environmental science-policy interface. For boundary organisation to maintain societal and political legitimacy and credibility, they must continuously adapt to changes in their wider social and political contexts, while at the same time managing to align both to policy actors' knowledge demands and the latest scientific standards (Pesch et al., 2012; Pielke, 2007). They are thus profoundly bounded by their social, political and scientific contexts. As such, the PBL finds itself in a somewhat paradoxical situation in which it has the ambition to innovate towards more reflexive modes of research to better address wicked and intractable socio-environmental problems, while at the same time the organisation cannot elude more traditional expectations on the role and function of policy researchers that reside in its societal and political context. This makes them of particular interest to study the interactions between policy researchers who seek to innovate their research practice and these contexts. I assume that this thesis' inquiry into the normalisation of knowledge co-production for policy evaluation at the PBL has relevance for other (environmental) boundary organisations with similar aspirations and position.

In this section I introduce the PBL and its aim to normalise co-production into its organisational research practice, and motivate my decision to select it as primary case in this thesis' research.

An introduction to the PBL and its ambition for normalising knowledge co-production

PBL describes itself as the national institute for strategic policy analysis in the fields of the environment, nature and spatial planning. The organisation strives to contribute to '*improving the quality of political and administrative decision-making by conducting outlook studies, analyses and evaluation in which an integrative approach is considered paramount*' (PBL, 2019). It was created in 2008 as result of a merger between the MNP Netherlands Environmental Assessment Agency (Dutch: *Milieu- en Natuurplanbureau*) and the RPB Netherlands Institute for Spatial Research (*Ruimtelijk planbureau).* Organisationally, the PBL is part of the Dutch Ministry of Infrastructure and Water Management (*Ministerie van Infrastructuur en Waterstaat*), which is also the institute's primary principal (although other ministries may also commission the PBL to conduct research in its fields of interest). Most of the organisation's research is government funded, supplemented with (international) research funding. Although the PBL technically falls under the authority of the Ministry, as mentioned earlier, PBL's status as independent, autonomous research institute is secured in the Protocol for the Policy Assessment Agencies (Government Gazette, 2012) and, as witnessed by the organisations' mission and vision statements (PBL, 2017a; 2017b), this independent and autonomous status is a hallmark of PBL's self-image.

In the PBL Vision 2025 (2017b) (and also addressed in the organisation's external inspection in 2017; Knottnerus et al., 2017) it is argued that various external developments demand new organisational orientations and an organisational 'transition'. These developments include processes of globalisation, Europeanisation, decentralisation and horizontalisation which have led to increasingly multi-level and multi-actor governance configurations. Accompanied by the progressive complexity of contemporary environmental issues, due to the interwovenness of social-economic and ecological aspects at multiple levels and scales, these issues more and more demand policies that are directed at societal 'innovation and transformation'. The vision statement additionally identifies post-truth discourses and the deterioration of public authority as important developments not just for governments, but also for the credibility, and perceived independence and trustworthiness of knowledge institutes. To address these developments, in the statement it is argued that innovation towards a practice of reflexive research is elemental in order to retain credibility and authority as national knowledge institute, and the PBL's ambition for more transdisciplinary,

deliberative and interpretative modes of research geared towards knowledge co-production is made explicit.

While the ambition for a more reflexive research practice was pronounced relatively recent and gained momentum with the initiation of several large-scale policy evaluation projects in which knowledge co-production was the central approach, the organisation's strive to practise co-production knows a longer history which has been studied empirically in the past (including Hage et al., 2010; Huitema & Turnhout, 2009; Kunseler, 2017; Pesch et al., 2012; Petersen et al., 2011). These studies have in common that they display the organisation's search for dealing with the external developments discussed earlier, given its own positivist-technocratic orientation. This search has resulted in the increasing adoption of knowledge co-production approaches and stakeholder participation to reflexively attend to the uncertainty that underlies environmental knowledge, and to enhance its legitimacy and quality (which led to the development of the Guidance for Stakeholder Participation, Dutch: *Leidraad voor stakeholderparticipatie*; Hage & Leroy, 2007). Despite the ambitions for knowledge co-production, it was found that concerns regarding PBL's independent status and the risk of getting snared in political power-play – as well as little trust in stakeholders' intentions regarding integer use of preliminary findings – led to instrumental rather than substantial participation in which stakeholders were consulted rather than involved as active research partners (Hage et al., 2010). Other scholars brought to light diverse dilemmas raised by PBL researchers on knowledge co-production, such as how to adequately balance distance from and engagement with policy actors, and the question whether stakeholders advance or rather deteriorate the scientific quality of research (Kunseler & Tuinstra, 2017). What the various studies on the PBL share is that they illustrate the bounded agency of its policy researchers. While they may seek to tailor their research approach to the characteristics of the environmental policy issue at hand, they cannot escape the societal and political institutionalised expectations regarding the conduct and outcomes of policy research, and risk losing credibility and authority in the eyes of the public. As Kunseler (2017) also points out, it appears that the innovation towards a reflexive research practice cannot happen overnight.

Motivation for the PBL as paradigmatic case

My decision for the PBL as paradigmatic case is motivated by the organisation's explicit ambition to practise reflexive research, and its search for methodological innovation towards knowledge co-production and associated deliberative and interpretive policy research approaches. As this organisational 'innovation' has been well studied in the past, this allows me to build upon earlier scholarly work on the

organisation's dilemmas with practising co-production and interpret these from a practice-based perspective.

Furthermore, the principle aim of studying practice is obtaining a thick empirical understanding of the practice under study which allegedly may only be obtained by active participation (Schmidt, 2016). Choosing the PBL allowed me to adopt a role of researcher-practitioner and access insider knowledge as I intermittently worked at the PBL in various research projects over the course of my studies. First, as parttime posted worker in 2017-2018, and from 2019 onwards as employee, which I combined with my work at the university (0.8 fulltime equivalent (FTE)/0.2 FTE, resp.). At the PBL, I was involved with a number of projects in which a co-production approach was adopted. In addition, I was tasked with supporting the professionalising and embedding of reflexive research within the organisation. This provided me with a 'view from the trenches', and allowed me to experience the practical concerns of researchers with enacting co-production first hand. As novel PBL employee I was submitted to a social process of learning about the organisations' normative orientations and its policy researchers' varying perspectives on what counts as legitimate and appropriate policy research. As Nicolini (2012) argues, as novices undergo a process of socialisation during which they are explained and shown how to conduct a practice, following novices is a particular useful exercise to study practice and its underlying logic. During the work this thesis describes and discusses, I was such a novice myself. Selecting the PBL as primary case allowed me to gain access to doings and sayings that might otherwise have remained invisible or obscure.

3.3 Research questions and methods

The main research question is addressed through four research sub-questions that concern different aspects of the normalisation of knowledge co-production, i.e., the challenges policy researchers experience with practising and normalising knowledge co-production (sub-question 1) and the activities they undertake to address these (sub-question 2), the role of contextual developments on the normalisation of co-production (sub-question 3) and the potential of reflexive monitoring to support the normalisation of co-production (sub-question 4). To answer the sub-questions, a total of four studies were conducted over the course of 2015-2021. I will discuss each of the studies and the respective research methods in this section. I attend to the four research sub-questions in the five chapters of this thesis (see **Table 3.1**).

The first two research sub-questions are:

> *Sub-question 1: What challenges do policy researchers experience with the implementation and normalisation of knowledge co-production and how may these be understood?*
>
> *Sub-question 2: What activities do policy researchers undertake to address these challenges, and what do these imply for the normalisation of knowledge co-production?*

To address these questions, two studies were conducted. The **first study** concerns a single case study of the Natuurpact (NP) research project's first cycle (2014-2017) (**Box 3.1**), a project during which a reflexive evaluation approach was adopted. The first study aimed to bring to light the challenges the project team experienced with knowledge co-production in policy evaluation and the activities they undertook to address these (or not). The PBL had commissioned researchers from the Athena Institute (VU University Amsterdam) – of whom I was one – to support their undertaking, because of the Athena Institute's expertise in deliberative research, reflexive monitoring and knowledge co-production. We took on roles of reflexive monitors to help the Natuurpact project team develop capacity for co-production and guide them in the process. Our roles were based on Reflexive Monitoring in Action, an interactive and action-oriented methodology for monitoring complex projects that aim to contribute to sustainable system innovation in the context of wicked problems (Van Mierlo et al., 2010). As such, we took part in project meetings, provided input for the research's design, helped design and facilitate multi-stakeholder workshops and reflected upon the project team's decisions to support them with developing co-production's key features in practice. As reflexive monitor, I also made participant observations which I recorded in field notes. In addition, 17 in-depth semi-structured interviews were held with members of the project team at the start, after the first year and after the first cycle (2014-2017) had concluded. The study brought to light different understandings within the team and within the PBL organisation of what constitutes 'good' policy research quality and impact, which had implications for how knowledge co-production was operationalised (published in a research report, Verwoerd et al., 2019).

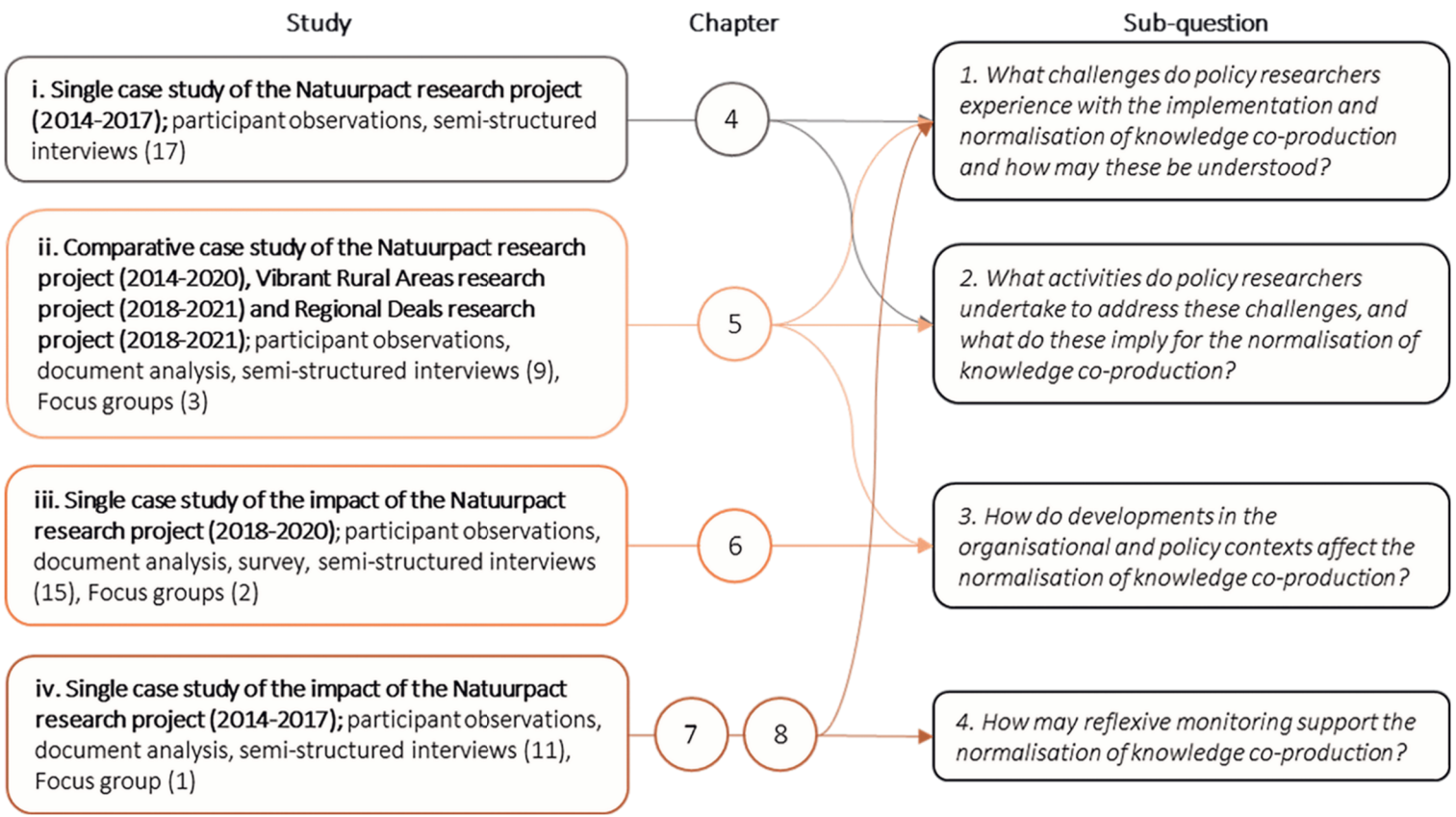

Figure 3.1 Overview of the studies (incl. methods), corresponding chapters and research sub-questions.

Box 3.1 The Natuurpact reflexive evaluation

In 2014, PBL was commissioned to conduct a longitudinal national policy evaluation of the Natuurpact agreement (2013). The pact finalised the decentralisation of nature policy to the 12 Dutch provinces and was signed by the Ministry of Agriculture, Nature and Food Quality (Dutch: Landbouw, Natuur en Voedselkwaliteit; LNV) and provincial governments (represented by the Interprovincial Council (Dutch: Interprovinciaal Overleg; IPO), and developed in close interaction with several societal organisations (including the Forestry Commission (Dutch: Staatsbosbeheer) and the Dutch Society for Nature Conservation (Dutch: Natuurmonumenten)). In the pact there was agreed upon nature policy ambitions which include halting biodiversity decline, increasing societal engagement with nature and strengthening the relationship between nature and economy. These ambitions were agreed to be attained by 2027. It was also recorded that, in parallel to the programme, PBL will conduct a policy evaluation directed at mutual learning to increase the impact of nature policy, executed in consecutive three-yearly cycles. So far, the first (2014-2017) and second (2018-2020) cycle have been concluded (PBL & WUR, 2017, 2020). Given nature policy's decentralised and increasingly networked character and the social-economic and ecological aspects of its ambitions, PBL and partner organisation Wageningen University & Research (WUR), together with national and provincial governments, decided upon a reflexive, participatory evaluation approach aimed at knowledge co-production. Under the header of 'reflexive evaluation' (Dutch: lerende evaluatie) it was argued that this approach would enhance the quality, usability and impact of the evaluation findings. Although hardly PBL's first endeavour regarding reflexive research (Section 3.3), the Natuurpact research project was the first policy evaluation of its scale, duration and profile that fully adopted a knowledge co-production approach. It provided a welcome opportunity for the organisation to further its reflexive aspirations.

The Natuurpact research project has provided the empirical setting in which much of the research in this thesis took place and has allowed me study first-hand how knowledge co-production was practised and normalised in context of an individual longitudinal research project.

In 2018 (the Natuurpact's second cycle had already started), PBL was commissioned for two more policy evaluations of large-scale national policy programmes, for which a knowledge co-production approach was deemed called for: the Vibrant Rural Areas (VRA) reflexive evaluation and the Regional Deals (RD) reflexive research programme (**Box 3.2** and **3.3**, respectively). This provided an opportunity for a comparative case study approach which could further our findings from the first study and test whether these findings were typical to the Natuurpact research project's first cycle, or applicable to other co-production projects too. For this **second study**, data was

collected by our involvement in the three projects (and by my general position at the PBL) as participant observers. In addition to our field notes, we conducted a document analysis of the projects' proposals and deliverables and nine in-depth semi-structured interviews were held with project leaders, members of the project teams and one supervisor, to improve insight into the challenges and barriers the teams experienced with practising co-production and how these were addressed (or not). Furthermore, focus group discussions were held with each project team to deepen our understanding on how decisions for how knowledge co-production was practised were made. Finally, a reflection session was hosted with all project teams combined, including other interested PBL colleagues, to further discuss the state of PBL's knowledge co-production practice and its barriers to further professionalisation and normalisation within the institute. As for the first study, the second gave answer to the first two research questions, albeit more in-depth. Specifically, the second study gave answer to the sub-research question 2's second part: the implications of challenges and actions for normalisation of knowledge co-production within the organisation. The findings of the second study were published in Verwoerd et al. (2021).

Box 3.2 The Inter-Administrative Programme Vibrant Rural Areas reflexive evaluation

The Inter-Administrative Programme Vibrant Rural Areas (2018-2021) (Dutch: Interbestuurlijk Programma Vitaal Platteland, IBP VP), aimed at advancing inter-administrative collaboration and instigating a transition towards a more 'vibrant' Dutch countryside. Its initiators – LNV, IPO, the Association of Dutch municipalities (Dutch: Vereniging der Nederlandse Gemeenten; VNG) and the Union of Water Boards (Dutch: Unie van Waterschappen, UvW) – had commissioned the PBL to conduct a research project directed at policy learning in support of the programme's aims of advancing inter-administrative collaboration and agricultural transition. With the Athena Institute once again as partner, the respective project team adopted a co-production approach directed at transformative learning. During this project, I was less directly involved due to my activities in the Natuurpact research project and another national policy programme PBL was commissioned to study: the Regional Deals programme.

Chapter 4 and **5** (and parts of **Chapter 8**) reflect on the first two research sub-questions. **Chapter 4** gives an account of how the NP project team operationalised features of knowledge co-production and seeks to explain the decisions that were made by showing how the team navigated two differing institutional logics – the modernist and reflexive – on policy research. **Chapter 5** continues with these observations and has particular attention for normative and relational structures that

were encountered by the NP, VRA and RD project teams and how they sought to align to these structures to develop space for their co-production.

Box 3.3 The Regional Deals for advancing regional wellbeing reflexive research programme

The Regional Deals policy programme for advancing regional wellbeing (2018-2021) had the ambition of enhancing the wellbeing in Dutch regions (Dutch: regionale brede welvaart) by closing so-called 'Regional Deals'. These Deals were focused on forging partnerships between national government and regions, represented by regional actors that include local governments, societal organisations, industry and businesses. LNV had commissioned the PBL to design and conduct a research project to run in parallel to the Regional Deals' execution and to inform the programme with relevant knowledge on policy for regional wellbeing. Once more, a co-production approach was adopted by the respective project team to inform the learning processes of involved actors with both theoretical and practical knowledge. Via my position at the PBL, I was involved in the Regional Deals research programme to support the co-production approach, design and facilitate multi-stakeholder workshops and to study the policy learning capacity of the actors that constituted the networked governance constellation manifest in the Regional Deals programme (my colleagues of the Athena Institute were not otherwise involved with this case).

The third research sub-question pertains to contextual developments and how these affect the normalisation of knowledge co-production:

> *Sub-question 3: How do developments in the organisational and policy contexts affect the normalisation of knowledge co-production?*

As part of our role as reflexive monitors, the Athena Institute was commissioned by the PBL to conduct an impact assessment of the NP research project's impact, with particular attention for its co-production approach, both after its first and second cycle. The **third study** is based on the impact assessment of its second cycle (I turn to our assessment of the first cycle shortly). We focused on policy researchers' and actors' views on what should constitute impact of policy evaluation and how they valued a knowledge co-production approach in light of this. Next to the participant observations that were made over the course of the project, we conducted a document analysis that included the project proposal and final report, minutes of meetings between the project team and formal commissioners, interview transcripts, and parliamentary papers that regarded the NP research project. A survey was conducted amongst the policy actors who had participated with the project to gain insight in their experiences and perspectives on the project's impact and quality. In addition, 15 in-depth semi-structured interviews were conducted with participants

ranging from national and provincial policy actors to societal organisations, and from more executive to strategic managerial levels. The interviews were followed by two focus groups during which national and provincial policy actors, societal organisations and the research project leaders participated, and which were directed at deepening our understanding and developing shared recommendations for how to improve the NP research's co-production approach and policy impact. The conclusions of this review were published in Brouwers et al. (2021).

Chapter 6 and parts of **Chapter 5** attend to this research question. **Chapter 6** reflects on the effects of policy and societal actors' views on the impact of the Natuurpact research project and their valuation of its reflexive approach for the design and implementation of the research project. It discusses how extrinsic political developments affect policy and societal actors knowledge demands and what this implies for the normalisation of knowledge co-production. **Chapter 5** discusses (un)conducive organisational developments and their importance for normalising knowledge co-production.

The fourth research sub-question concerns reflections on our own role as reflexive monitors during the Natuurpact research project. The fourth sub-question is as follows:

> *Sub-question 4: How may reflexive monitoring support the normalisation of knowledge co-production?*

The **fourth study** includes a theoretical exploration of the potential of reflexive monitoring for supporting the implementation and normalisation of knowledge co-production in research projects, and an empirical exploration in which we put these ideas to practice. It is based on our experiences as reflexive monitors during the first cycle of the NP research project and our impact assessment of this cycle. In addition to our field notes, we conducted in-depth semi-structured interviews with 11 national and provincial policy actors and societal organisations that had participated, followed by a focus group discussion with a formal working group comprised of 12 provincial nature policy actors who had actively collaborated with the project team over the course of the NP project. The interviews and the focus group were directed at obtaining their view on the impact of the project and on the features of the co-production approach attributable to this. Also, transcripts of multi-stakeholder workshops and interviews that were held in light of the NP research project were analysed, as were formal project documents, including the project's proposal and final report and (provincial) policy documents. This led to a review report on the perspectives of the policy actors who participated with the co-production process on

its impact and quality, which included recommendations for the improvement of the co-production approach for the project's second cycle (Verwoerd et al., 2017).

Chapters 7 and **8** both contribute to answering the fourth sub-question. **Chapter 7** does so from a more conceptual-theoretical perspective. In this chapter we discuss the potential of reflexive monitoring to simultaneously promote and asses knowledge co-production projects. In **Chapter 8** we put our hypotheses to the test. It presents a similar discussion, but is more explicitly grounded in the empirical material that was collected during the Natuurpact research project.

3.4 Research validity

Being embedded as a researcher-practitioner in the practice reflected upon in this thesis enabled me to experience the challenges and dilemmas researchers face when attempting to normalise reflexive research. This allowed for a deep and acute understanding of this practice – arguably more so than could have been reached when observing the practice from an outsider's perspective. Simultaneously, such close proximity to the object of research also presents potential risks for the validity of the research, especially regarding researcher bias. Keeping a wide breadth of interpretive possibilities may be difficult when an organisation, its researchers and research projects have become so familiar. To remedy this and to maximise research validity, various strategies were adopted.

To start, triangulation of research methods was applied to buttress validity and to help ensure that biases that may arise from use of a single method, source or researcher are overcome (Carter et al., 2014; Noble & Heale, 2019). The central notion of triangulation is that methods, sources and researchers that lead to the same results enhance the validity of the findings. The research methods that were employed included combinations of participant observations, document analyses, surveys, interviews and focus group discussions. Combining various methods allowed me to benefit from their individual strengths, while guarding against potential bias caused by their individual weaknesses. In addition, researcher triangulation was applied; each study was conducted in a team of three researchers or more, all of whom were actively engaged with the empirical context of the study and who collaborated on the study's design, execution, and reporting, while not all always were similarly embedded in the practices at hand.

In addition to triangulation, a self-reflective stance was pursued by various activities to enhance reflexivity. To start, the teams involved with the various studies were expanded with researchers who had not been previously involved with the study nor its empirical context, and who functioned as so-called 'critical friends' (Schuijer, 2020;

Van der Meij, 2017). These critical friends were tasked with asking 'naïve' yet critical questions regarding the study's design, conduct, analysis and conclusions, and bringing in fresh theoretical and/or societal perspectives. Furthermore, as a team of researchers, we frequently and deliberatively reflected upon our own standpoints and normative positions regarding the studies' outcomes, and how these could potentially affect the ways in which we interpreted the findings. We challenged each other's interpretations, putting assumptions to the test and formulating alternative hypotheses. For studies one and four (which concerned assessments of the policy impact of the Natuurpact research project) reflexivity was further enhanced by appointing external review committees that comprised scholars in fields of reflexive governance and reflexive policy analysis who critically reviewed the process and outcomes of our studies.

Finally, we sought to guard internal validity through member checks (Birt et al., 2016; Lincoln & Guba, 1985), formally with all respondents by asking their feedback on our interpretations of the interviews, focus group discussions or workshops, but also in less formal settings for instance with PBL researchers by discussing whether our interpretation of occurrences corresponded to their views. While we held prerogative over the studies' findings and conclusions, their feedback was taken seriously and provided valuable information to us as regards their experiences and perspectives on co-production and its normalisation within the organisation.

3.5 Outline of this thesis

This thesis continues as follows: in **Chapters 4** to **8** I present the findings of the studies I described above. **Chapters 4** and **5** focus on the emerging practice of knowledge co-production in policy evaluation at the PBL. **Chapter 4** discusses the challenges and dilemmas with practising knowledge co-production in policy evaluation as experienced by the NP research project team during its first cycle (2014-2017), paying particular attention to how institutional logics play a role in guiding research practice and what this implies for normalisation. **Chapter 5** continues this discussion in light of a multiple case study (which includes the NP project's second cycle (2018-2020), and the VRA and RD projects) and explores more in-depth the type of activities that policy researchers undertake to negotiate space for their aspired co-production approach, as well as the conducive developments for normalisation in the projects' organisational context. **Chapters 6** delves into how the impact of knowledge co-production in evaluation is understood and appreciated in the eyes of involved policy researchers, and policy and societal actors, and how this affects their view on the legitimacy of knowledge co-production. **Chapter 7** presents a theoretical exploration for the potential of Reflexive Monitoring in Action to promote and assess

knowledge co-production. This discussion is furthered in **Chapter 8** in which this potential is empirically explored by reflecting upon the merit of our role as reflexive monitors during the NP research project's first cycle.

4 | NAVIGATING LOGICS ON POLICY EVALUATION

Abstract

The first empirical chapter discusses this thesis' study of the Natuurpact research project's starter cycle (2014-2017). Positioned in the discourse on reflexive evaluation aimed at societal change and sustainable development, this chapter reflects upon the first two research sub-questions regarding the challenges policy researchers experience with knowledge co-production in evaluation – i.e., reflexive evaluation – and the actions they undertake to address these.

To advance understanding of the challenges policy researchers encounter when seeking to practise reflexive evaluation, the chapter takes as point of departure the prevalence of two ideal-typical institutional logics that feature in policy research – the modernist and the reflexive logic – that guide policy practices in many countries in the Global North. While hybrid evaluation practices are increasingly common, modernist evaluation logics that focus on performance management continue to be privileged, hampering the normalisation of (more) reflexive logics revolving around system change. Normalisation Process Theory (NPT) is used to analyse the activities the research project team of the PBL Natuurpact reflexive evaluation's first cycle (2014-2017) undertook to accomplish alignment between the prevailing modernist and the proposed reflexive logics. Ad hoc alignment strategies and insufficient investment in mutual sensemaking regarding reflexive evaluation are found to hinder normalisation. It is argued that alignment requires that legitimacy for reflexive evaluation is developed in the context of application, while guarding the integrity of its (theoretical) ideals. The chapter concludes by reflecting on the use of NPT for studying the normalisation of reflexive evaluation.

The findings in this chapter provided important input for the study presented in the following chapter, in which concepts such as legitimacy and integrity, and the actions policy researchers undertake to navigate these concepts, are explored empirically and more in-depth in a comparative case study.

Published as:
Verwoerd, L., Klaassen, P. & Regeer, B.J. (2021). *How to normalise reflexive evaluation? Navigating between legitimacy and integrity.* Evaluation. 27(2), 229-250.

4.1 Introduction

Contemporary policy processes increasingly occur in complex, multi-actor and multi-level governance contexts. This has led to a proliferation of views on the purposes and roles of policy evaluation and suiting approaches and methodologies. Evaluation literature often distinguishes schools in policy evaluation along the lines of 'modern' versus 'post-modern' science (Funtowicz & Ravetz, 1993), 'technical' versus 'deliberative' models (Owens et al., 2004), 'technocratic' versus 'participatory' approaches (Chouinard, 2013) or 'modernist' versus 'reflexive' logics (Kunseler & Vasileiadou, 2016). We endorse the rejection of such distinctions as being overly rigid, siding with scholars who point out a trend in evaluation practices towards a tailored choice of functions, methodologies and tools in which elements of allegedly opposed schools are combined (see, for example, Green et al., 2015; Van Hemelrijck & Guijt, 2016). However, evaluation practices can hardly be expected to be free from the influence of institutional, political and societal conventions and preferences, and we observe that there are often limits to the extent to which evaluations can be tailored to the policy issue at hand. This is attested by the factual existence of hybrid evaluation practices that manifest a continued privileging of so-called modernist evaluation approaches in many Western countries (Chouinard, 2013; Fitzpatrick et al., 2008) – i.e., evaluations largely based on assumptions about the linearity of policy processes, emphasising objectivity, accountability and performance management (Nieminen & Hyytinen, 2015) at the cost of, e.g., inclusivity, usefulness or learning.

As the (older) modernist knowledge tradition has provided the technocratic script for the science–policy interface for decades, norms and conventions describing the appropriate function and form of evaluations often derive from modernist foundations. Policy researchers are hence faced with a dilemma: in seeking ways to adequately inform policy processes and to address complex societal problems, they are drawn towards more systemic and reflexive evaluation approaches, while simultaneously not being able to entirely elude historically entrenched organisational, political and societal expectations of modernist approaches to policy evaluation.

This paper explores an attempt by evaluators from the Netherlands Environmental Assessment Agency (PBL), with a longstanding modernist tradition of policy evaluation, to normalise a more reflexive evaluation practice. To do so, we adopt a practice-perspective and combine the concept of institutional logics (Berg Johansen, 2017; Friedland & Alford, 1991) with Normalisation Process Theory (NPT), a theory developed for studying the normalisation of innovations of organisational practice (May & Finch, 2009). Our case comprises a series of steps in the process of normalising a novel approach to evaluation which was manifest both in the

organisational context of PBL, and in context of the nature policy program that was evaluated. We specifically focus on what normalisation of a reflexive evaluation practice entails when undertaken in contexts more readily amenable to so-called modernist evaluation logics. What we find is that this process of normalisation is best described as a trajectory in which evaluators continuously negotiate and navigate between two different logics, modernist and reflexive.

4.2 Background

'Logics' of evaluation and their implications

Contemporary societal issues, including climate change, global poverty, and loss of natural resources, are increasingly understood as complex or wicked (Rittel & Webber, 1973) societal problems. These require reconsideration of (in)formal rules, dominant ways of thinking and doing, problem solving and resource management, as these are in many ways part of the problem (Beck et al., 1994). Scholars have made a case for adopting a systems approach in the governance of these problems (e.g., Geels, 2004) and it is argued that such approaches should include a 'reflexive perspective' as well. This means that things that are usually taken for granted are scrutinised in ways that challenge their historically grown self-evidence, thereby creating possibilities for system change (Loeber et al., 2007:84). To support the design and analysis of policy or interventions for system change towards sustainable development, evaluation approaches have emerged that also include such a reflexive perspective (e.g., Van Mierlo et al., 2010).

Reflexive evaluation diverges from modernist evaluations, which in their essence present an instrumental tool for warranting accountability and compliance. Following a linear inputs-outcome-outputs-impact framework, such approaches tend to overlook complexity (Nieminen & Hyytinen, 2015) and fall short in drawing attention to systemic properties that delimit the issue at hand (Arkesteijn et al., 2015). Modernist evaluation logics seem to dominate policy and program evaluation practices of many Western countries (Chouinard, 2013; Fitzpatrick et al., 2008; Nieminen & Hyytinen, 2015), despite rising calls in academic literature for more complexity-oriented and reflexive evaluation in research and attempts to do so in practice (Patton, 2010; Van Mierlo et al., 2010).

To gain understanding in the limited uptake of reflexive evaluation we draw inspiration from social practice theorists, including Giddens (1984), Schatzki (2002) and Shove (2010). Rather than individual people and institutions and their opinions or intentions, or the surrounding social structures, we take practice itself as the basis focus of inquiry: the observable, collective and organised behaviours and actions that

people purposively and routinely perform and consider to be 'normal' ways of doing (Nicolini, 2012; Reckwitz, 2002). In this view, the limited uptake of reflexive evaluation approaches is not seen as the result of individuals' intentions or beliefs being hampered by contextual barriers, but as a manifestation of institutionalised social practices (Warde, 2005). Scholars have argued that different evaluation practices can be viewed as the material embodiment of different 'institutional logics' (Dahler-Larsen & Schwandt, 2012; Kunseler & Vasileiadou, 2016), which can broadly be understood as sets "of material practices and symbolic constructions [that] constitute organising principles" (Friedland & Alford, 1991:248) within particular institutions. Such logics operate as behavioural guidance, as they supply actors with the 'rules of the game' (Jones et al., 2015). These logics are not static but are practiced and shaped at nested, interacting levels: amongst individuals and teams within organisations, at the organisational level and within their wider societal context. Scholars have demonstrated how – sometimes contradictory – institutional logics may be at play and create friction, giving rise to plurality and potential space for institutional change, and the emergence or development of novel practices (Berg Johansen, 2017).

Scholars have rejected an empirically discernible dichotomy between modernist and reflexive evaluation approaches (or, e.g., 'modern' versus 'post-modern' science; Funtowicz & Ravetz, 1993) and have pointed out that such schools are rarely practiced in their pure form (Owens et al., 2004; Vaidya & Mayer, 2014). Indeed, theoretical understandings of evaluation are often equally hybrid, such as Beck et al.'s (1994) ideal of reflexive modernisation which predominantly veers towards the reflexive side of the spectrum but which also contains modernist elements. While we endorse this view, for the purpose of this study we distinguish between a modernist and a reflexive institutional logic of evaluation, each observable in the enactment of distinctive material practices and the discursive and non-discursive traces of complexes of beliefs, norms and rules about what evaluation is, 'should do', and how this is best achieved (Kunseler & Vasileiadou, 2016). We propose to understand these logics as useful, though overstated, generalisations, much like Weberian 'ideal-types' (Shils & Finch, 1949), to conceptualise the logical space within which evaluation practices can exist and from which practitioners may draw. A clear-cut conceptual distinction between modernist and reflexive evaluation is used as an epistemic tool to gain analytical depth and apprehend the influence of these implicit logics on policy evaluation practice. In their hypothetic and ideal-typical form, the institutional logics on evaluation are rooted in fundamentally different epistemologies and political theories. We argue that analysing the significance of these underlying differences helps to understand the challenges practitioners experience when attempting to practically reconcile them.

The modernist logic can be traced back to eighteenth-century Enlightenment ideals and nineteenth century ideas of technocracy and rationality, as well as even older ideas on scientific objectivity, value-free science and impartiality (Kunseler & Vasileiadou, 2016; Lentsch & Weingart, 2011). Built around epistemologies that present reality as objectively knowable (positivism, empiricism) and as substitute for religion and tradition, scientific knowledge was argued to be the best foundation for public decision-making. The highest quality of scientific facts could be obtained through an independent and closed-off scientific apparatus, free from social values. Modernist logic treats science and policy as strictly separated domains and the resulting science was believed to linearly advance progress and public welfare (Dahler-Larsen, 2012; Jasanoff, 2011). This modernist outlook on scientific knowledge and its proper relations with policy was further institutionalised with the rise of new public management in the 1980s as Western governments increasingly called for accountability and performance measures to enhance the performance of the public sector (Chouinard, 2013; Pollit et al., 2007). In this context, evaluation primarily focuses on regulation and compliance, while serving as a management instrument to assure accountability. Ideally, evaluators provide empirically confirmed and logically consistent statements to enable evidence-based decision-making, keep the domains of science and policy separated, and prevent the infringement of scientific fact with values (Sternlieb et al., 2013). Golden standards for evaluators based on a modernist logic comprise objectivity, scientific autonomy, impartiality and detachment from societal values and politics. Its preferred methods include technical-analytical processes that follow logic-model thinking, such as impact assessments, cost-effectiveness analyses and modelling (Kunseler & Vasileiadou, 2016; Verwoerd et al., 2019) (**Table 4.1**). Second, the paper considers a reflexive logic of evaluation, which gained increasing traction over the past 20 to 30 years. Linearly advancing public welfare had been demonstrated to have unintended (negative) side-effects, thereby disclosing the complex and uncertain interdependencies of ecological, social, economic, political and institutional processes that modernist approaches are arguably poorly equipped to understand (Duijnhoven & Neef, 2016). Drawing on social-constructivist epistemologies (pluralism, relativism, pragmatism), reflexive logic builds on the premise that scientific knowledge is not produced in isolation, but is deeply intertwined with cultural understandings of socio-economic and -ecological relations (Jasanoff, 2011). Science and policy are considered inevitably entangled domains. Ideally, their relations are open, transparent and deliberative, enabling social learning and the co-production of socially robust answers to complex problems. Evaluation approaches explicitly geared towards engaging with these complex problems take a system perspective (Moore et al., 2019; Patton, 2010) and aim to support policies or interventions by stimulating their reflexivity on their relationship to these systems

(Arkesteijn et al., 2015) and enhancing their ability to challenge various systemic (societal, institutional, political) contexts to allow for system change (Beers & Van Mierlo, 2017). Given the limitations to their ability to produce ultimate 'truths' to dictate public decision-making, evaluators working from the reflexive logic take on an attitude of humility and organise inclusive and productive interactions on the science-policy (and society) interface to facilitate interactive learning and learning-by-doing. Norms and principles include intersubjectivity (moving from 'one objective knowable truth' to 'understanding the world together'; De Jaegher et al., 2017), acknowledgement of perspective plurality and uncertainty, and transparency. Reflexive logic is partial to deliberative-analytical methods that are context-sensitive (Rog et al., 2012), real-time (Marjanovic et al., 2017) and include joint fact finding, knowledge co-creation, system analyses and reflexive monitoring.

Myriad evaluation practices have been derived from these logics for evaluators to suit the policy issue at hand. However, when practices rooted in different institutional logics are combined, something that appears self-evident within one logic may be highly problematic from within the other. For instance, where interaction between science and policy during evaluation may be considered inherent and crucial according to reflexive logic, modernist logic would argue this to be detrimental to the scientific quality of the research because its objectivity and disinterestedness might be compromised (Turnhout et al., 2013). Furthermore, the modernist logic, being the historically more established of the two, predominately guides evaluation practice in many Western countries, giving rise to a culture of accountability in their policy and program contexts (Fitzpatrick et al. 2008). Consequently, it is hard for both evaluators and policymakers to move away from this dominant logic towards more reflexive approaches, as this requires facing unaccommodating beliefs, norms and rules.

Table 4.1 Modernist and reflexive evaluation logic (adapted from Kunseler, 2017; Verwoerd et al., 2019).

	Modernist logic *"Speaking truth to power by bridging the gap"*	Reflexive logic *"Enhancing reflexivity for system change by organising productive interactions and interactive learning"*
Epistemological foundations	Positivism, empiricism, scientism	Social constructivism, pluralism, pragmatism
Perspective on science-policy interface	Technocratic and bureaucratic relations between science and policy; science and policy as strictly separated domains	Open, transparent, and deliberative relations between science and policy; science and policy as ultimately entangled
Purpose of evaluation	Evaluation is an instrument for accountability assurance, performance assessment, and compliance	Evaluation is a mechanism to enhance the reflexivity of policy or interventions for system change to deal with complex societal problems
Role of evaluators	Evaluators provide empirically confirmed and logically consistent statements to inform evidence-based decision-making. They mediate domains of science and policy and work to keep them apart	Evaluators facilitate processes of interactive learning, learning-by-doing and reflection on systemic properties that hamper or facilitate system change. They organise productive interactions on the science-policy interface to develop socially robust answers to complex societal problems
Norms, principles and disposition	Objectivity; neutrality; impartiality and detachment from societal values and politics	Inter-subjectivity; acknowledgement of perspective plurality and uncertainty; transparency; humility
Methodology and tools	Technical-analytical processes that follow logic-model thinking such as impact assessments, cost-effectiveness analysis, modeling	Deliberative-analytical processes that include a system perspective such as joint fact finding, knowledge co-creation, system analysis, reflexive monitoring

Thus, although the two logics are rarely practiced in their pure form, that does not mean that it is straightforwardly clear *how* to combine them well, nor how those beliefs, norms and rules of the modernist logic that are particularly unaccommodating towards reflexive evaluation can be engaged with in such a way that reflexive forms of evaluation become more normalised. It is our understanding that when a practice 'out there' is newly introduced into an organisation, a certain amount of modification is required for it to align to its practitioners and their context (Fullan & Pomfret, 1977).

When the practice finds uptake into the formal and informal structures of an organisation in a way that the practice's original integrity is maintained and viewed as legitimate, this is considered successful normalisation (May & Finch, 2009). This paper empirically investigates the process through which practitioners conduct a large-scale evaluation of a nature policy program, drawing on reflexive logic in an organisational and policy context that are partial to the modernist logic. We thereby aim to make recommendations on how normalisation of reflexive evaluation practices can be encouraged.

Institutional evaluation logic at the PBL: the Natuurpact reflexive evaluation

The empirical material on which we draw comprises the evaluation of a Dutch nature policy program called the Natuurpact, conducted by the PBL. The PBL is a public knowledge institute charged with independent, scientific policy assessments 'in the fields of the environment, nature and spatial planning' (PBL, 2019). It has established itself within a technocratic, modernist paradigm and has an authoritative status on the science–policy interface (Halffman, 2009). Given its own modernist orientation, PBL actively strives to practice a more reflexive logic in giving scientific advice, by attempting to innovate its research repertoire with more deliberative and interpretative modes of research. The co-existence of different logics at the PBL and its endeavours towards a more reflexive practice are well studied (Kunseler & Vasileiadou, 2016; Kunseler & Tuinstra, 2017; Petersen et al., 2011). The Natuurpact evaluation is the organisation's first large-scale longitudinal *evaluation* for which the entire approach was designed following reflexive principles.

In 2013, the Dutch national government, governments of the 12 provinces, and various societal organisations signed the Natuurpact. Through this pact, nature policy became the responsibility of provincial governments, and its signatories formulated and agreed upon high shared ambitions for nature policy. It was also recorded that PBL would conduct an 'ex durante' evaluation to allow for timely policy adjustments. As we discuss below, the program presented a welcome opportunity for PBL to further its reflexive aspirations.

The Natuurpact evaluation will run until 2027 and comprises three-yearly evaluation cycles. The first cycle ran from 2014 to 2017; the second (2018–2020) has just concluded. The current authors were involved as external academic experts for our knowledge and skills in participatory research, reflexive monitoring and evaluation research. We also performed the role of reviewers of the impact of the Natuurpact reflexive evaluation after the first cycle had ended. We discuss our role in more detail below.

Normalisation Process Theory

To study how institutional logics shape evaluation practice, we draw inspiration from the field of implementation science and view the implementation of reflexive evaluation – and with it, an underlying reflexive logic – as an innovation of PBL's organisational practice. Already in the early seventies, scholars in education innovation have pointed out the importance of conceiving implementation as a process that occurs within an institutional and wider context to understand why some innovations succeed to become standard practice, and why others don't (Fullan & Pomfret, 1977; Havelock, 1971; 1970; Huberman & Miles, 1984). May and Finch (2009) develop this further and with Normalisation Process Theory consider the mechanisms that facilitate or hamper the normalisation of an innovation into its 'host context'. NPT focuses on the work actors do for implementation to gain an understanding of the complex dynamics of implementing and institutionalising new technologies or innovations in organisational practice (May et al., 2018; McEvoy et al., 2014). The theory includes a practice perspective: rather than individual people or institutions and their opinions or intentions, the basic focus of inquiry is the observable, collective and organised behaviours and actions that people purposively perform (Nicolini, 2012). As such, NPT arguably allows for an analysis of how a new practice becomes normalised into its social context as those involved make sense of it, buy into it, agree on how it's done and appraise how it has value, and potentially provides a useful theoretical lens to understand how context influences normalisation.

Central to NPT are four core mechanisms that comprise the work implementers of an innovation do to normalise it: *sensemaking, engagement, enactment*[5] and *appraisal work* [6]. Furthermore, May et al. (2016) underline the importance of context for implementation processes and argue how these are the emergent outcomes of interactions and negotiations between components of the innovation and elements of the host context. Actors who aspire a novel practice are thus required to negotiate a level of 'fit' between the prevailing standards and ways of working, and the aspired

[5] In earlier versions of the NPT framework enactment was the core focus, as it comprises the acts and behaviours of performing a practice, and therefore the practice itself (in alignment with other social practice theorists who consider enactment and practice two sides of the same coin, see e.g. Giddens (1984) or Reckwitz (2002)). For sake of clarity, in this paper we use the term 'enactment' only in relation to the third NPT mechanism, which – in concordance with sense-making, engagement and appraisal – may lead to the *normalisation* of innovative practices.

[6] While the inventors of NPT expand each mechanism with multiple sub-concepts, for the purpose of our study we have chosen to focus our analysis around the four core mechanisms of NPT. We refer readers to May & Finch (2009) for a seminal account.

ideals and procedures, in order for the latter to have viability and workability. In other words, the work belonging to the four core mechanisms largely takes the shape of what we call *alignment work* (**Figure 4.1**).

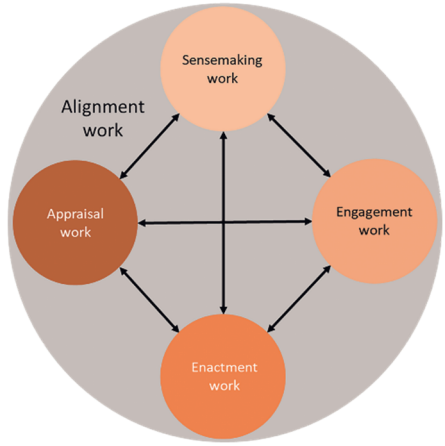

Figure 4.1 The four NPT mechanisms with "alignment work" as underlying determinant for how the work of each mechanism is conducted.

We hypothesise that NPT functions as an appropriate lens through which to study how institutional logics at play in the contexts in which the Natuurpact evaluation was conducted, shape evaluation practice. Below we elaborate the NPT mechanisms in more detail and show how each applies to the case in question, and how the need for alignment shaped the work that was done by the research team. First we turn to the research methodology.

4.3 Methodology

Case study

Our case study concerns the Natuurpact evaluation's first cycle, executed by researchers from PBL and partner Wageningen Environmental Research (WER), supported by the current authors (Athena Institute). The interdisciplinary team consisted of six researchers (including two project leaders). Few had prior experience with deliberative or reflexive evaluation approaches. The team met bi-weekly to discuss the evaluation's progress and research activities.

The evaluation's primary participants consisted of provincial and national policy actors responsible for the development and implementation of nature policy. The main form of interaction with the participants consisted of bi-monthly meetings with a selected working group of 12 representatives from provincial governments, and eight

multi-stakeholder workshops that occurred over the course of the evaluation. The purpose of these workshops differed according to the evaluation phase, and included making an inventory of the evaluation needs of public actors, joint interpretation of preliminary evaluation findings and drawing joint conclusions for action perspectives for change. The team also met with the evaluation's commissioners (administrators from national and provincial government) twice a year to ensure the evaluation was still on track in terms of timing and budget. **Figure 4.2** presents a schematic overview of the evaluation's first cycle.

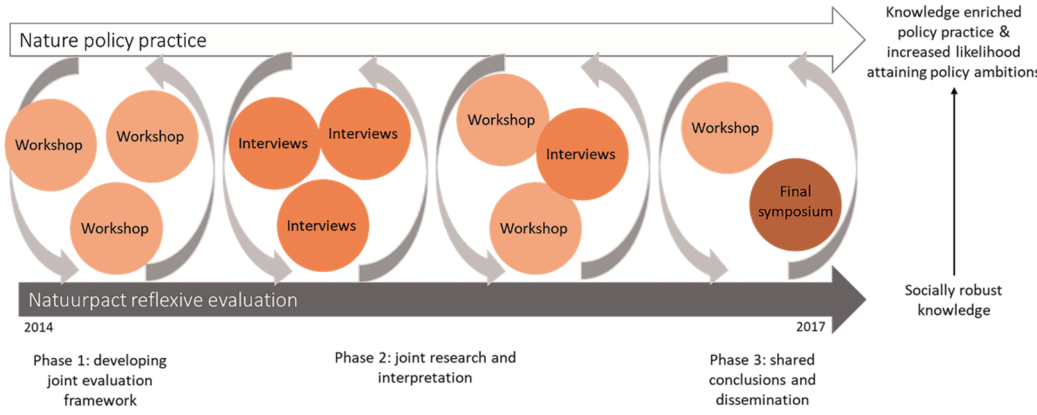

Figure 4.2 Schematic representation of the first cycle of the Natuurpact reflexive evaluation, including its three main phases and types of interactions between participants and researchers.

Material and methods

Participatory action research

Authors A1 and A3 were (variably) members of the project team as participatory action researchers: they supported the design and execution of the reflexive evaluation, and simultaneously studied this process. They were tasked with developing a theoretical framework for reflexive evaluation based on academic literature including process principles and ideal outcomes. During project meetings the authors would draw from this framework and their previous experiences with reflexive research to support the team with operationalising these principles, and organise critical reflection by the team on the evaluation's progress. The authors monitored the challenges the team encountered with conducting reflexive evaluation on a so-called Dynamic Learning Agenda (DLA; Van Mierlo et al., 2010). When the first evaluation cycle had concluded, the authors also assessed the policy impact of the reflexive approach, as commissioned by the project coordinators.

Material

All observations made by the authors as part of the project team were recorded in field notes, including the DLA. Additional field notes were kept during evaluation activities, including workshops, and seminars that were held at the PBL to inform the organisation on the project's progress. Data was also collected through 17 in-depth interviews with members of the team at the start, after the first year and after the finalisation of the first cycle. All members were interviewed at least twice over the course of three years. The interviews focused on their experiences with reflexive evaluation, and were flexible and open-ended in order to gain in-depth understanding of the rationale behind their actions regarding the implementation of reflexive evaluation. The interviews were audio-recorded and transcribed.

Data analysis

All data were analysed by authors A1 and A2 and corroborated with A3. We used content analysis (Hsieh & Shannon, 2005), making use of sensitising concepts derived from the NPT framework (the core mechanisms). A2 had not previously been involved with the Natuurpact case and brought in a different perspective to help sharpen the analysis.

4.4 Results

The findings show that all four of NPT's core mechanisms could be identified and several themes emerged for which alignment work was necessary, which in turn determined *how* the work for each core mechanism took shape. Time and again the team had to negotiate its reflexive approach in order to align it sufficiently with the prevailing modernist norms and customs regarding policy evaluation within their home organisation and nature policy practice. In the following, we discuss each NPT mechanism and how each mechanism became manifest in the form of alignment work undertaken by the team.

Sense-making work

The first core mechanism of the NPT framework is *sense-making work*: the work practitioners do to develop a shared understanding of the new practice and why it is important in relation to other practices (May, 2015). Coinciding with PBL's preceding interest in reflexive research, the agreement in the Natuurpact that its evaluation would have an ex durante character led to the decision to adopt a reflexive evaluation approach. The findings show that during the first stages of the project, the work the team did to make sense of what reflexive evaluation entailed beyond a general idea of its purpose and general process was limited. In what followed, the team set the approach apart from other approaches primarily with reference to its timing and the

level of interaction with its intended participants. This is illustrated by the project leader's explanation of reflexive evaluation: '*What is really different, is that we work with policy actors immediately from the start of the evaluation: there is a lot more interaction than there would be during a regular policy evaluation'* (PBL project leader, early 2015). Furthermore, the team differentiated reflexive evaluation on the basis of its purpose, explaining it as an approach that searches to '*reconcile evaluation purposes of policy learning and accountability'* to allow for transformative learning and change, as opposed to exclusively focusing either on learning or on accountability (as during responsive evaluation and impact assessment, respectively; Evaluation Plan, 2015).

Table 4.2 Excerpt from the theoretical framework for reflexive evaluation developed to guide the Natuurpact evaluation (Van Veen et al., 2016).

Evaluation element	Characteristics
Evaluation phases	*i.* Developing joint evaluation framework; *ii.* Joint research and interpretation; *iii.* Drawing shared conclusions and dissemination.
Guiding process principles	*i.* There is multi-stakeholder participation in the design and conduct of the evaluation; *ii.* The evaluation addresses the stakeholders' evaluation needs in light of the policy program's ambitions and draws attention to systemic properties that hamper or facilitate change; *iii.* The evaluation facilitates productive interactions amongst participants and experts; *iv.* Within this multi-stakeholder context the evaluation simultaneously allows for interactive learning, accountability and performance management, and draws attention to systemic properties that hamper or facilitate system change.
Outcomes	*i.* Socially robust knowledge and a knowledge-enriched policy practice; *ii.* Identification and challenging of systemic barriers and opportunities to increase the potential for system change; *iii.* Concerted action by involved stakeholders.

As part of sense-making, the team commissioned the current authors to develop a theoretical framework for reflexive evaluation (**Table 4.2**). The framework served to guide the team in their evaluation design and implementation. Our idea was that its principles required further operationalisation to make them work in the particular

project context. However, initially the framework found relatively little traction with the team; in a way, the team acted as if it had outsourced the work of sense-making by involving us as experts in reflexive research. After being given the formal go-ahead, the size and scale of the project caused the team to quickly become more occupied with 'doing' the evaluation than with 'thinking' about it.

Engagement work

We observed that the amount of work done for sense-making of reflexive evaluation increased when the team was faced with the need to do *engagement work*. NPT's second core mechanism concerns the work actors undertake to ensure engagement and commitment of others with the new practice, for instance through stressing its urgency and orchestrating managerial endorsement, but also through arranging sufficient capacity for others to become involved.

To encourage participation from nature policy actors as well as support from the PBL community – which were both vital for normalisation – the team first needed to make the purpose and value of reflexive evaluation explicit in a way that spoke to the prevailing conventions on useful and high-quality evaluation, all of which constitutes sense-making work. To do so, the team employed a strategy of reframing: developing two co-existing, mutually inclusive narratives that they used in different contexts. The first narrative emphasised the evaluation objective as: '*…allowing for mutual learning* [between policy actors and researchers] *from experiences with nature policy to timely inform policy processes and benefit the progress that is made on the nature policy ambitions'* (Evaluation plan, 2015) and was mostly used when communicating with policy actors. The second narrative describes the benefits of reflexive evaluation in terms of '*increased research quality and impact'* (PBL supervisor, start-up seminar early 2015) and was mostly confined to internal communications within the PBL organisation. In both narratives, the team framed the benefits of reflexive evaluation in a way that aligned with modernist ideals on the purpose of policy evaluation within their respective contexts: to enhance program performance (context of policy practice) and to promote research quality and impact (organisational context).

This course of events demonstrates how the need for engagement work gave rise to sense-making, and resulted in growing mutual understanding of reflexive evaluation among the team members, as well as the policy actors and the PBL community. This could be observed by a gradual decline in critical inquiry from actors from both host contexts after the approach was adopted, and by an increase in the team's confidence in explaining what reflexive evaluation is and does. Importantly, *how* sense-making and engagement work were done was determined by the need to align with the dominant frames within the intended contexts to implement the reflexive approach.

The project team reflected that developing commitment from both host contexts required a significant investment in engagement work: '*Convincing my constituency of the value of reflexive evaluation really required some work. I had a lot of informal chats with them, all to make them shareholders of this new approach'* (PBL project leader; 2016). Another example of how engagement work was informed by the need to align with modernist norms may be perceived in the team's decision to involve the current authors. While the original argument for our involvement was substantive, the team also mobilised our involvement strategically: to the PBL and policy communities they presented the purpose of our involvement as 'guarding the scientific rigor of the reflexive evaluation approach'. Through this strategic argument, the team used our involvement to underline the scientific integrity of the approach, adhering to the technocratic belief in truth claims that predominates within the PBL. The PBL project leader explained: *'I really used* [the current author's] *involvement to show: look, we are serious, this is also a scientifically sound method. This was critical to convince* [PBL colleagues] *that this was a valid evaluation approach.'*

How the engagement work to convince actors to buy into reflexive evaluation was done – by emphasising the scientific integrity of the approach –, was strongly shaped by the need to create alignment.

Enactment work

The third mechanism consists *enactment work* and comprises all acts of the team to 'do' reflexive evaluation. A starting point for this enactment work was the operationalisation of the four reflexive evaluation principles outlined in Table 2. We observed that operationalisation was again shaped by the need to align the evaluation design with both modernist customs and reflexive ideals. To explore this, we draw on the operationalisation of the first principle as an example: *There is multi-stakeholder participation in the design and conduct of the evaluation.* This principle is based on two premises. First, that complex problems require various stakeholders with different perspectives to become involved with a social learning process (Patton, 2010) and, second, that participation in evaluation design and conduct positively influences the relevance and shared ownership over the findings in light of the program's ambitions (O'Sullivan, 2012). Our example focuses on the latter. The idea of involving stakeholders throughout the entire evaluation evoked a strong response from both the team and the PBL community. Objections centered around the potential compromise of the team's objectivity and independence, and it was felt the credibility of the institute was on the line. These concerns were outed formally during the project's start-up seminar, but also in bilateral interactions with the project leadership. Consequently, participation was demarcated to specific evaluation phases, and the topics about which participants would have a say were limited. For instance, policy

actors were involved during the first phase, during which the scope and the main research questions of the evaluation were determined. Also, during the third phase, interactive stakeholder sessions were organised during which preliminary findings were collaboratively analysed and interpreted. The participants were not, however, allowed to discuss research methods during the second phase, nor given a say on the substance of the conclusions in the final report. Regarding research methods, one team member explained: *'It is up to us to decide what methods are most appropriate. It would do our independent judgement no good if we let policy actors decide how they want to be evaluated. We are, in the end, the experts.'* The PBL project leader confirmed: *'Letting them co-decide on methods, I can't account for that. It would be like allowing butchers to inspect their own meat.'* The participants considered this no issue at all: *'Research methods, that is really a topic for the researchers. I would have no idea',* one policymaker reflected. Modernist conventions – within both host contexts – on the distinct and separate roles of researchers and policymakers appeared beyond the scope of negotiation, regardless of the reflexive ideal to enable participation during all evaluation phases. Instead, this principle for reflexive evaluation was stretched in a way for it to sustain legitimacy in the eyes of the PBL and policy communities. In its re-negotiated interpretation, the principle came to mean something akin to 'participation not during "all" but during "most" evaluation phases'.

A similar discussion occurred concerning the dissemination of the findings. The team suggested to co-publish the final report with the participants, to underline their joint efforts. But because they required a visibly independent evaluation report for the recommendations in the evaluation to have strategic-political value, the policy actors strongly opposed co-publication. Interestingly, this was the same argument that had prohibited them from co-deciding on research methods earlier on, namely that, to their constituencies and the public, their participation might be perceived as compromising the scientific autonomy, objectivity and, ultimately, the credibility of the evaluation.

In these examples, the demarcation of participation to particular research phases had bend the reflexive principle in a way that neatly fitted in with modernist ideals of scientific autonomy and political distance. At the same time, it initially appeared to hold true to the principle's reflexive ideal: it still allowed for sufficient participation to facilitate social learning processes and generate shared ownership over its findings. However, during appraisal work, to which we turn later, it became evident that this principle had been stretched too far.

While the norms and procedures for objectivity and scientific autonomy were quite rigidly maintained at times, there were also moments when there was more room to

contemplate how to adhere to these norms while allowing for on-going science-policy deliberations. Notably, the procedures for guarding objectivity broadened over time. Rather than maintaining literal (physical) distance at all times, the team members distributed roles (some interacted with the participants while others focused on desk research and running models) to prevent researcher bias. Furthermore, for the final report, the peer-review community was extended to policy and societal actors. In doing so, the team implicitly expanded the norm of objectivity to intersubjectivity. While these broadened norms and procedures initially elicited critical remarks during project seminars, these remarks toned down as the evaluation progressed. Over time, convictions about the roles of evaluators and policy actors as strictly separated became less and less articulated, suggesting that actors from both contexts became more used to the changed relationship. Illustrative of this development is the wide uptake by the PBL community of the term 'requester' to replace 'principal' when indicating the commissioner of a project, suggesting a more equal-level footing between researchers and policy actors while simultaneously acknowledging researchers' autonomy.

Appraisal work

The fourth NPT mechanism concerns *appraisal work* and comprises activities that judge the value and effectiveness of a new practice during and after its enactment. In light thereof, the current authors were tasked with assessing the different ways the reflexive approach had had an impact on nature policy practice. This review was regarded with serious formality and weight, which in itself is illustrative for the degree of buy-in to the approach by the team and its supervisors. Notably, the review implicitly served a dual purpose: first, to learn from the experiences to improve the following cycle's execution and, second, to legitimise the continuation of the reflexive approach within the PBL community by demonstrating its value. This was manifest in our task to assess the approach to institutionally shared beliefs on what would constitute such value: enhanced research quality and policy impact. While our assignment was commissioned on the basis of a modernist logic, we strived for a more reflexive and deliberative approach, and brought to light effects more characteristic of reflexive evaluation. These included, for instance, enhanced policy learning, a strengthened nature policy community and a knowledge-enriched policy practice (Verwoerd et al., 2020; Verwoerd et al., 2017). Such effects beyond traditional linear ideas on policy impact have found uptake within the organisation and are sometimes used to discuss potential impact of new studies.

The review was critical on the operationalisation of some reflexive principles, which was argued to be ineffective in some respects. We return to the example of the operationalisation 'multi-stakeholder participation', which had been demarcated to

specific phases due to concerns about autonomy and the credibility of the PBL at large. The review identified that findings' relevance and usability was limited: a mismatch was observed halfway through the project between the scale at which the public actors' evaluation needs transpired (regional) and the scale at which the computational model that had been used provided findings (national). As a result, the participants felt the findings were of limited use to inform their nature policy plans as they provided few perspectives on regional action. Regarding this mismatch, the WER project leader noted: *'That this model would be used was decided upon before we had even started. […] It cost us a lot of time and effort to explain and repair this mismatch to policy actors. The decision which model to use should have been informed by the demands of our intended end-users.'* In retrospect, by not involving the participants in the discussion on research methods, their evaluation needs were initially only partly met. Demarcating the first principle thus also compromised the second, and reduced the usability of the evaluation for policy learning and change. To remedy the mismatch, the project team undertook a significant amount of work to produce findings on a more relevant scale.

The team later reflected that while opening up the determination of the methods to policy actors had been regarded as being beyond discussion, some members in addition had defaulted into working in a 'research-driven' as opposed to a 'practice-driven' fashion. In hindsight, the initial shared understanding at the start of the project of what reflexive evaluation is and how 'it is done' appears superficial. Different understandings of reflexive evaluation emerged during its actual implementation and materialised in the form of conflicting modes of working. The team's preoccupation with 'doing' reflexive evaluation absorbed time for thinking and discussing the approach, allowing different modes of enacting reflexive evaluation to remain untouched and deeply embedded routine ways of working unchallenged.

Despite these difficulties, after the first evaluation cycle had concluded the overall feeling amongst those involved was one of enthusiasm, and the reflexive approach was continued during the second cycle. Moreover, at the time of writing, several other large-scale reflexive evaluation projects have been initiated at the PBL, and notions such as 'reflexive thinking' have found some uptake in the organisation's vocabulary.

4.5 Discussion and conclusion

This paper aimed to empirically investigate the process through which evaluators attempt to conduct and, in so doing, normalise an evaluation practice that draws on a reflexive evaluation logic in a context partial to modernist logic. In this section, we reflect on our findings and, starting off from the idea that developing legitimacy without compromising integrity is vital to successfully normalise reflexive evaluation, make some suggestions for evaluators seeking to conduct and normalise reflexive

evaluation. Before we present these, we briefly reflect on the value of using NPT as theoretical lens for understanding and facilitating such normalisation.

Over the past decades, a wide variety of views on the purposes and roles of policy evaluation have proliferated in evaluation literature, accompanied by myriad evaluation approaches and methodologies (Stern et al., 2015). Some more recent approaches share that they are stakeholder-oriented and search to better address the complexities of the phenomenon under study while contributing to its goals and ambitions (e.g. Arkesteijn et al., 2015; Verwoerd et al., 2020). Much of the literature on these emerging practices concerns their theoretical underpinnings, experiences with applications in specific cases, or practical or theoretical differences between distinct approaches (Fetterman et al., 2015; Moore et al., 2019; O'Sullivan, 2012; Rolfe, 2019). There is a dearth of literature on challenges involved with implementing novel practices in institutional settings that are not necessarily conducive to it (Chouinard, 2013; Guijt, 2010; Kunseler & Vasileiadou, 2016; Petersen et al., 2011). An increased understanding of the processes through which evaluators address these challenges may support evaluators aspiring a reflexive logic to create room for such work in contexts where modernist approaches are privileged and to do so without compromising the reflexive ideal. In response, this study analysed the work that was done by an evaluating project team to normalise reflexive evaluation, using NPT's core concepts.

Our findings confirmed that the core mechanisms for normalisation do not occur linearly, but rather in conjunction: each mechanism caused iterations in the others, and vice versa, continuously strengthening each other (McEvoy et al. 2014). In our case, many of the challenges in engagement, enactment and appraisal work could be traced back to lack of mutual in-depth understanding and agreement amongst evaluators (and participants) on how (not) to 'do' reflexive evaluation. Our findings suggest that as much of the work that was undertaken towards normalisation occurred relatively ad hoc and in response to urgent unaccommodating structures or aspects of political and organisational culture – such as beliefs about who has a say in determining research methods – implicit conventions and evaluation routines factually remained unchallenged. Our findings resonate with challenges identified for the introduction of participant-oriented evaluations to development projects, where lack of participatory sense-making on the appropriate evaluation approach was found to conduce the defaulting into modernist approaches (Van Hemelrijck & Guijt, 2016). Nieminen and Hyytinen (2015) suggest this is reinforced due to limited methodological repertoires that evaluators who seek to practice more systemic and reflexive approaches tend to draw on, consequential to the institutionalisation of modernist logic and the epistemological differences that hinder different logics'

hybridisation. Although unrelated to reflexive evaluation per se, scholars of implementation science also underline the importance of inclusive and mutual sensemaking *prior* to implementation for a new practice to become successfully normalised (Mair et al., 2012).

As illustrated by the continuation of the Natuurpact reflexive evaluation, and the initiation of various other reflexive evaluation projects, normalisation has been at least partly successful. Despite the ad hoc character of their actions, the team succeeded in establishing alignment, or a 'fit', between the two evaluation logics. Crucially, this fit did not appear as a fixed state, but rather as a negotiated, emergent and dynamic accomplishment. Indeed, the alignment work for normalising reflexive evaluation encompassed navigating and negotiating *reflexive evaluation legitimacy* on the one hand and *reflexive evaluation integrity* on the other. Specifically, our study has shown that alignment work comprises various strategies, including reframing the purpose of reflexive evaluation for it to make sense from the point of view of the dominant institutional logic. Another strategy concerned emphasising the scientific rigor of the approach to demonstrate its validity as a peer-reviewed research approach. Such strategies 'work' as they ensure the legitimacy of the reflexive evaluation approach from the modernist logic perspective. At other times, such strategies proved insufficient to acquire legitimacy: regardless of framing or appeal to scientific rigor, in our case the involvement of policy actors in deciding on research methods, drawing conclusions or co-publication, were topics largely beyond discussion both for evaluators and participants. Here, the dominant modernist logic's norms could not be negotiated to develop legitimacy.

May et al. (2016) have studied the degree to which contextual structures and cultures may be negotiated to normalise a novel practice and refer to contextual *elasticity*: the ease with which contexts accommodate new ways of working. They propose that the greater the elasticity of contextual structures and cultures, the less work is required from practitioners of a new practice to develop legitimacy for it. In relation to evaluation logics, this implies that the more the modernist logic is institutionalised and embedded, the more work is required from evaluators to normalise a reflexive practice – i.e., to develop understanding, buy-in, agreement on enactment and appraisal. From our findings it appears that when the roles of researchers and policy makers seemed to become too intertwined, the reflexive logic was disciplined by actors from either context, or both.

We also observed that some normative structures became more accommodating over time, particularly when the value of the reflexive approach became materially tangible, as happened when the lack of participation in deciding upon research methods

resulted in only limitedly useful findings for policy learning. Such observations confirm that normative structures more accommodating to the reflexive logic can be developed (Verwoerd et al., 2020). More in-depth studies of the elements that constitute contextual elasticity and of how more accommodating contextual structures and cultures may be developed, may be fruitful to further the understanding of normalisation of reflexive evaluation. For instance, past research has shown that participation – as constitutive element of reflexive evaluation – can be obstructed by diverse political, administrative and social barriers and power inequalities (Engel & Carlsson, 2002; Gregory, 2000; Lehtonen, 2014). It would be very useful to study the role of such contextual elements in the context of attempts to normalise reflexive evaluation logic.

Moreover, we observed that when contextual structures and cultures were unaccommodating, evaluators navigated these by altering components – principles, procedures, purpose – of the reflexive approach. May et al. (2016)'s concept of innovation *plasticity* comes in useful for understanding this, as plasticity refers to the malleability of components and the discretion practitioners have to develop alignment between the contexts and the innovation. The greater the degree of plasticity of innovation components, the less work is required from practitioners to normalise it (ibid). Already in 1977, Fullan & Pomfret discuss that a level of on-site modification may be required for innovations to effectively meet context-specific needs. However, adaptation beyond the 'zone of drastic mutation' (Roitman & Mayer, 1982:3) may compromise the fidelity to, or integrity of, the original intentions. Indeed, in our case, some reflexive components were stretched too far: not involving participants in making methodological decisions led to a decreased chance of policy learning. It appears that there are certain parameters that determine the plasticity of reflexive evaluation components (i.e., the zone of drastic mutation) and the leeway enactors have to retain reflexive evaluation integrity. These parameters were obscure (it was not immediately clear if or when a line was crossed) and, as for contextual elasticity, what precisely determines these parameters, and thus the plasticity of reflexive evaluation components, requires additional inquiry.

Contextual elasticity and innovation plasticity are useful concepts to think about the required alignment work for negotiating and navigating reflexive evaluation legitimacy and integrity. Using both concepts in conjunction may shine new light on the explanation as to why it is so difficult for practitioners to innovate towards a more reflexive evaluation practice. This is important as scholars have previously problematised successful participation and it has been pointed out that, despite ideals for new roles for science and policy, novel repertoires often deviate little from their technocratic counterparts and traditional ideas on the relationship between scientists

and policy makers (Reinecke, 2015; Turnhout et al., 2013). New 'in vogue' terms, in our case reflexive evaluation, may even cover up this reality, promoting legitimacy for 'innovative' approaches while actual practices remain unchanged. As Van Der Hel points out, limits to elasticity and plasticity seem a pertinent explanation of why researchers are inclined to do *"more of the same under a different name"* (2016:173). As researchers are pushed back by the contexts they operate in, they cannot help but default into modernist logic routines, and end up (unwittingly) greenwashing, tokening and technocratising participation (Chilvers, 2008). This is relevant, as such occurrences may promote participants' subjugation rather than their empowerment (Aarts & Leeuwis, 2010; Turnhout et al., 2013), thereby reinforcing the modernist status quo.

We conclude with a reflection on the use of NPT for studying normalisation of reflexive evaluation. Our findings confirmed the theory's potential: applying NPT provided analytical depth to investigate how reflexive evaluation was operationalised and helped identify the evaluators' challenges with implementation. Consistent with others' experiences, applying NPT was not instantly intuitive (McEvoy et al., 2014). Although its numerous (sub)concepts make the theory versatile in use, these also required significant translation effort to render them applicable to the specific context of our study, an issue also addressed by Finch et al. (2012). For the purpose of our study, we decided early on to centre our analysis around NPT's core concepts, which, following the example of McNaughton et al. (2019) and with the aim to better engage with them, we re-labelled into more accessible language (e.g., 'coherence' was re-labelled as 'sense-making').

Others have pointed out NPT's undue emphasis on individual and collective agency, discarding the influence of organisational and relational contexts in which implementation occurs (Clarke et al., 2013). We engaged with this shortcoming by including the concept of institutional logics, which drew attention to the interacting and nested contexts (team, organisation, societal interactions) in which these logics were enacted and through which they shaped agency.

Overall, we conclude that timely investment in mutual sense-making is recommendable for normalisation of reflexive evaluation to occur and seems especially pertinent for joint understanding of the components of reflexive evaluation, their plasticity and the parameters that safeguard reflexive evaluation integrity. In doing so, the methodological repertoires evaluators draw from may be increased and strengthened. In a similar vein, we conclude that considering unaccommodating contextual structures and aspects of political and organisational culture (and their elasticity) from the onset may help evaluators anticipate these to timely orchestrate

alignment strategies to develop evaluation legitimacy. For both cases, using NPT prospectively and *in action* (de Brún et al., 2016), may be fruitful to guide structural reflection and learning-by-doing and hence can facilitate the effective navigation of reflexive evaluation legitimacy and integrity in the normalisation process.

5 | NEGOTIATING SPACE FOR KNOWLEDGE CO-PRODUCTION

Abstract

Building on the preceding chapter, this second empirical chapter explores more in-depth the alignment activities undertaken by policy researchers to assure a fit between their selected knowledge co-production approach and conventional norms and customs for knowledge production they encounter at the science-policy interface. This chapter reflects upon this thesis' second study: a comparative case study of the Natuurpact (first and second cycle), Vibrant Rural Areas, and Regional Deals research projects, and the challenges the respective project teams encounter with implementing their selected co-production approach (sub-question 1) and how they seek to address these (or not) (sub-question 2).

In the chapter, two types of alignment activities are identified: contextual restructuring of normative and relational customs, and modification of co-production features to accommodate a contextual fit. The findings on the implementation of knowledge co-production approaches in the three projects furthermore showed that policy researchers developed co-production capacity over time and that they become more skilled at contextual restructuring and creating a fit without compromising co-production integrity. Additionally, distinct organisational developments proved conducive to the normalisation of knowledge co-production, including the instalment of a Community of Practice that provided a platform to reflect upon the projects, as well as taking on a new 'chief scientist' with affinity with deliberative and reflexive policy research approaches. It is argued that investment in both policy researchers' co-production capacity and their ability to recognise and navigate (un)conducive structures is required to create space for knowledge co-production in science-policy systems. An action-reflection cycle is proposed as heuristic tool to help build capacity and support policy researchers with *observing* whether co-production has sufficient legitimacy in its intended contexts, *reflecting* on (un)conducive contextual structures and on co-production integrity, *planning* activities to achieve alignment and *acting* out these planned activities to orchestrate a contextual fit.

The chapter also touches upon the views of involved policy actors with knowledge co-production and on developments in the policy context in which the projects occur, and how these affect the normalisation of knowledge co-production. This is further explored in the following chapter, which discusses the policy context of the Natuurpact research project in particular.

Revisions in progress
Verwoerd, L., Brouwers, H.J.H., Kunseler, E.-M., Regeer, B.J., de Hoop, E. *Negotiating space for co-production.* Science & Public Policy.

5.1 Introduction

Knowledge co-production is an increasingly popular research approach to address the unprecedented environmental and sustainability challenges of our time. Adopted in (policy) projects and programmes geared towards socially robust transformation for sustainability problems, its key characteristics comprise the inclusion of (a diverse set of) stakeholders, focus on reflection and mutual learning for actionable knowledge, integration of knowledge, and emergent and responsive research designs (Pohl et al., 2010; Scholz and Steiner, 2015a; Turnhout et al., 2013).

Despite increased popularity, knowledge co-production practices are also contested. Various authors identify a theory-practice gap (Boon et al., 2019;Flinders et al., 2016) and highlight how achieved impacts do not live up to theoretical expectations (Jagannathan et al., 2020). In addressing the discrepancy between co-production outcomes in theory and real-life settings, various scholars consider the role of societal, political, cultural and organisational contexts therein. They demonstrate how prevailing norms, procedures, values and conventions regarding knowledge production in these context render such them less conducive to theoretical-ideal practices of co-production (Felt et al., 2016; Klerkx et al., 2017; Turnhout et al., 2020). Indeed, as Flinders et al. (2016, p. 262) point out, co-production '*rubs up against traditional social norms and roles, which may mean its potential as a radically innovative form of research encounters problems both in theory and practice'*. In response, scholars have argued that co-production requires institutionalisation – or normalisation – to meet its transformative potential (e.g., Braun & Könninger, 2018; Jahn et al., 2012; Schneider et al., 2019). In the literature, however, little detail is available on how such processes of normalisation might precisely occur in practice (Arnott et al., 2020). It is to this particular knowledge gap this paper seeks to contribute.

Our objective is to advance theoretical understanding of the influence of context on the process by which researchers put theoretical-ideal features of knowledge co-production to practice. Additionally, we aim to offer practical insights to further promote the normalisation of knowledge co-production in policy science. As we will further elaborate upon in the theoretical background (Section 2), we conceive the normalisation of a practice as an ongoing and dynamic process in which researchers seek to establish alignment between prevailing contextual structures – i.e., norms, rules, etc. – on the one hand, and the theoretical-ideal features of co-production on the other. By establishing alignment, policy researchers arguable craft space for knowledge co-production's practice and normalisation. How precisely they may succeed in establishing alignment – i.e., what activities do they undertake and why, and at what are these activities directed – comprises the central focus of this paper.

Additionally, as it is unlikely that the conditions conducive to co-production are all in place at the outset (Broerse, 1998), previous studies have underscored the importance of learning by policy researchers to successfully practice co-production in dynamic, contested and multi-stakeholder contexts (Lang et al., 2012; Norström et al., 2020; Verwoerd et al., 2020, 2021). Hence, we also have attention for the building of policy researchers' co-production capacity that is arguably required to develop such conditions.

We delimit our scope to science-policy co-production; i.e., the collaborative efforts of policy researchers, policy actors and relevant (societal) stakeholders to co-produce knowledge that is credible, legitimate and salient for informing policy decision-making processes (Cash et al., 2002). To study science-policy interactions, boundary organisations may be of particular interest. Such organisations operate at the science-policy interface from where they play an important role in linking science and policy decision-making (Guston, 2001). To maintain their authoritative position on the science-policy interface, they have to sustain their societal legitimacy and credibility by constantly adjusting to changes in their wider institutional and political contexts, while simultaneously ensuring adherence to both policy makers' demands and scientific standards (Pesch et al. 2012; Pielke 2007). As boundary organisations are strongly bound to their institutional, political and scientific contexts, they potentially offer a unique opportunity to study how these contexts are negotiated to craft space for co-production. Empirically, we draw from the experiences of researchers from the PBL Netherlands Environmental Assessment Agency (in Dutch: *Planbureau voor de Leefomgeving;* PBL), a boundary organisation in pursuit of normalising knowledge co-production in policy evaluation and its embedding in the organisation's research repertoire.

5.2 Background

Knowledge co-production on the rise: observing a gap between theory and practice

This paper understands knowledge co-production as a knowledge-production methodology characterised by intentionally opening up scientific knowledge production processes to the perspectives of stakeholders involved with the problem.[7]

7 Originally, the term co-production was introduced in the field of Science and Technology Studies (STS) to understand interactions between science, technology and society. This analytical understanding of co-production emphasises how knowledge and social order co-shape and co-evolve, underscoring that

We thus follow scholars who consider it as a mode of research in which the social contract for science as disconnected is replaced with a more inclusive, socially robust, and deliberative research culture to bring about societal transformation and sustainable development (Arnott et al., 2020; Gibbons, 1999). According to the literature, key features of co-production include the collaboration between researchers and a diverse set of public actors and stakeholders involved in the problem, inter- and transdisciplinary knowledge integration, an emergent research design that allows researchers to be responsive to developing insights or other external developments, and the development of actionable knowledge to bring about societal transformation and sustainable development (Hoffmann et al. 2017; Regeer and Bunders 2009; Scholz and Steiner 2015a).

The increasing popularity of co-production has been accompanied by an expanding body of literature, much of which is focused on theoretical foundations (e.g., Scholz and Steiner, 2015a), epistemological underpinnings (e.g., Regeer and Bunders, 2003), and conceptual and methodological advancements in terms of principles, mechanisms, and check-lists and guidelines to enhance co-production effectiveness and quality (Bergmann et al., 2021; Jahn and Keil, 2015; Lang et al., 2012; Norström et al., 2020; Polk, 2015). Other parts have been dedicated to studying the societal impact and quality of knowledge co-production and dealing with assessing the different types of long-term, contingent, and sometimes unintended effects of co-production (Verwoerd et al., 2020; Walter et al., 2007; Wiek et al., 2014). Somewhat more recently, scholars have started to point out observed discrepancies between the outcomes co-production is alleged to achieve in theory and those that are realised in practice (Flinders et al., 2016; Oliver et al., 2019). In their recent review of 21 co-production projects described in literature, Jagannathan et al. (2020) demonstrate a gap between co-production projects' outcomes reported in practice and the theoretical scope of their ambitions. While outcomes such as enhanced knowledge use were frequent, the societal transformation that co-production promises in theory was less frequently reported than may be expected based on theory. Other authors have reported outright failures, where co-production projects paradoxically reinforced the status quo for marginalised actor groups rather than transforming the power relation that the projects intended to do (Felt et al., 2016; Turnhout et al., 2020). Arguably, it raises the question whether knowledge co-production projects and programmes

knowledge is simultaneously 'a product of social order and […] constitutive of forms of social life' (Jasanoff, 2004 in Van der Hel, 2016).

should be accompanied with more ambition or more caution (Flinders et al., 2016; Lemos et al., 2018).

Why context matters for knowledge co-production implementation

While part of the observed theory-practice gap may be explained by a lack of (and difficulties with) rigorous long-term evaluation of co-production outcomes (Schäfer et al., 2020), a growing body of work points towards societal, political, cultural and institutional barriers (Scholz and Steiner 2015b). For example, in their literature review on the political and power dimensions of co-production, Turnhout et al. (2020) show how prevailing power relations and politics affect how co-production unfolds in practice and what outcomes are achieved. They show how attempts to depoliticise co-production practices leave unequal power relations unchallenged, perpetuated and even reinforced, thereby decreasing the likelihood of empowerment and societal transformation. Barriers to co-production's key feature of collaboration between researchers and practice experts have also been identified, such as organisational and institutional differences that may result in disparate perceptions of time, knowledge requirements, goals and mentalities (Boon et al., 2019; Choi et al., 2005), and epistemological differences on judging the value of knowledge (Hegger et al., 2012). Additionally, diverging from more regular modes of science may (be perceived to) pose risks to academic careers, especially for young scholars (Felt et al., 2013).

It has been argued that science-policy systems in the global North are largely organised by techno-bureaucratic institutions that aim to produce scientific facts that are free from political values, perpetuating the boundary between science and policy (Chouinard, 2013; Flinders et al., 2016). This has given rise to institutionalised normative dispositions towards the appropriate relationship between researchers and practice experts, and the appropriate role of practice experts in scientific knowledge production (van der Hel, 2016). Especially during political crises, it may be in the interest of politicians and policymakers to maintain the boundary between science and policy, as it allows them to shift blame and avoid responsibility when negative policy research is published (Kowalczewska and Behagel, 2019). Consequently, while environmental science-policy research may start with the intention of co-production, the level of 'collective experimentation' that is required to develop actionable knowledge is hardly achieved (Felt et al., 2007:26).

Notably, the above examples all describe *contextual* barriers and their implications for the practice and outcomes of co-production. Here, we side with Felt et al. (2016) who argue how contexts (they refer to 'places') matter: they contain implicit norms, values, customs and power dynamics and actor constellations that prescribe actors' behaviour and are rooted in shared '*experiences with issues related to technoscience, sense-making*

narratives, experiences and recognised processes of non-expert participation, as well as routine ways of assessing and handling knowledge claims' (p. 738). Such situated contextual properties affect the degree of space for knowledge co-production.

Because the contextual limits to co-production are arguably of a systemic character (Flinders et al., 2016), institutional change is required for it to become part of standard research practice. However, as Arnott et al. (2020) point out, little knowledge is available on how contextual structures rooted in conceptions of knowledge production as a linear process and with a strict separation between scientists and practice experts, may be successfully navigated to support and promote co-production practices on the science-policy interface. It is to this particular gap this paper seeks to contribute.

Normalisation as a dynamic and emerging process

This paper aims to gain insight into how policy researchers practise co-production in contexts that are more partial to linear conceptions of knowledge production and science and policy as strictly separated worlds, and how this affects co-production practice and outcomes. In light of this we suggest to consider the introduction of knowledge co-production as an innovation in organisational (research) practice. In the following text we propose a theoretical lens through which to study the work policy researchers do to implement and normalise co-production and discuss its operationalisation for the purpose of this paper.

Since 50 years, scholars have highlighted that approaching implementation as a process that takes place in interaction with its context allows for understanding why some innovations find uptake into standard practice while others do not (Fullan and Pomfret 1977; Havelock 1979; Huberman and Miles 2013). May and Finch (2009) further conceptualised normalisation processes as the outcomes of dynamic interactions between the innovation's components and its intended 'host context'. For an innovation to normalise, a level of 'fit' or alignment needs to be established between the innovation and its context, and May et al. (2016) have identified two possible routes to do so.

Restructuring normative and relational contextual structures

We expect that policy researchers who seek to practise and normalise knowledge co-production encounter unconducive normative and relational contextual structures in various settings. The first route to establish alignment concerns negotiating these dominant normative and relational structures. Through collective action and the mobilisation of resources, policy researchers may encourage normative and relational restructuring: changes in norms, conventions and routines, and in the entrenched

ways that actors interact (or not), respectively.[8] When successful, restructuring increases the space for practising co-production without making any modifications to its features.

We expect that normative and relational structures in practice are interrelated and share a certain degree of overlap, i.e., what is considered appropriate considering how knowledge is produced likely has consequences for how relational structures, such as interactions between scientists and policymakers, are enacted. Furthermore, as our focus of inquiry is directed at the implementation of knowledge co-production in projects that occur in contexts of a policy research organisation and its wider policy setting, we expect the structures to be of an *epistemic* nature. Epistemic norms for instance consider how different ways of knowledge production are valued and considered valid and legitimate. The distinction nevertheless provides analytical depth to our inquiry.

Modifying knowledge co-production components

Sometimes, it is argued, contextual structures may appear too rigid to restructure. In such situations, alignment may be achieved by modifying the innovation's components, such as co-production's key features as identified in the literature. This is the second route identified by May et al. (2016). It is argued a certain degree of modification is generally required for an innovation to be workable and aligned to contextual structures that cannot be restructured.[9] However, scholars warn that plasticity beyond the 'zone of drastic mutation' (Roitman and Mayer 1982:3) may result in the loss of the innovation's original intent, compromising the innovation's integrity (May et al. 2016). For instance, Felt et al. (2016) point out how 'structural fixes' in case of a misfit, such as reducing stakeholder participation, are often deployed rather than instigating radical change in attempts to open up research to society. This example may be explained by the over-modification of a key-feature of co-production (i.e., stakeholder participation) in the face of too-rigid structures, by which the integrity to co-production's ideal might be lost (such as stakeholder empowerment).

We expect that when policy researchers encounter normative and relational structures that are too rigid to negotiate and restructure, they will modify certain features of

[8] May et al. (2016) refer to contextual *elasticity* to address the ease with which these structures may be restructured to allow space for a new practice.
[9] May et al. (2016) refer to the innovation's *plasticity*: the degree of malleability of its components and the discretion practitioners have to alter these to fit the context without compromising its original purpose.

knowledge co-production in order to establish the required fit for its implementation and, eventually, normalisation. Here, we expect researchers to risk compromising co-production integrity when co-production features are altered in a way that fidelity to its original purpose is lost.

The above conceptualisation of normalisation of co-production corresponds with findings in the field of sustainable development that demonstrate a degree of redesigning necessary for an innovation's social embedding. Hoes and Regeer (2015) describe how the adoption of a novel innovation may be promoted through functional alignment activities (by adding or removing new components of the innovation or by adding new rules) and conceptual alignments activities (by coupling or uncoupling the innovation to existing problems, polities and politics, and reframing how the novel innovation fits). Alignment activities help researchers to act 'in-between': they assure a contextual fit, while they slowly transform conventional institutions, identities and approaches of knowledge production at the science-policy interface (Kunseler 2017).

5.3 Methodology

Embedded case study

We applied an embedded multiple case study design to study how policy researchers negotiate space for knowledge co-production and its normalisation. Embedded case studies are appropriate to study phenomena in their real-world context, and involve multiple 'sub-units of analysis' to allow for a detailed level of inquiry (Scholz & Tietje, 2002; Yin, 2003).

Knowledge co-production at the PBL Netherlands Environmental Assessment Agency (in Dutch: *Planbureau voor de Leefomgeving,* PBL) serves as paradigmatic case (Flyvbjerg, 2006). PBL is a national policy assessment agency that identifies itself as an authoritative and impactful knowledge institute on the (environmental) Dutch science-policy interface and has formulated its mission as '…contributing to the quality of political-administrative decision-making', stressing the independence and scientific grounding of its research (PBL, 2019). The majority of its researchers have backgrounds in the natural sciences. While established under a positivist technocratic paradigm, the complexity of contemporary policy issues, accompanied by multi-level and multi-actor governance configurations, has led PBL's managers to start aspiring for co-production to become normalised within the organisation's research practice. In line with this pursuit, a large-scale co-production project to evaluate nature policy was initiated in 2014. Embarking on this project sparked discussions within PBL on the validity of co-production, as many PBL researchers felt uncomfortable with co-production for evaluation purposes as it seemed to compromise the organisation's

independence and credibility as scientific government advisor (Kunseler, 2017). In the years that followed, two more co-production projects for policy evaluation were conducted at the PBL. All three projects mentioned here figure as cases to study the activities PBL researchers undertake to normalise co-production in this paper.

Table 5.1 Overview of the projects' characteristics.

Case	Case 1: Knowledge co-production for the evaluation of the Natuurpact programme	Case 2: Knowledge co-production for the Regional Deals policy programme	Case 3: Knowledge co-production for the evaluation of the Inter-administrative Programme for Vibrant Rural areas
Respective policy programme and its ambitions	The Natuurpact programme, aimed at halting the biodiversity decline and improving societal relations with nature	The Regional Deals programme, aimed at making 'deals' between national governments and Dutch regions to advance regional wellbeing	The Inter-administrative Programme for Vibrant Rural areas, directed at enhancing inter-administrative collaboration and instigating a transition towards a vibrant countryside
Policy context	Decentralised nature policy from the national government to the 12 provincial governments; increasingly horizontal governance of provincial nature policy	Inter-administrative partnership between the national government and 30 Dutch regions	Inter-administrative partnership between national government, provincial government, municipalities and water board directives, manifest in 15 rural areas
Period of the project	2014-ongoing (comprised of consecutive three-year cycles, of which the first (2014-2017) and second (2017-2020) had been concluded at the time of this study)	2018-2021	2019-2021
Formal commissioners	2 commissioners: The Ministry of Agriculture, Nature and Food Quality (Landbouw, Natuur en Voedselkwaliteit; LNV), the Interprovincial Council (Interprovinciaal	1 commissioner: LNV	4 commissioners: LNV, IPO, Association of Dutch municipalities (Vereniging der Nederlandse Gemeenten; VNG), Union of Water Boards (Unie van Waterschappen; UvW)

	Overleg; IPO)		
Composition of project team in terms of scientific backgrounds	Environmental (3); public administration (2); co-production (1)*	Economics (2); spatial planning (3); public-administration (1); co-production (1)	Environmental (2); public administration (4); co-production (3)
Experience with co-production	The public administration and co-production scientists had previous experience with co-production (albeit the former two on a smaller scale)	The public-administration and co-production scientists had been involved with case 1	Three scientists had experience with co-production, among which 1 project leader

* The composition displayed here concerns the initial composition of the case 1's team, which changed slightly over time.

Co-production for policy evaluation constitutes a particularly opportune setting to study how researchers seek to navigate or accommodate challenging and rigid structures to practice knowledge co-production. Policy evaluations inherently have political dimensions (Bovens et al., 2009) that harbour institutionalised societal and political expectations regarding what policy research is and ought to do. As a consequence, the contextual structures in which policy evaluation takes place are often particularly rigid, reducing the space for knowledge co-production.

The cases

The three selected projects evaluated multi-level and multi-actor policy programmes, and aimed to develop usable, relevant and socially robust knowledge and recommendations, to both evaluate and support the programmes in reaching their goals. Research questions were developed in response to questions from practice experts and data was collected via 'transformative' interviews and reflective multi-stakeholder sessions. Researchers and practice experts jointly worked towards producing socially and politically robust action perspectives based on the projects' research conclusions, which were subsequently reported by the researchers in the final reports. The projects differed with regard to the different policy programme evaluated, commissioners, duration, the composition of the project teams and the degree of their experience with co-production (see **Table 5.1**).

Participant observation

The data for this research is derived primarily from participant observation and supported by interviews, document analysis and reflection sessions.

Participant observation

All authors played several roles in the selected cases. Firstly, all authors were involved as researcher-practitioners in one or more of the selected cases (author 1 in case 1 and 2, author 2 and 3 in case 3, author 4 in case 1 and author 5 in case 1 and 3). As research-practitioners, we co-developed the projects with PBL and supported co-production's implementation. Secondly, we were also involved in the process of normalising knowledge co-production at the organisational level. All authors have been in contact with all cases studied, and with other relevant projects at PBL. Such encounters include author 2 attending seminars and project meetings of case 1 and 2 respectively, author 5 conducting scientific peer review for another reflexive research project at PBL, and author 1, 2, and 5 mapping the societal impact of case 1. Lastly, author 1 and 4 currently work at PBL and are tasked with supporting the organisation's pursuit of rendering PBL research practices more reflexive and advancing the normalisation of co-production. Taking on roles of participant-observers for this paper, a method built upon traditions of action research (Wittmayer & Schäpke, 2014), our embeddedness allowed for a view 'from the trenches': in-depth insight into the challenges and barriers researchers encountered when implementing co-production and the actions they take to overcome these, as presented in this paper.

Additional data collection and analysis

In addition to our participant observations (recorded in field notes), we gathered additional data through document analysis, ten in-depth semi-structured interviews and four reflection sessions. The documents analysed include project proposals and documents containing the official requests as formulated by commissioners, providing us with fundamental information about the intention, goals and nature of the project. Interviews were conducted with the cases' project leaders, several team members and the supervisor of two of the three cases. The interviews covered key motivations to conduct a co-production project (the goal and nature of the project and a reflection on the team's and PBL's role in the policy program), the involvement of stakeholders (who, in what way and to what extent), and the way in which the project team collaborates (in terms of knowledge integration between different disciplines and encouraging reflexivity). After conducting the interviews, a case description was developed for each case, describing their most prominent challenges and how these were addressed. Drafts thereof were member-checked by the respective research teams to strengthen the validity of our observations (Birt et al., 2016). Subsequently, reflection sessions were organised, one for each project team, to reflect on the findings of the interviews and the case description. During a fourth and final session, members of all three cases, supervisors and the chief scientist of PBL discussed action perspectives to address PBL's challenges in normalising knowledge co-production.

5.4 Results

Zooming in: researchers' alignment activities for implementing co-production

We first adopt a zoomed in vantage point and present the actions undertaken by each of the three project teams to establish normative and relational alignment between contextual structures researchers experience and the aspired co-production approach. As we will show, epistemic friction plays a major role therein: most PBL researchers have a natural sciences background, rooted in positivist codes of conduct and a technocratic understanding of science-policy interactions while knowledge co-production draws from social-constructivist epistemologies and a deliberative orientation to science-policy interactions. Actions to establish normative and relational alignment could be divided into the two distinct activities introduced in the background: *restructuring* prevailing structures, and accommodating to these same structures, by *modifying* co-production features.

Normative alignment

The term normative alignment refers to alignment between prevailing norms – conventions, rules and resources (May et al. 2016) – for knowledge production within the multiple contexts of implementation and the theoretical ideals for knowledge co-production. Established normative structures regard objectivity, independence and scientific autonomy as golden standards for scientific rigor and policy impact within the PBL organisation. These (predominantly epistemic) norms were for example enacted during internal project seminars (a common format used at the PBL to discuss projects' design or intermediate results with interested colleagues beyond the project team), where colleagues critiqued the projects for putting researchers' independence at risk, and, consequently, the organisation's credibility. This was especially the case for case 1, as the first policy evaluation of its scale and duration that adopted a co-production approach, but could also be observed at the start of case 3. Case 2 was less visible in the organisation as no project seminars were organised, and therefore experienced less critical inquiry within the organisational context. Nevertheless, the overall organisational legitimacy of and support for co-production was low. In all three cases, unconducive norms were also encountered within the project teams, complicating interdisciplinary collaboration. We discuss the activities the teams undertook to address this initial lack of legitimacy that resulted from these normative frictions between conventional norms and co-production ideals, and that were directed either at restructuring or accommodating conventional normative structures. **Table 5.2** displays these activities.

Table 5.2 Activities undertaken by the project teams directed at normative alignment.

Activities to establish alignment between existing norms and the theoretical ideals of knowledge co-production, by...	
restructuring of normative structures	• Coupling co-production to impact frame • Coupling co-production to a frame of scientific quality • Creating support network • Establishing boundary objects •
modifying co-production features	• Demarcating collaboration to specific research phases

Restructuring normative structures

A first alignment activity to create organisational legitimacy, undertaken by the case 1 and 3 teams, was *coupling co-production to the impact rhetoric* that is prominent within the organisation by arguing that the complex, uncertain, multi-actor multi-level governance character of the policy issue at hand demanded a more deliberative approach to '…increase the quality, usability and impact of the produced knowledge' (final reports case 1, PBL & WUR, 2017:27; 2020:54) Additionally, case 1 invested in assessing its policy impact and in publishing the results of this assessment in various (Dutch) scientific and policy publications. By framing co-production in terms of impact, the teams appealed to the organisation's identity as an impactful knowledge institute.

Secondly, we observed that the case 1 and 3 teams mobilised resources to involve the current authors as 'co-production scholars' throughout the projects to ensure expertise and scientific grounding in academic literature on co-production. Here, the teams *coupled the co-production approach to a scientific frame* of demonstrating scientific quality through academic affiliations and scientific publishing to legitimate and justify the use of a co-production approach, thereby adhering to the organisation's core value of generating scientifically sound research. The case 1 team had involved author 1 and 5 to scientifically ground their co-production approach in academic literature by developing a theoretical framework to guide their work (also see the second restructuring activity). This framework was used as an important instrument to defuse epistemic controversy and developing legitimacy for the approach. Rather than presenting co-production as the opposite of 'regular' impact assessments for accountability purposes (which follow a positivist logic), and siding with responsive evaluation approaches oriented to learning (which follow a social-constructivist logic), the approach was characterised as a pragmatic combination of the two: a hybrid of 'regular' impact assessments and responsive evaluation. Epistemic conflict was

circumvented by arguing that this hybrid would be able to perform both policy research's function of accountability and learning by adopting a reflexive perspective.

A third alignment activity could be observed in case 1's project leader's pursuit of actively sharing his experiences and perspective on the value of co-production, both inside and outside the organisation, to create legitimacy via *building a support network*. He drew attention to co-production's potential for policy impact (related to activity 1) and coupled conventional standards for objectivity and independence to the teams' approaches to address such standards from a co-production perspective (building on activity 2). These actions encouraged familiarisation with and buy-in to co-production within the organisation and its wider policy contexts.

At face value, the first two of these three activities may appear to accommodate to normative structures rather than restructuring them. However, in practice they created space for co-production without having to adapt co-production features. The third activity built on the first two to further expand the space for co-production, thereby increasing legitimacy both inside and outside the organisation. Indeed, over time, we observed normative epistemic restructuring, in the sense that the validity of producing knowledge based on co-production principles became less debated within the organisation. For instance, during the final seminar of case 3 (February 2021), the organisation's chief scientist (responsible for supporting the organisation's scientific quality assurance, and who had recently been newly appointed) presented co-production as an inherently sound approach to policy science: *'Of course, we don't need to discuss the appropriateness of this research approach, for we know by now that is it a valid method.'* This was, in part, also the result of diverse conducive organisational developments to which we turn later.

A fourth alignment activity was directed at alignment within the project team-setting. Each team comprised researchers from various disciplinary and epistemic backgrounds (see **Table 5.1** in case descriptions), each with distinct preferred ways of working, skill-sets, and levels of familiarity with co-production. This contributed to epistemic controversy within the teams on the research conduct and how to derive valid and valuable knowledge (and what this entails). We observed that teams *established boundary objects*, which allowed for interdisciplinary collaboration by functioning as a bridge between various epistemic frames. For example, the case 1 team invested in constructing a shared storyline to inform writing the final report. Each of the researchers shared the results from their respective sub-research projects, and reflected on each other's and their own epistemic frames through the process of working out how the different pieces could fit together. Similarly, case 2 (after realising compartmentalising different parts of the research compromised knowledge

integration, which we discuss in **Section Relational alignment**) and case 3 teams developed shared conceptual frameworks to allow for interdisciplinary collaboration (also see **Section Relational alignment**).

Modifying co-production features to accommodate to normative structures

Some conventional normative structures experienced within the teams and within the organisational context could not be negotiated due to concerns with the level of objectivity of the knowledge produced. To illustrate this, we discuss how this played out in the context of one of the key features of co-production, namely the collaboration with practice experts. This feature challenges PBL's norms of objective and independent knowledge production. To PBL colleagues who were not involved in co-production research, the notion that practice experts would have a say in the development of the research programme's methodology and in the formulation of conclusions diverged too much from these norms and posed a risk to the integrity and credibility of PBL research. Importantly, the desire to demarcate collaboration did not only come from PBL researchers who were not involved in co-production research, but also from practice experts. For them, particularly for policymakers involved with case 1 and 2, it was important that the research results had an unmistakably independent status, so the findings could be dubbed 'evidence-based' and function as solid grounding on which to base policy decisions.

To address this concern, we observed that the teams *demarcated collaboration to particular research phases*, thereby delimiting when and about what practice experts would (not) provide input to the research process. In case 1's first cycle, a clause was included in the research framework that delineated such roles and responsibilities, for example stating that determining research methods and the writing of the final conclusions would be the researchers' prerogative. The research plans of case 2 and 3 also contained such a clause. Indeed, including such a clause in the research plan has become part of the organisation's understanding of how co-production is to be done in a sound and integer manner.

While the institutionalisation of such demarcation of research phases constitutes a modification to accommodate to existing normative structures, we also observed normative structures stretch as projects and experience progressed. For example, the case 1 team experienced that locking out practice experts from discussing methods during its first cycle had resulted in decreased usability for some of them: the computational model that was used only produced findings on a national scale, while a large part of the practice experts operated on a provincial level. To them, these findings were of lesser use (Verwoerd et al., 2020). The case 1 team then realised that

discussing the methods with stakeholders could potentially have prevented such a mismatch, and they did so in the subsequent research cycle.

Table 5.3 Activities undertaken by the project teams directed at relational alignment.

Activities to establish alignment between existing relational structures and collaboration ideals of knowledge co-production, by...	
restructuring of relational structures	● Building interpersonal relations ● Developing common ground ● Establishing managerial endorsement ● Organising interdisciplinary knowledge integration
modifying co-production's relational features	● Demarcating stakeholder diversity ● Loose coupling of co-production and other research activities

Relational alignment

Relational alignment pertains to establishing a fit between existing structures and customs for science-policy interaction and co-production ideals of collaboration. As discussed, collaboration between researchers and practice experts during all research phases is a key feature of co-production. While we discussed the normative structures clashing with this feature in **Section Normative alignment,** relational structures were also less conducive to this aspect of co-production. **Table 5.3** shows the diverse actions the co-production researchers undertook to establish relational alignment.

Restructuring relational structures

All project teams initially focused on organising interactions with the commissioners of the research. Commissioners were also those actors who were the formal owners of the respective policy programmes, either national policy makers or associations who represented regional governments on a national policy level (see **Table 5.1** for an overview of the commissioning actors). However, especially for case 1 and 2, collaboration with commissioners was not self-evident. Despite that co-production was commissioned, some policymakers had concerns that the policy research would be used for national accountability purposes – with potential backlash in terms of budget and task allocation – which made them less cooperative and unwilling to share insight into their policy practices. For example, in case 2 the commissioners at national government initially refrained from sharing their policy dilemmas, which made it difficult for the team to develop a research design that addressed their knowledge needs.

Three activities could be distinguished to further develop prevailing relations with policy actors. First, the teams invested in *building interpersonal relations* to develop mutual trust. To this end, they organised more frequent face-to-face interactions during which they explained the purpose of the research and the importance of gaining the policymakers' input. An open and transparent way of working to inspire trust was actively pursued, and, at least to a certain degree, also achieved: over time, the respective policymakers opened up and more easily shared insight into their practices. For case 3, establishing such open and transparent relations with its commissioners was in general less challenging. The policymakers who commissioned the learning evaluation took their programme's ambition of learning and promoting collaboration to heart. They therefore stimulated openness amongst policy actors involved and were keen to involve the researchers in their efforts from the start.

The second activity we identified was that of *developing common ground*. During the research, the teams ensured they aligned to the practice experts' frames of reference in terms of language, problem framings, research needs and concerns to establish a shared practice and mutual ownership over the research. This could for example be observed in case 1's development of a shared policy theory on the definition and naming of nature policy strategies and their potential for biodiversity levels.

The third activity was directed at *encouraging managerial endorsement* for practice experts to become involved with the co-production projects. We observed project leaders and supervisors organise strategic meetings with top-managerial policy actors to discuss the co-production approach, who later – e.g., during meetings, workshops or presentations – promoted co-production, thereby creating legitimacy and endorsement for their constituencies to become more actively involved. The normative alignment strategy of building a support network appears to be conditional to this relational activity (see **Normative alignment**).

Fourth, as for normative alignment, we observed actions directed at establishing alignment within the project team setting. To facilitate more fruitful interaction amongst researchers with distinct preferred ways of working and skill-sets, the teams actively *organised interdisciplinary knowledge integration* by planning regular meetings during which time was reserved to reflect upon their joint work and identify the overlaps and synergies between the sub-research projects. For instance, the case 1 team planned several writing weeks' during which the team physically got together and collaborated on jointly writing the final research report to establish integration between the project's various themes. Similarly, the case 2 team also allocated time

and resources in pursuit of knowledge integration and organised several reflection sessions in light of this.

Modifying co-production features to accommodate to relational structures

In some situations, relational structures were only (partly) negotiated after they had first been accommodated to. We illustrate this by the teams' search for involving a diversity of stakeholders with their research, another key feature of co-production. Despite the evident multi-level and multi-actor governance contexts of each policy programme, the case 2 and 3 teams especially had difficulty with establishing collaborations with more regional actors beyond the commissioners, due to structural barriers that proved challenging to negotiate, which included unfamiliarity with the regional context, the overwhelming diversity of actors and the lack of previously established relational structures at regional level to draw upon. While we observed actions on developing common ground – i.e., the teams attempted to attune the framing of their research and findings to the regional frames of references during multi-stakeholder workshops, thereby seeking to negotiate relational structures and encourage regional actors' involvement – activities to develop relational structures with regional actors to involve them more structurally with the research, initially remained largely absent. Rather, we observed how the teams *demarcated stakeholder diversity*, for instance by making explicit the research would focus primarily on aligning to the knowledge needs of its commissioners (as for case 1 and 3). Stakeholder diversity in case 1 did broaden over time. As the team built relations with societal organisations through interviews and workshops, and as the team experienced how diverse stakeholder involvement was important for the development of relevant and socially robust knowledge, an advisory board was initiated through which they could give their input on the research scope, findings and conclusions.

A similar process can be observed with regard to the interdisciplinary collaboration within the teams. The investment in interdisciplinary knowledge integration did not fully resolve matters of epistemic misalignment within the teams. . To avoid complex epistemic discussions,, the teams used the strategy of *loose coupling* of sub-project and themes, which allowed for conflicting epistemologies to co-exist within one research project. For example, in case 2, co-production only became the main mode of working in a particular sub-project, while 'regular scientific' knowledge-production approaches were used in the other parts of the project. Similar observations were made for case 1's first cycle. Loose coupling also allowed researchers to default into their standard research routines. However, as noted earlier, interdisciplinary knowledge integration was considered crucial to produce usable knowledge and, as the projects progressed, both teams' members realised that loose coupling prevented knowledge integration. In response, in case 1's second cycle and the case 2 team, additional restructuring

activities of establishing boundary objects for organising interdisciplinary knowledge integration (fourth activity) were undertaken.

Zooming out: the cases in their wider contexts

In this section of the results, we zoom out to appreciate the implementation processes of co-production in the individual cases in conjunction and their wider organisational context.

Different degrees of contextual rigidity and modifications lead to different co-production practices

The alignment activities undertaken by the teams share a common denominator: each was directed at developing legitimacy and buy-in to co-production from the perspective of the predominant normative, relational and epistemic structures experienced by the teams. When legitimacy could not be developed – i.e., contextual structures were too rigid or the teams lacked the capacity to successfully negotiate these –, the teams modified co-production features. Interestingly, the cases show different degrees of experienced structural rigidity, leading to different decisions regarding the modification of features. For example, case 1 experienced highly rigid normative structures within the organisational context (e.g., there was much debate about the independent position of researchers in co-production), leading to actions for normative restructuring as well as the demarcation of stakeholder involvement to certain research phases to accommodate too rigid structures. This was less the case for case 2 and 3: the normative restructuring instigated by case 1 had reduced rigidity (e.g., the debate on researcher independence and other perceived risks of co-production had slightly ceased), paving the way for subsequent co-production projects. Another difference was seen in the rigidity with which relational structures were maintained. Case 1 and 2 experienced of rigidity, leading to additional alignment activities to build relationships as well as a demarcation in stakeholder diversity, than case 3 (where policy actors explicitly requested for co-production and were intrinsically motivated to become involved). In this example as well, case 1 could be observed to have set a precedent as there was partial overlap in practice experts that were involved with case 1 and 3. Their experiences with case 1 had contributed to their enthusiasm for co-production.

The above observations show that different experiences of the degree of contextual rigidity (in different settings) cause diversity in alignment activities and degrees of modification, leading to distinct operationalisations of co-production and thus co-production practices. While outside the scope of this study, it is likely that the success with which intended co-production outcomes are achieved is also dependent on the degrees of rigidity and modifications.

Building co-production capacity

As discussed in **Section 5.2**, according to the literature a certain level of on-site modification when novel practices are implemented in existing contexts is likely to be required for the practice to have workability. However, when a practice gets modified too much its original intent might become lost and the integrity of the practice is compromised. In the cases, we observed this in how key features such as stakeholder collaboration and diversity were only partially satisfied, which compromised the quality standard of socially robust knowledge to some extent. Interestingly, however, we also observed that rigid structures that had initially led to modifications of co-production features became re-modified as experience with co-production increased, and informed renewed (and successful) efforts for interdisciplinary collaboration. For example, while in case 1 the experienced normative structures initially did not allow practice experts to have input into research methods, the realisation of the consequences for the project's usability caused methods to become less off-topic in its following cycle. Similarly, in case 1 the experienced relational structures led to the demarcation of stakeholder diversity to policy actors and the exclusion of societal organisations. During the project's second cycle it was realised that their involvement was crucial for their support and social robustness of the findings, leading to the instalment of a societal advisory board. Finally, we observed how the case 1 and 2 teams became aware that the initial delimiting of co-production to particular sub-projects or themes, due to rigid epistemic structures, compromised the produced knowledge's usability, leading to renewed efforts to normative and relational restructuring.

In all these examples, initial modifications were re-altered over time and the integrity with which co-production was implemented was enhanced. This underscores that restructuring and modifying are interacting modalities that should be regarded dynamically over time and space. Moreover, the examples show how researchers with little initial experience with co-production built co-production capacity as their projects progressed.

The development of co-production capacity was also visible in how researchers whose ways of working were primarily rooted in positivist epistemologies shifted towards co-constructivist epistemic norms during the projects. For example, where some first regarded preliminary research findings as something not to be shared outside PBL, based on concerns with scientific rigour, they started regarding preliminary research findings as suitable material for discussions with policymakers and other stakeholders. Researchers who were new to co-production started realising the limits of their own knowledge production practices and that practice experts, too, had a wide variety of relevant knowledge to bring to the table: *'Our expert knowledge is also corrected by*

regional practical knowledge' (project leader case 2). Co-production capacity was also observed to increase as researchers gradually became more sensitised to policy frames. Frequent discussions on the research project with policymakers allowed researchers to become more familiar with policymakers' mentalities, work preferences and perceptions of the research goals, timeliness and research quality. We observed how the tone of at least some researchers towards policymakers changed from a critical and evidence-oriented perspective ('policymakers should build on a comprehensive engagement with evidence when making policy') towards a more responsive understanding of policymakers' worlds, including the political fields in which they operate and the challenges they encounter (such as the multiple accountabilities policy makers face, towards the programme, government, their minister and the public in general).

The above observations suggest that while awareness of co-production features and the skills to develop conditions conducive to co-production were not always in place from the start, these are developed over time. In such situations, researchers were initially inclined to default into embedded routine ways of working – e.g., focusing on meeting the knowledge demands of the projects' commissioners or delimiting co-production to certain parts of the research – not recognising they were modifying key co-production features. As they became aware of the consequences of these modifications for the aspired co-production outcomes, they learned to better guard co-production integrity by re-modifying certain features. By 'learning by doing' and sharing these experiences across teams, capacity-building for co-production is supported, accompanied by organisational developments conducive to the implementation of co-production practices, to which we turn next.

Conducive organisational developments

In addition to the alignment activities undertaken by the teams, several organisational developments with a likely effect on the experienced rigidity of contextual structures could be identified. To start, several PBL employees were observed to direct efforts into legitimising co-production. They functioned – more or less intentionally – as ambassadors for the approach within the organisation. For example, the supervisor of case 1, and later also case 3, actively shared the success of the co-production projects within the organisation. She refuted counter-arguments to and misunderstandings about the approach, for instance during meetings at managerial level, and committed herself to the embedding of co-production with the organisational research practice. Additionally, in 2018 and 2019 several other key actors came into play. For instance, multiple department managers were appointed with affiliation with deliberative and participatory research. This further reduced normative and epistemic discussion on

co-production within the organisation. Also, a new chief scientist (responsible for the quality control of the organisation's research) with similar affiliations was hired. Under his supervision, an organisation-wide 'vision on quality' was developed in which the organisation's regular quality standards for policy research – e.g., objectivity, independence, legitimacy and relevance – were adapted to also be suitable for co-production approaches. The document presenting this new 'vision on quality' for example provides tools to maintain independence while working close to practice experts, such as extended peer review and researcher triangulation. This development appears to be conducive to normative restructuring by creating legitimacy for co-production via a scientific route (see also the second activity in **Table 5.2**). As such, the new chief scientist played an important role for embedding co-production within the organisation.

Similarly, we witnessed organisational developments that appear conducive for relational restructuring. We observed how researchers who were pioneering with co-production (amongst which the aforementioned supervisors of case 1 and 3), initiated, with support of PBL's Executive board, a Community of Practice (CoP) 'Multi-level and multi-actor Governance' in 2016 to learn about and share experiences with knowledge production with regional policy actors and local stakeholders. PBL researchers and representatives from regional boards participate in the CoP, as do the project teams included as case in our study. This provides them with a platform where they can share and reflect upon the normative and relational challenges and dilemmas they encountered with their peers. The CoP appears to be conducive to both dimensions, although most apparently to relational structures by the engagement of local policy representatives in the CoP.

Parallel to the other developments, epistemic alignment between knowledge co-production and conventional research approaches was encouraged on an organisational level through the introduction of an organisation-wide 'reflexivity frame' (Kunseler & Verwoerd 2019; Kunseler et al. 2020) (Kunseler and Verwoerd, 2019; Kunseler et al., 2020). This frame introduced 'working reflexively' as a concept to denote researchers' awareness and proficiency to tailor their research approach to the characteristics and context of the policy issue at hand, underscoring the relevance and legitimacy of each approach in its own right. By introducing co-production as a valid research approach within a larger repertoire of approaches, its embedding within the organisation was encouraged. High-profile scholars and policy actors were invited to share their expertise and experiences with co-production and reflexivity during internal workshops and lectures, further legitimising the reflexivity frame. Author 1 and 4 additionally supported its enactment by translating the frame's principles to hands-on experiences and guidance (e.g., Kunseler & Verwoerd 2019). Illustrative of

the conducive effect of this reflexivity frame was the turnaround of a department manager who, while previously highly sceptical about the validity and value of reflexivity and co-production, agreed to present how, in his research on the Dutch Climate Accord, reflexive working and features of co-production were adopted to produce knowledge with impact.

By themselves, organisational developments can be understood as part of the emergence of a wider organisational discourse on the changing role of the organisation on the science-policy interface. While one-to-one causal relations between these organisational developments and the experienced structural rigidity are difficult to assess, it is likely that these developments were conducive to the space experienced by the teams for implementing co-production.

5.5 Discussion and conclusion

Despite co-production's increasing popularity, a discrepancy between its theoretical outcomes and those achieved in practice is observed, also referred to as a theory-practice gap (Flinders et al., 2016; Jagannathan et al., 2020). Scholars have pointed towards institutional and political barriers that interfere with the implementation of co-production, causing it to diverge little from their technocratic counterparts in practice despite its radical ambitions of societal transformation and sustainable development (Turnhout et al., 2020; Van der Hel, 2016). Our aim in this paper has been to advance understanding about how the observed theory-practice gap comes into being, by exploring the influence of context on the work policy researchers do to implement co-production. In this section we reflect on the theory-practice gap based on the premise that researchers who seek to practice co-production cannot elude normative and relational structures, often with regard to the epistemological underpinnings of the different knowledge production practices at hand, that occur in the contexts in which co-production is implemented. To gain legitimacy and buy-in to their approach, co-production researchers need to align their practices to conventional structures. Drawing from the work of implementation scientists May and colleagues (2016), we posited that researchers may do so via either attempting contextual restructuring or by accommodating prevailing structures by modifying co-production features.

Our study empirically unpacked and demonstrated how differences in the rigidity of contextual structures and the degree of modifications that are consequently required to implement co-production may play a role in the emergence of different forms of co-production practices and in the way the normalisation processes unfolds over time and between settings. From this follows our first reflection on the theory-practice gap. This gap should not be understood as static, but rather as continuously negotiated

accomplishment, and that co-production's normalisation process, narrowing this gap, can take place incrementally (Borquez et al., 2017; Felt et al., 2016; Hegger & Dieperink, 2014; Van der Steen et al., 2018). We argue that researchers in co-production projects need to learn to recognise and modify both project-level and long-term restructuring dynamics to further increase the space for co-production. When researchers and practice experts exchange experiences with co-production, and reflect on (un)conducive organisational and policy developments that affect the co-production they can enhance the normalisation process. Since we observed the implementation of co-production as a negotiated and dynamic process, we conclude that normalisation does not necessarily imply that co-production is moving towards the 'ideal' co-production standards, but may result in a hybrid approach to satisfy both conventional and co-production structures.

Our second reflection on the theory-practice gap therefore problematises the inclination of scholars to find the ultimate solution in resolving or closing the gap by institutionalising co-production ideals. These institutionalising attempts often fail to acknowledge the level and depth of change required to live up to the promises of these ideals (Van der Hel, 2020). It is therefore more promising to explore ways of 'doing it differently' without striving for ideal-typical applications. Based on our findings we would argue that less-rigorous application of co-production features may be permissible – and even necessary to navigate unconducive structures and to develop legitimacy – as long as integrity is ensured to an acceptable degree and restructuring is initiated in parallel to enhance integrity over time. Instead of problematising the suboptimal outcomes of co-production when key features – such as sustained stakeholder engagement or stakeholder diversity – are compromised (Van Epp & Garside, 2019), our study thus adds nuance to this idea. Our findings showed that researchers developed co-production capacity over time and became more skilled at contextual restructuring and making modifications without compromising co-production integrity. This is pertinent, as it points towards the importance of a learning-by-doing perspective on the normalisation of co-production, which resonates with work of scholars who have underscored the importance of learning by researchers (Fielke et al., 2017) and other actors involved with co-production practices for their successful outcomes (Lang et al., 2012; Norström et al., 2020; Van Epp & Garside, 2019). As it is unlikely that the conditions conducive to co-production are all in place at the outset (Broerse, 1998), we argue that investment in both researchers' co-production capacity and their capacity to recognise unconducive structures is required to develop skilled co-production practice. Our study shows that these capacities can be enhanced by improving knowledge on what the key features

of co-production are and by developing skills to implement these key features, especially when stakeholder participation is concerned.

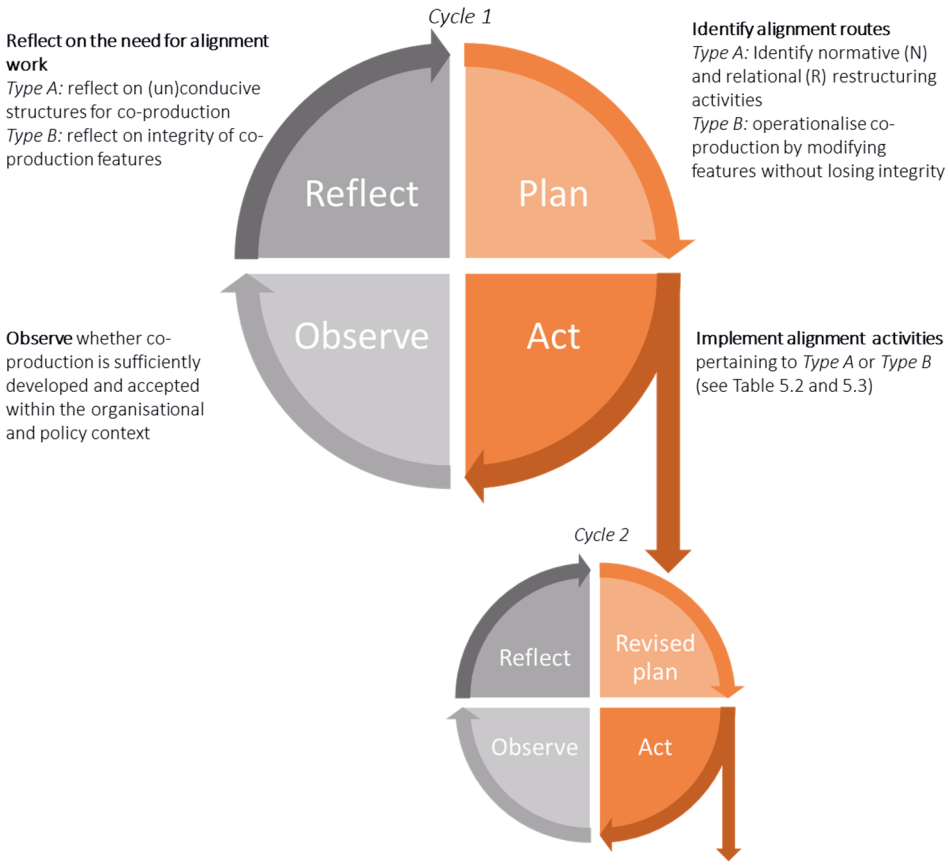

Figure 5.1 Heuristic tool for creating space for knowledge co-production.

In support of capacity-building, we propose a framework to guide reflections and inform . action learning to know when and what modifications are permissible or when more investment in restructuring is required to to make space for their aspired co-production practice (see **Figure 5.1**). The action learning cycle (Kemmis & McTaggart, 2000) may function as a heuristic tool in light of this, guiding reflection and thereby contributing to capacity-building. It consists of four interrelated steps that co-production practitioners can iteratively navigate to align co-production aspiration to the implementation context.

The step 'Observe' asks for awareness of the (epistemic) normative and relational structures in the organisational and policy contexts in which co-production is

implemented, and which affect the space available for co-production in these contexts. Moreover, it asks for awareness of the availability of expertise and capacities to operationalise co-production features. The step 'Reflect' enables co-production practitioners to reflect on the need for alignment work to create space for co-production or deal with the limited availability of expertise to implement co-production features. The type of reflection can be either oriented towards the (un)conducive structures in place (Type A) or towards the integrity of co-production features (Type B). This interaction between these types should be perceived as dynamic in time, and comprehending this interplay requires that co-production practitioners actively reflect on their choices in relation to the context in which they operate. The distinction between Type A and Type B reflection in our framework is thus a mere simplification of this dynamic interplay, and they should take place in tandem. The step 'Plan' can be explicated as a decision-making process in which the alignment routes are identified, in accordance with Type A and B: to reduce the rigidity of contextual structures, normative and relational restructuring are needed (Type A) while modifying co-production features should not compromise integrity (Type B). The 'Act' step is supported by the list of restructuring and modifying activities that we identified in this study (see **Tables 5.1 and 5.2**), knowing that it is non-exhaustive and can be complemented by other scholars and practitioners who practice co-production in their implementation context. We encourage the use of this heuristic tool, as it enables co-production practitioners to build capacities for reflexively implementing the co-production process in alignment with the social-political settings in which co-production occurs.

We further argue that researchers also need to build capacity on how to deal with unconducive structures in the context in which they operate, because it would be naive to simply presume that teams and organisations that invest in researchers' co-production capacity would, by themselves, be successful in overcoming deeply entrenched structures. Such structures may for example include power dynamics and cultural differences at the science-policy interface that compromise co-production's potential (Turnhout et al., 2020). At the same time, in our study, we have identified various conducive organisational developments that, through interactions with the projects, positively affected normalisation. On an organisational level, normalisation can thus be furthered by purposefully organising such interactions, for example by installing communities of practice, in which co-production practitioners may learn from their mutual engagement about how to improve their practice (Wenger, 1998). When researchers learn to reflect on their own practices of alignment they become aware of prevailing routines and norms and in which ways they do (or do not) change as a result of their alignment practices (Nicolini, 2012). Inspired by each other's

reflections researchers may stretch the internalised 'boundaries' of what science-policy work entails, and enhance organisational conduciveness to knowledge co-production.

Although wider contexts were beyond the scope of this study, relevant developments in, for example, the policy arena in which co-production is located may also affect normalisation of co-production practice, since in this study we found that the preferences of policymakers in how to interact with researchers largely inform the potential for normative and relational restructuring to implement co-production features without losing integrity. For researchers aspiring to conduct co-production practices, this means that they do well to explore to what extent policymakers desire collaboration with researchers, to what extent other actors can be engaged and in which research phases. The normalisation of co-production also requires capacity-building and alignment work on the policy side, to enable policymakers to fully engage in co-production and experience the benefits thereof.

To conclude, our study has demonstrated the importance of understanding the organisational and policy context in which co-production is implemented not as a passive backdrop that may constitute sources of interference, but as actively shaping and being shaped by novel practices (Hawe et al., 2009; Squires et al., 2015). Normalising co-production involves probing and improvisation on how to best fit and adapt innovative aspirations to the implementation context, while further enhancement of co-production in policy evaluation projects has to go hand in hand with conducive contextual developments that are outside the direct sphere of influence of the project. We have seen that this is, indeed, taking place, not in the least because the growing popularity of policy programmes on complex multi-level multi-actor issues increasingly demands continuous learning and co-constructive knowledge building.

6 | IMPACT ASSESSMENT IN CONTEXT: EXPLORING THE INTERPLAY BETWEEN POLITICAL DIMENSIONS OF KNOWLEDGE CO-PRODUCTION AND ACTORS' UNDERSTANDING OF IMPACT

Abstract

The previous two chapters reflected upon the experiences of policy researchers with practising knowledge co-production in evaluation, specifically the challenges they encounter and how they seek to address these. From both chapters it became apparent that context – and the actors who constitute this context – plays a significant role in the space policy researchers experience for practising knowledge co-production. In this current chapter, specifically the policy context in which co-production occurs is discussed to address research sub-question 3. The chapter reflects upon the third study of this thesis, based on the impact assessment of the Natuurpact research project's second cycle (2018-2020).

The chapter aims to explore how policy actors and researchers involved with the Natuurpact project understand its impact and value its deliberative and reflexive approach, and the implications of their appraisal for the project's design and execution. Four ideal-typical impact rationales are identified which national and provincial policy actors, societal actors and policy researchers draw on to understand and appraise the project's impact: the accountability, instrumental, network and transformative rationale. At face value, these rationales appear fundamentally incompatible, as they encompass conflicting ideas on the appropriate function of evaluation (e.g., accountability vs. learning), roles of evaluators (e.g., distant experts vs. critical friend) and subsequent quality criteria. In practice, however, actors mobilise multiple rationales interchangeably. Additionally, which rationale is recruited is observed to be influenced by political developments, including nearing nature policy goals deadlines and developments regarding adjacent policy domains (e.g., agriculture and climate change). It is found that such developments simultaneously delimits and strengthens the interest of actors regarding knowledge co-production, as both the accountability and transformative rationale are drawn upon more frequently in the face of these developments. This presents difficulties for policy researchers who seek to adhere to actors' knowledge demands to design impactful knowledge production processes. It is suggested that to address these coexisting impact rationales, there should be more attention for organising constructive conflict between impact rationales during the knowledge production process.

Resubmitted

Brouwers, H.J.H., Verwoerd, L.[10], Loeber, A.M.C., Regeer, B.J., Klaassen, P. *Impact assessment in context: exploring the interplay between political dimensions of knowledge co-production and actors' understandings of impact.* Environmental Science & Policy.

[10] Shared first authorship

6.1 Introduction

'Wicked' sustainability issues characteristically combine high-stake decisions with high technical uncertainty at the system level (Ney & Verweij, 2015). The co-production of knowledge can help address this by ensuring that a wide range of societal discourses are included in discussing potential solutions (Frame & Brown, 2008). In co-production processes, researchers actively engage stakeholders in generating shared knowledge to best inform strategic (policy) action (Arnott et al., 2020b). Ideally, this deliberative process is also reflexive, that is, it challenges system properties that seem self-evident, but are in many ways part of the problem (Voss et al., 2006). With the increasing popularity of deliberative and reflexive knowledge co-production processes comes questions about how to measure their impact and how concerns over their impact relate to their design (Schäfer et al., 2020).

Various scholars have tried to answer these questions. Some have identified generic impact pathways that can inform future endeavours (e.g., Chambers et al., 2021; Frantzeskaki and Kabisch, 2016; O'Connor et al., 2019; Schneider et al., 2019; Wiek et al., 2014). The literature on evaluation and knowledge utilisation is rich in conceptual and empirical work on designing knowledge production processes with impact (James & Jorgensen, 2009; Kirkhart, 2000; Weiss, 1995). Scholars in these fields highlight how the context surrounding a knowledge production process affects how its impact unfolds (Dunn & Laing, 2017; Hansson & Polk, 2018; Tangney, 2017; Tangney & Howes, 2016). Political dimensions in particular affect how and what type of impact is generated because the knowledge needed and the value placed on knowledge production processes and outcomes are subject to (changing) political priorities (Kowalczewska & Behagel, 2019). Indeed, Tangney and Howes (2016, p. 1128) show that the *'political acceptability of science'* is crucial for knowledge to impact policy decision-making.

The literature, however, tends to overlook how actors involved in knowledge co-production appreciate the process and its impact (O'connor et al., 2019). This is striking because this is important for understanding how context influences the way impact emerges and is valued. Such understanding is particularly important for co-production processes, as these involve diverse actors and potentially a myriad understandings of what impact (should) entails. We contribute to fill this omission by empirically investigating how actors' understandings of a knowledge co-production process' impact are influenced by the political contexts in which the process occurs. Furthermore, we study how actors' understanding and valuing of the process and its impact influence co-production process design.

We draw on a Dutch three-year (2018–2020) reflexive evaluation of the national nature policy programme, conducted by the PBL Netherlands Environmental Assessment Agency, a national institute for strategic policy analysis that advises the government on environmental issues and nature policy. The evaluation (Dutch: *Lerende Evaluatie Natuurpact*; LEN) was a knowledge co-production process: national and provincial policymakers, societal organisations and policy researchers collaborated to produce actionable knowledge on the nature policy program's achievements in light of its goals. The LEN took place in a politically-laden context because the nature policy had been decentralised, shifting responsibilities from the national government to the provincial level in 2013. Since then, little progress had been made, while the 2027 deadline for achieving nature policy goals (some of which were set to meet the European Union's biodiversity goals) was drawing near. PBL commissioned three of this article's authors to review the LEN's impact. They set out to assess how the involved actors valued the knowledge co-production process and understood its impact.

Below, we first develop an analytical lens for understanding how actors perceive impact, drawing on literature regarding policy-oriented knowledge production and impact assessment. We then describe the case and our research methods, followed by a discussion of the four 'impact rationales' we constructed from our empirical material, including how these interact and coexist. In section 5, we discuss the coexistence of multiple impact rationales and its implications in light of relevant literature. We also offer considerations for designing knowledge co-production processes that address (and embrace) that coexistence.

6.2 Understanding the impact of knowledge co-production

In the literature on policy-oriented knowledge (co-)production, what counts as 'impact', what types of impact can be observed and how impact can be assessed is highly debated (e.g., Chambers et al., 2021; Loeber et al., 2011; Schneider et al., 2019; Wiek et al., 2014). Interestingly, discussions about impact show a relation between two issues: perceptions of what impact entails determine the types of phenomena studied to assess the effects of a knowledge (co-)production process, and this in turn determines whether the process is deemed a success. In other words, like beauty, impact is in the eye of the beholder. This is especially the case in knowledge *co*-production processes. What their impact is or should be is usually a subject of deliberation. To understand impact of knowledge co-production, researchers must determine whether 'impact' should be conceptualised from the outside observer's perspective (the *etic* point of view; Guba and Lincoln, 1989) or from the involved actors' perspectives (their *emic* point of view; as argued by O'Connor et al., 2019).

Because a co-production process aims to mobilise the epistemic diversity stakeholders bring, it seems pertinent to consider their views in assessing its impact.

This can begin by exploring perspectives on the functions of knowledge in policy decision-making processes. Ever since Weiss (1981) observed that evaluations serve different purposes in policy processes, the number of empirically identified functions has multiplied (e.g., Schoenefeld and Jordan, 2019; Leeuw and Furubo, 2008; Hansen, 2005). Amongst the most obvious is meeting accountability and compliance demands (Nieminen and Hyytinen, 2015; Chouinard, 2013): using evaluations to enable policy actors to take responsibility for their actions and to be held accountable is highly institutionalised, particularly in the New Public Management era (Bovens et al., 2009). Other political functions include improving policies and programmes and legitimising (previously made) policy decisions (e.g., Schoenefeld and Jordan, 2019). Additionally, policy evaluations increasingly function to inform collaborative and reflexive governance (Regeer et al., 2009; Sanderson, 2002) and to help in developing communities or multi-stakeholder networks around particular policy issues (Walter et al., 2007). Hence, the first dimension we consider in making sense of how actors understand impact are their views on the intended function of a knowledge co-production process (*function of knowledge co-production*).

Underlying those views are actors' often implicit assumptions about how societal changes happen, that is, what the 'pathway to impact' is. Various authors have empirically identified and conceptualised a variety of ways via which impact is produced. Schneider et al. (2019) distinguish three (not exclusive) mechanisms through which impact is generated: creating new knowledge leading to better, more informed and equitable policy and decision-making; enhancing social learning and increasing collective action; and developing competencies for reflective leadership. Chambers et al. (2021) build on this to identify two main impact pathways: establishing knowledge co-production processes aimed at producing scientific knowledge, which 'is expected to shape policy and/or practice' (2021:4), and adopting a broader approach to knowledge production in such processes, which may induce social learning and a sharing of experiences. Also, knowledge co-production processes may be seen as complex mechanisms that may advance the reflexivity of the actors involved by increasing their capacity to understand and challenge the societal, political and institutional context in which they operate in (Arkesteijn et al., 2015).The perceived pathway to impact has a reciprocal relationship with how one perceives the function of knowledge co-production. For example, if accountability is seen as an important steering mechanism in policy making – that is, *'what gets measured gets done'* (Van Twist et al., 2015, p. 598) – meeting such accountability demands may be regarded as the primary function of a knowledge production process. Alternatively, if

the function of the process is to inform reflexive governance, the impact pathway is to advance actors' reflexivity. Given the diversity in (implicit) understandings of how knowledge processes can lead to impact, we consider actors' understanding of how impact emerges to be a second relevant dimension for analysing their understandings of impact (*pathway to impact*).

Finally, the third dimension to consider in studying actors' understandings of impact concerns the way actors identify the problem at stake, as it determines the role they see for knowledge production in addressing it. How societal problems are viewed has been evolving. Policy problems can be qualified as 'wicked' (Rittel and Webber, 1973) that is, as dynamic and essentially unique, without a final solution increasingly. A new class of notably (global) environmental problems even qualifies as 'super wicked' (Levin et al. 2012) since time for resolving these is running out, and no single actor has sufficient power or discretion to address them effectively. Some argue solutions involve fundamental revisions of standing practices and associated beliefs, rules and regulations because it is this very constellation that perpetuates their persistent character (Schuitmaker, 2012; Smith & Stirling, 2007). Approaching governance as part of the problem that requires change is arguably the way forward ('reflexive governance', Voss et al., 2006) (*problem identification*).

The three dimensions discussed above are interwoven: how actors variously define the issue (the third dimension) is related to the diversity of views on effective pathways to impact (the second dimension) and the function of knowledge production (the first dimension). One reason for the varied spectrum of views is that demands for accountability and compliance (firmly institutionalised in classic governance modes) have not disappeared with the emergence of reflexive governance and the associated collaborative knowledge production. Some have observed that policy actors have not actually embraced new modes of governance (Nederhand et al. 2019a) . Thus, multiple views on appropriate governance approaches coexist amongst policy and societal actors (Nederhand et al., 2019b).

As stated in the introduction, the political context surrounding the knowledge production process influences the type of impact that is achieved because political developments affect what actors consider usable and acceptable knowledge for policy decisions (e.g., Tangney and Howes, 2016). Kowalczewska and Behagel (2019) add that political circumstances also affect actors' views on the knowledge production process itself. They show that political developments cause policy actors to favour classical modes of knowledge production over deliberative and reflexive ones, as the former are widely assumed to provide independent and objective knowledge that is hard to dispute. They also show that, while policy actors often value knowledge co-

production and expect it to result in usable, action-oriented knowledge, they consider the perceived lack of objectivity as potentially jeopardising the legitimacy of the policy decisions informed by that knowledge. Moreover, if political stakes are high, policy actors tend to prefer knowledge that legitimises and confirms their policy agenda (Flinders and Buller, 2006). In other words, political circumstances may influence what kind of knowledge production process is considered appropriate and how it is expected to affect policy decision-making. This implies that the understanding of what impact is or should be is dynamic in both time and space.

The three dimensions form an analytical lens for exploring how actors in this case study understand the impact of knowledge co-production processes. Using this lens, we look at their views on (i) the function of the knowledge co-production process they consider appropriate, (ii) the pathway to impact they expect and (iii) the problem they identify. We also explore how their understandings of impact are influenced by the political context in which the knowledge co-production process is situated. We now turn to the research methods used in studying this case.

6.3 Methodology

The case and its context

In 2013 the Dutch national and provincial governments signed the *Natuurpact,* a covenant on nature conservation, in which they agreed to collaborate on nature policy goals for 2027, including halting declines in biodiversity (as required by the EU Birds and Habitat Directives) and enhancing societal engagement with nature. With the Natuurpact, the 12 provinces, rather than the national government, became responsible for nature policy formulation and implementation. However, the national government remained accountable to 'Brussels' for the EU biodiversity goals. Additionally, the Natuurpact established that PBL (with partner Wageningen University & Research; WUR) would evaluate progress towards achieving the nature policy's goals in three-year cycles (2014-2017, 2018-2020, and so on). To support constructive interactions and the intended collaboration between national and provincial governments, the evaluation (the LEN) adopted a reflexive and deliberative approach aimed at enhancing policy learning. It is thus an example of a knowledge co-production process.

The second cycle of the LEN (2018-2020) functions as our case of a knowledge co-production process. Its main participants comprised national and provincial policy actors tasked with developing and executing nature policy, and, to a lesser degree, societal organisations (nature organisations and farmer collectives). Participants interacted with the LEN research team in several ways: biannual meetings with the

LEN's formal commissioners (the Ministry of Agriculture, Nature and Food Quality and the Interprovincial Council) to discuss the LEN's progress and findings; bimonthly meetings with a workgroup of provincial policy actors to ensure the relevance of the LEN's scope, process and findings; and large multi-stakeholder workshops to discuss findings and formulate action perspectives.

One of the main conclusions of the LEN's second cycle was that provincial policy plans were insufficient to attain the biodiversity goals in time. While undisputed amongst the participants, this conclusion was politically sensitive because it could be interpreted to mean that the decentralisation process had failed, straining the already tense relations between national and provincial governments. Additionally, at the time, the urgency for nature policy to restore nature quality was further exacerbated by what became known as the 'nitrogen crisis'. Nitrogen deposition norms, for which the provinces are responsible, interfered with the biodiversity goals, for which the national government is accountable to the European Commission. Moreover, elections for the House of Representatives were coming up, increasing the national government's need to demonstrate progress towards nature policy goals. These developments, combined with already strained intergovernmental relationships, led to a tense political context in which the LEN was conducted.

Research methods

We (LV, BR and PK) were involved in the 2018–2020 LEN as external academic experts on reflexive monitoring, evaluation and knowledge co-production, and supported the project's design and execution. As participant-observers, we documented relevant internal and external developments in field notes. We (HB, LV and PK) were also commissioned by PBL to review this LEN's impact on policy.

Data for the study into the LEN's policy impact was collected through document analysis, surveys, semi-structured interviews, focus groups and self-evaluations. The document analysis included the LEN evaluation proposal and final report and meeting minutes (from LEN project team meetings and from meetings between the LEN project team and LEN participants), which we used to map out the knowledge co-production process and contextual (political-administrative) developments. We also analysed national and provincial nature policy plans and parliamentary papers referencing the LEN and scrutinised them for changes in line with the LEN's findings. An exploratory survey was given to over 200 participants (specific group response rates: 100% of the four commissioner representatives, 50% of the 12-member provincial workgroup, 32% of the 16-member societal advisory board; overall response rate, including actors less actively involved: 17%) to gain initial insight into their experiences and views on the LEN and its impact. These insights guided 15 in-

depth semi-structured interviews with national (2) and provincial (7) policy actors and societal organisations (6). The interviewees ranged from LEN commissioners to participants and came from administrative, strategic-managerial and operational levels. Subsequently, two focus groups were held with in total 11 policy actors as participants and the LEN project leaders (3). Because we were focusing on participants' views on impact and on the value of the reflexive approach for achieving impact, we used an action-value attribution framework (Hellström, 2015). Respondents were asked what impacts they perceived the LEN to have had on their (policy) practice, which particular elements of the knowledge co-production process had contributed to this (or not) and what they perceived as the strengths or weaknesses of the process. Finally, the researchers' perspectives were studied using their field notes, a self-evaluation of the LEN research team (guided by LV and HB) and discussions on the scope and character of the next LEN cycle. All data were analysed via content analysis.

6.4 Four rationales on impact of reflexive evaluation

In examining the LEN's impact we identified patterns in actors' views on what the LEN's impact is or should be. From these, we constructed four ideal-typical impact rationales: *accountability, instrumental, network* and *transformative*. As ideal-types, they do not correspond exactly with empirically observable phenomena but instead provide analytical concepts that may be used to inform future theory and research (Shils & Finch, 1949). We named these *rationales* because they are sets of reasons that help explain individuals' sayings and doings vis-a-vis particular questions about knowledge co-production and its impact. In the following subsections, we first elaborate on each rationale and then reflect on their mutual relations. We also discuss our observation that although some actors (and actor groups) generally adhered to one particular rationale, they also frequently employed others. To protect respondents' anonymity, they are identified as R1, R2, and so on.

The accountability rationale

> *The hard evaluation of goal attainment has not been conducted: how much has been realised [referring to the goals], how much money has been spent. The accountability.* (R10)

The first rationale, the *accountability rationale*, was most frequently used by national policy actors and some societal actors.

When discussing the LEN's impact, some national government respondents indicated that they lacked knowledge about the provincial nature policy plans: what steps are the provinces taking to meet the biodiversity goals? They also lacked information

about the differences between the provinces in this regard. Thus, these actors expressed a need for more transparent information on the provinces' strategies and their effectiveness. Those voicing this need tended to emphasise the growing urgency for reaching the Natuurpact goals, referring to the fast-approaching 2027 EU biodiversity deadline and the nitrogen crisis, for both of which the national government is accountable to the EU.

In this rationale, transparency about provinces' efforts in working towards the nature policy goals is key to achieving impact and attaining goals, as this makes it possible to hold the provinces accountable. Clarity in numbers is considered essential to the (political) commitment to act, as one respondent illustrates in reference to the LEN's first cycle: *In the first LEN, the provincial representative at that time said: 'I was shocked by this, it means we need to take a few extra steps.' What she referred to were the numbers [in the LEN]* (R10). Accordingly, respondents see the LEN's primary function as providing transparency into provincial policies to promote accountability.

These respondents also generally critique the LEN's final report for its perceived 'soft' tone. They tend to view its conclusions as *'too friendly'* (e.g., R7 and R11), *'sensitive to policy and politics'* (R10), and sometimes ambiguous, due to the LEN research team's efforts to serve the various actor groups involved (e.g., R11). To some, the LEN has lost its credibility: they find its constructive tone irreconcilable with the continuing decline of biodiversity. A firmer message, they hypothesise, would reduce ambiguity, hold provinces accountable for the nature policy goals, help gain political attention, elicit the required urgency for nature policy issues, and increase attainment of those goals.

The general feeling of a lack of accountability was amplified by the LEN research team's decision (made in accordance with the involved actors) to not include comparisons between the provinces' progress or the cost-effectiveness analyses of particular strategies, both of which are viewed as paramount in this rationale. Such comparisons had so far not been made for political reasons: given the (relatively) recent decentralisation and the ecological differences between provinces, the provinces were not keen on eliciting benchmarks and 'competing'. To respondents embracing this rationale, however, these concerns are not sufficient justification for leaving out cross-provincial comparisons. Regarding cost-effectiveness, methodological constraints play a role: the relatively recent investments and policy implementations combined with the complex and slow changes in nature values make unambiguous measurements of (cost-)effectiveness difficult. Nevertheless, these respondents argue that the LEN should at least give approximations.

The instrumental rationale

> *The great advantage of a reflexive evaluation is that it has become a tool to get better results.* (R2)

The second rationale, the *instrumental rationale*, was used by national and provincial government representatives.

The most distinctive characteristic of this rationale is how the function of the knowledge co-production process in the LEN is understood: the LEN is instrumental in providing insights that can be used for policy improvement. Most respondents embracing this rationale agree that the LEN fulfils this function, both through its final report – *[The recommendations] are an inspiration* (R7) – and through the workshops held throughout the project – *The sessions allowed for a broader perspective and therefore relevant new insights that we can use in our policies. They also motivated us to have these discussions in our organisation* (R11).

This group considers the LEN's impact to be that provinces learn and then use what they learn to improve their nature policy plans. Although they value the organisational discussions resulting from the LEN, they see its real impact in the direct use of its findings in decision-making. As one respondent shared: *The LEN recommendations have almost seamlessly given us the outline for the new Nature Programme* (R2). However, some respondents consider the provinces' low success in attaining goals to indicate not only that the provinces' current strategies are insufficient for reaching the Natuurpact goals by 2027 (as was concluded in the LEN) but also that the LEN itself may be ineffective because it does not provide the substantive resources decision-makers need to redirect current policy plans.

Respondents using this rationale are generally more content with the LEN than those using the accountability rationale. In this view, the knowledge co-production process is valuable for generating impact because the multi-stakeholder workshops provide timely insights in parallel to policy processes. It also allows for better tailoring the research to policy actors' knowledge needs, enhancing the produced knowledge's instrumentality: *I appreciate the recommendations being so close to provincial policy practice. With a regular evaluation, you could never have formulated those recommendations in this way – and make such a relevant report* (R7). However, respondents in this group also question how much value actually emerges from the workshops. Participants generally attend only one or two workshops, leading to large but changing groups. Moreover, the workshop discussions are largely case-specific, making it difficult to transfer insights to one's own organisation. Respondents therefore noted that the LEN

'yields much to those directly involved' (R2) but has trouble expanding its relevance beyond that.

The network rationale

I believe in [an evaluation that is about] the way you work together, the way you want to make it better. (R4)

The third rationale, the *network rationale*, was embraced mostly by the provinces, although societal actors also sometimes used it.

At the heart of this rationale is the conviction that nature policy is a shared responsibility between the provinces, national government and societal actors, although each party may have different roles: *What a lot of work we have to do together, each from their own involvement and responsibility* (R14). Putting the multi-level, multi-actor complexity central in its understanding of nature policy, this rationale regards collaboration and coordination in a network as crucial for obtaining the nature policy goals.

In this rationale, the pathway to impact is understood broadly, but the focus on joint learning is paramount: *If we take a little more time for that [joint consultation and learning] with each other, that's incredibly significant. Then you're not reinventing the wheel separately in twelve provinces* (R14). Impact is perceived to occur through frame reflection via interaction with other actors, improved relationships between actor groups (e.g., provinces saw particular strategic value in improving their relationship with the national government) and increased (societal) support for the LEN's findings and provincial nature policy, all of which are promoted through the knowledge co-production process. Consequently, the LEN should function as a platform through which policy and societal actors engage in joint learning about a range of topics.

In addition, respondents using this rationale speak highly of the 'energy' and 'enthusiasm' that the LEN sparked, especially its workshops. This was frequently the first thing that came to respondents' minds in the interviews. In their view, energy and enthusiasm are important conditions for policy learning and collaboration. Thus, respondents believe the LEN's tone should be constructive and appreciative, so actors experience space for learning: *You should feel free to say what you have experienced* (R1).

However, respondents also thought that the LEN missed the opportunity to stimulate learning about policy coordination in the context of decentralisation, particularly regarding the role of the national government and how the monitoring process (currently different in each province) should be coordinated. Regarding the latter

though, the LEN research team experienced little possibility to address this, as provinces desired to maintain their autonomy in the monitoring process.

The transformative rationale

The connection with other domains – agriculture, urban planning, climate adaptation. (...) I think that's very much needed to achieve the goals. (R7)

The fourth and final rationale, the *transformative rationale*, was especially adhered to by the LEN research team. As R16 noted: *We see a transformation is on-going, and necessary, in nature policy.* A few provincial policy and societal actors also used this rationale.

The transformative rationale builds on the understanding that nature policy is complex and runs into systemic barriers, especially in interaction with other policy domains. This systemic complexity is seen to hamper progress on the nature policy goals: *If you look at the forces for realising [the goals in nature policy], it's working with housing, the energy transition, the large drivers that change the use of space* (R7).

Moreover, users of this rationale see the recent nitrogen crisis as evidence of the need for policy learning on a systemic level because, as the fragile legal basis of the Dutch nitrogen legislation was well-known in policy circles, it did not lead to a changed course of action. Moreover, they see the nitrogen crisis as an illustration of how nature policy is impeded by its context: as building permits are withheld to protect nature areas, nature becomes framed as impeding economic development in the Netherlands, generally causing low(er) political commitment to invest in nature. Thus, in this rationale, linking other policy domains and transforming the context of nature policy is considered essential to improving nature policy.

The LEN's function, according to this rationale, is to enhance the reflexivity of nature policy and its involved actors by facilitating system learning, that is, by increasing awareness of systemic barriers and identifying ways to overcome them. The reflexive character of the LEN is paramount to this rationale. Rather than seeing the LEN as a method for enhancing knowledge uptake (instrumental rationale) or improving relationships (network rationale), respondents using this rationale see the LEN and its reflexivity as an essential tool for guiding policy and societal actors in transforming nature policy and its wider context.

The one adjacent policy field that was incorporated in the LEN (agriculture) and the workshops around themes such as 'nature policy and climate change' and 'the barriers of nature policy' were greatly appreciated by most actors involved. However,

respondents adhering to the transformative rationale felt the LEN should not focus on just one domain or a few workshops but should more strongly incorporate systemic barriers. For instance, various respondents reflected on the rigidity of the European goals (as formulated in the Birds and Habitat Directives) and how it compromises the flexibility they see as required to integrate nature policy goals with the policy objectives of adjacent fields, such as agriculture, climate change and spatial planning. They felt the LEN should have included (more) research on these policy domains and their implications.

Four impact rationales and their interactions

Table 6.1 summarises the four impact rationales by highlighting the three dimensions of how actors understand impact. The rationales are not standalone – they are co-constituted by their relationality: each is intrinsically intertwined with (more or less) explicit judgements about the others, and they all reflect both the positionality of the actors who use them in a given institutional and political context and the dynamics present therein. This subsection explores the relationships and tensions between the rationales.

We found that the accountability and network rationale seemed most at odds with each other. This was especially evident in actors' struggles to understand each other's priorities regarding the LEN's functions. For example, respondents using the network rationale dismissed statements about a too-soft framing of the LEN's findings and instead emphasised that decades of stringent accountability-oriented evaluations had not resulted in achieving the biodiversity goals. Those using the accountability rationale countered that a too-soft tone and a focus on learning would decrease the sense of urgency needed for timely goal attainment. In addition, they claimed that the close collaboration between researchers and policy practice would risk compromising the quality (i.e. objectivity and independence) of the LEN – a risk that ultimately might lead to dwindling political acceptance of its findings: *You can [do co-production] for a while, but at some point [researchers] become part of the [policy] process and can't fulfil the role of evaluator anymore. I expect that parliament will be even less willing to accept that the LEN findings are an objective representation of what is going on* (R13). Respondents' ideas about the appropriate role for policy researchers also conflicted: those using the accountability rationale (and, to a large extent, those using the instrumental rationale) thought an independent, distant expert was required, while

Table 6.1. Schematic representation of how actors understand the impact of knowledge co-production along three dimensions, leading up to four distinct impact rationales.

Rationale / Dimensions	Accountability rationale	Instrumental rationale	Network rationale	Transformative rationale
(i) Perceived function of the knowledge co-production process	Providing clarity in numbers: transparency	Providing instrumental knowledge	Providing a platform for joint learning and deliberation	Advancing the reflexivity of the policy programme and involved actors regarding their relationship to the policy system
(ii) Perceived pathway to impact	Transparency holds provinces accountable, creates learning opportunities and political commitment	Instrumental knowledge leads to enhanced policy effectiveness and efficiency	Social learning improves policy coordination and supports the findings of a knowledge co-production process	Social and system learning support joint actions to overcome systemic barriers and elicit nature policy transformation
(iii) Problem identification	Current (political) transparency and commitment is not sufficient	Current policies are not sufficiently effective	Current policy coordination is not sufficient	Current policy system is not sufficient

those using the network rationale favoured a facilitator of learning processes interacting closely with policy actors.

The accountability rationale was also at odds with the transformative rationale. Actors using the accountability rationale considered including other policy domains to be distracting, and they thought doing so made goal attainment unnecessarily complicated. However, this 'too-narrow' approach frustrated respondents using the transformative rationale: *[It seems that] if it doesn't serve Natura 2000 areas, it's not important* (R7). The transformative rationale also differed from both the accountability and instrumental rationales regarding perceptions of policy researchers' place in the policy process. To those using the transformative rationale, policy researchers were seen as welcome guides who support actors in their collective search for improving nature policy rather than as distant researchers: *In particular, I remember the lead researcher. How he struggled – but with a smile. I think that's super positive, because I don't have the answer either* (R3).

Despite the strict differences in how the LEN's impact is understood within the ideal-type rationales, in real life, the rationales were not as incompatible as they may seem: actors were well aware of the different views on problem identifications and impact pathways, and they appreciated multiple impact rationales. For instance, respondents embracing the accountability rationale, nevertheless valued the impacts considered most important by those using the instrumental and network rationales, such as provincial policy actors learning about nature policy, the constructive discussions and improved (governmental) relationships. Also, many respondents who embraced the network or transformative rationale believed that a stronger framing of the LEN's conclusions would help ensure political commitment, which they also considered important. Moreover, despite their appreciation for networking and learning, some indicated that accountability remains important because it is inherent to policy practice: *I don't find accountability exciting, or scary. I don't see that as reckoning either. It has to be done because it's public money* (R14). Lastly, one respondent using the transformative rationale emphasised the importance of collaboration but also doubted how much it advances goal attainment: *We have to find a meaningful dialogue. Whether that leads to increased goal attainment; I have my doubts about that* (R3).

Although we observed some strong divergences between individual actors in specific actor groups, national policy actors tended to prefer the accountability rationale, whereas provincial policy actors mostly used the network rationale. This difference was in line with their positions in the nature policy system and their subsequent accountabilities. Societal actors did not appear to favour any one rationale, while policy researchers used the transformative rationale most frequently.

Which rationale respondents used was influenced not only by intrinsic motivations but also by external, political circumstances. For example, because of a perceived lack of urgency amongst provinces, especially given the nearing EU biodiversity deadline and the current nitrogen crisis, the accountability rationale had become more prominent. Respondents attributed this to the accountability culture amongst national policy actors. They claimed that national politicians were increasingly seeking to control nature policy progress, despite the decentralisation: *[the House of Representatives] struggles with the fact that it can no longer actively manage nature policy* (R13). This accountability pressure was also felt amongst other respondents: *At some point, we cannot escape saying something about the effectiveness of the policy pursued in recent years* (R2).

Another example involves the transformative rationale, which was increasingly used by actors who considered adjacent policy domains important for nature policy, such as the nitrogen crisis and climate change. They argued that systemic and transformative action was required to achieve the goals in all these domains. This perspective was rooted in the current broader policy discourse on integrative policies in the Netherlands (and beyond). Interestingly, even some respondents who used the accountability rationale were not opposed to integrating adjacent policy domains in the nature policy itself. However, they questioned what, if any, role the LEN should play in integrating them. Indeed, it appeared as if also some respondents using or wanting to use the transformative rationale felt constrained from doing so. For example, several respondents said that despite the perceived importance of system learning, the focus should be kept on the existing nature policy frameworks: *'It is an evaluation of the Natuurpact. That's the policy framework'* (R15). Thus, they felt there was little space to work outside those boundaries.

The prevalence of the accountability rationale has implications for how much discretion policy researchers felt they have to include domains or topics not explicitly recorded in the Natuurpact agreement. The LEN research team generally preferred moving towards a transformative perspective: *We feel system learning is necessary, and would like to approach the next evaluation from a more transformative angle* (R16). Yet they experienced, in preparing the next LEN cycle (2021–2023), that support amongst national and provincial governments for this perspective is not self-evident.

Although respondents criticised certain (lack of) elements of the LEN, they all also recognise its value. Thus, our overall findings suggest that the LEN generally managed to accommodate all four rationales simultaneously.

6.5 Discussion & conclusions

This paper aimed to elucidate how actors' understandings of the impact of knowledge co-production are influenced by the political contexts in which that co-production takes place. In this section, we reflect on the paper's contributions to the literature and assess what our findings suggest for the design of future knowledge co-production processes.

We constructed four ideal-type impact rationales – *accountability, instrumental, network* and *transformative* – that respondents used when appraising the LEN's impacts. Each rationale manifested through different ideas about the function of knowledge co-production, the perceived pathway to impact and the identified problem (see Table 1). The accountability and instrumental rationales were similar in their linear understanding of the impacts of a knowledge co-production process and their technocratic view on science–policy interfaces, whereas the network and transformative rationales both understood impacts as being embedded in a knowledge co-production process and they shared a more deliberative perspective on science–policy interactions. The rationales also reflected diverging ideas about the appropriate functions of knowledge co-production and the roles of policy researchers. For example, according to the accountability and instrumental rationales, policy researchers should be undisputedly impartial, objective and separated from policy practice, all of which are considered at risk during knowledge co-production. In the network and transformative rationale maintaining a strict separation between science and policy is deemed counterproductive because doing so could obstruct the deliberation and collective search seen as required for advancing nature policy. These differences and tensions between rationales compare with those described by researchers distinguishing between modernist and reflexive knowledge production 'imaginaries' or 'logics' (Dahler-Larsen, 2012; Kunseler & Vasileiadou, 2016; Verwoerd et al., 2021), which have been found to guide policy researchers' practices.

While we saw patterns between actor groups and rationales, no single rationale was uniquely associated with a distinct actor group. This corresponds with others' findings that actor type is not necessarily predictive of perspective (Cuppen et al., 2010). Moreover, individuals seldomly adhered strictly to one rationale and how they used rationales appeared to be dynamic over time. For example, usage of the accountability rationale, in which more classical science–policy interfaces and focus on accountability are preferred, increased over time as the urgency of the 2027 deadline increased. This finding confirms the observation that political developments may lessen policy actors' interest in deliberative knowledge co-production processes (Kowalczewska & Behagel, 2019) because more distinctly separated science–policy relations may allow them to avoid political accountability for policy decisions (Flinders & Buller, 2006). However, we also saw the transformative rationale's use increase: the

nitrogen crisis was viewed as proof of the interwoven character of multiple policy domains, implying that system-level action and learning is necessary. Thus, although political developments elicit a response to use the historically more embedded accountability rationale (e.g., Chouinard, 2013), they may also increase the perceived importance of system transformation and thereby increase actors' interest in knowledge co-production.

Our findings have implications for the design and assessment of knowledge co-production processes aimed at influencing policy processes. This regards frameworks that aim to formulate criteria for usable knowledge, such as the CRELE framework (*credibility, relevance* and *legitimacy*; Cash et al., 2002), ACTA (*applicability, comprehensiveness, timing* and *accessibility;* Dunn and Laing, 2017), Weiss' (1995) criteria for usable knowledge (*relevance, conformity, quality, action-oriented* and *challenging*) and Dewulf et al.'s (2020) recently proposed logics of *consequentiality*, *appropriateness* and *meaningfulness*. Some scholars contest the practical use of these frameworks, arguing that the proposed criteria neglect the complex interplay between a project, its context and the surrounding politics (Dunn & Laing, 2017; Hansson & Polk, 2018; Tangney, 2017; Tangney & Howes, 2016). Impact is thus not solely dependent on a process' internal quality. Our study adds to this by underscoring previous arguments that it is largely impossible to fully predict which criteria will be considered important or what knowledge will be considered relevant, and by whom (Huitema & Turnhout, 2009b). Our findings show that policy researchers' face complexity as to what actors deem 'usable' which is rooted in different rationales that change over time due to external political developments and, in turn, affect how they value co-production itself.

How, then, should policy researchers deal with coexisting impact rationales when designing knowledge co-production processes? Given that such reflexive, deliberative processes are rooted in literature on deliberative, reflexive and transformative research, which arguably shows more affinity with the network and transformative rationales, researchers may be tempted to discourage the accountability and instrumental rationales. However, this dismisses the political pressures that are shown, in our findings and previously by Kowalczewska and Behagel (2019) and Flinders and Buller (2006), to encourage the use of accountability and instrumental rationales. We argue that even though using a unified perception of impact may initially seem more helpful, the epistemic roots of knowledge co-production mean that the multiplicity of ways to look at impact should be embraced. This is in line with the literature on dealing with contemporary sustainability challenges, which considers including diverse perspectives important for improving the quality of knowledge for policy

decisions: that is, co-producing knowledge (Funtowicz & Ravetz, 1993; Norström et al., 2020; Pohl, 2008).

Embracing the four impact rationales requires explicit and open conversations with policy researchers and amongst participants about their different needs and expectations regarding the desired impact. These conversations will undoubtedly give rise to new conflicts. However, such explicit conversations are essential to address the recent argument that co-production processes are often devoid of critical reflection on how dominant actors and complex politics shape the co-production process and reinforce such existing inequalities (Turnhout et al., 2020). In addition, scholars such as Cuppen (2012b) and Laws and colleagues (2014) argue that conflicts can be 'constructive'. Conflicts can enhance learning by challenging embedded and implicit assumptions (2012b), and they can increase and sustain actors' engagement with the problem in question (Laws et al., 2014). For a constructive conflict, these scholars emphasise the importance of first identifying and understanding the existing perspectives on the issue. In that sense, the four impact rationales identified in our study provide a useful starting point for conversations on what type of impact a particular knowledge co-production process 'should' generate.

Accepting the four rationales' coexistence also implies that the conversation about the co-production process' desired impact is a continuous one. This suggests that such conversations should focus on learning, and consensus should not be a goal in itself (Cuppen, 2012b). For learning to happen, all actors (both participants and policy researchers) must acknowledge that the four rationales each have their own 'unique contributions' and generate different types of resources and impacts (Brugnach & Ingram, 2012). Learning about and acknowledging each other's contributions may also enhance the 'interpretative flexibility' (Pinch & Bijker, 1984) of a knowledge co-production process. In other words, the conversation about the four impact rationales should be geared towards the question of how a co-production process can better accommodate all four perspectives. Policy actors who already use multiple rationales can aid in this pursuit by acting as 'boundary spanners' between different rationales (Williams, 2002).

In further dealing with the rationales' coexistence we draw from scholars' suggestions for using formative evaluation (Lux et al., 2019; Norström et al., 2020) or reflexive monitoring (Botha et al., 2016; Van Mierlo et al., 2010) to guide co-production processes in progress. This would preferably be *'conducted by an extended peer group comprising experts from both science and practice'* to support iterative learning for adapting the co-production design as policy progresses and needs change (Norström

et al., 2020). We would add that this peer group should represent the multiplicity of impact rationales identified in this study.

7 | THE POTENTIAL OF REFLEXIVE MONITORING IN ACTION TO SUPPORT RESPONSIBLE RESEARCH AND INNOVATION

Abstract

After first attending to policy researchers' challenges with and activities for knowledge co-production in evaluation, and the influence of the contexts in which their projects occur on normalisation, I now turn to a reflection on my own role as reflexive monitor during the Natuurpact research project. This current chapter approaches the potential of this role for supporting the normalisation of co-production from a theoretical point of view (sub-question 4).

In this chapter, the research on reflexive monitoring is positioned in the field of Responsible Research and Innovation (RRI), a co-produced research & innovation (R&I) process that occurs in interaction with a broad range of involved actors. RRI can be seen as an approach aimed at a more democratic, responsible mode of knowledge production that seeks to align R&I processes and outcomes with societal value and needs. It is argued that the developments as regards RRI's conceptualisation and enactment have increased the need to establish whether RRI leads to the assumed outcomes. Looking beyond its aspirations, can RRI's environmental and societal merit be proven? The case is made that for assessing such claims, classical evaluation approaches assume a too-linear view on RRI processes and its outcomes, leading to evaluation findings that are seldomly suitable to feed back into the process to improve its quality. This chapter explores conceptually what the potential of Reflexive Monitoring in Action (RMA) could be for monitoring and assessing RRI. RMA is an interactive and action-oriented methodology for monitoring complex projects that aim to contribute to sustainable system innovation that encourages and supports experimental and reflexive learning. With RMA, an appointed reflexive monitor observes and reflects on the RRI process, promoting RRI's key features of anticipation, openness, diversity and responsiveness by facilitating recurring and systematic reflection on these features. Ultimately, it is argued that RMA could be a relevant methodology for simultaneously promoting and assessing RRI projects.

In the following chapter this is put to the test and the potential of RMA is empirically explored by reflecting upon the merit and intricacies of my role as reflexive monitor during the Natuurpact research project's first cycle (2014-2017).

Published as:
Klaassen, P., Verwoerd, L., Kupper, F., & Regeer, B. J. (2021). *Reflexive Monitoring in Action as a methodology for learning and enacting Responsible Research and Innovation.* In E. Yaghmaei & I. Van de Poel (Eds.), Assessment of Responsible Innovation: methods and practices (pp. 222–243). Routledge.

7.1 Introduction

In recent years, a considerable amount of conceptual and practical work has been done under the umbrella term *Responsible Research and Innovation* (RRI). From its onset, RRI has been framed and developed in both policy arenas and academic spheres, leading to a rich variety of interpretations and approaches (Ribeiro et al., 2016). Certain actors emphasise the framing of RRI as a governance framework, whereas others frame it as actively co-shaping research and innovation (R&I) trajectories in interaction with a broad range of actors involved. In general terms, RRI is seen as an approach to align the values and purposes that drive science and innovation with the values, interests and needs of society in order to steer towards outcomes that are societally beneficial, ethically acceptable and environmentally sustainable. The conceptual work sometimes aims at tracing RRI's roots or localising it in the geography of (like-minded) ideas, movements or disciplinary fields (Blok & Lemmens, 2015; Gianni et al., 2018; Iatridis & Schroeder, 2016; Oftedal, 2014; Rip, 2014; Van Lente et al., 2017). Although sometimes such conceptual work is rather critical of RRI — whether *dismissive* (Zwart et al., 2014) or in the sense of investigating the conditions in which RRI might be truly responsible (Grinbaum & Groves, 2013) — this arguably stimulated the route RRI travelled from being a rather esoteric subject to becoming almost a household name, at least in fields pertaining to the governance and execution of research and innovation. In parallel to the conceptual work undertaken on and in RRI, increasing attempts were also made to *practise* RRI. Much of this work builds on older traditions of thought and action, which can be found in disciplines or fields such as (constructive) technology assessment, (applied) ethics, science communication or gender studies, and the conceptual work RRI borrows from, adds to or relates to this. Indeed, in our view one of the merits of RRI is precisely located in the fact that it brings together valuable conceptual and methodological resources that previously were dispersed across various communities.

With this merit, however, there is also an associated risk — namely, that with the ensuing complexity of RRI it becomes difficult to monitor or assess how people, institutions, companies, fields or the research and innovation system as a whole are influenced by RRI. In this chapter, we argue that Reflexive Monitoring in Action (RMA), an interactive and action-oriented methodology for monitoring complex projects that aim to contribute to sustainable system innovation in the context of 'wicked problems', could well be used to monitor and assess RRI. As such, our approach resonates with the 2015 report by the European Commission's (EC) Expert Group on Policy Indicators for Responsible Innovation, which emphasised the importance of experimental and reflexive learning in assessing RRI (Roger et al., 2015).

7.2 Background

As we see it, the emergence of RRI as a governance framework for research and innovation makes immediate sense in light of two trends. First, research and innovation play an increasingly important role in society (Acs et al., 2013). That is to say, in today's knowledge economy, a growing number of people are working in research and innovation (Pyöriä, 2005), and the fruits of such work become ever more deeply ingrained in all aspects of our lives. A large part of research efforts concern the production of so-called 'pure' or 'fundamental' knowledge with, at its core, norms of objectivity and independence. It is, however, equally true that research and innovation are often driven by values of applicability and making a positive societal, environmental or economic impact — working towards improvements for, as it is often put, people, planet and profit. This holds true also for research that is *not* primarily or purposively done with an eye to application or improvement, as this too can unexpectedly find its way to the market — think, for instance, of Alexander Fleming's serendipitous discovery of penicillin (Bosenman, 1988).

In parallel to the expansion of the role of scientific knowledge in developing, for instance, technical solutions, health interventions or evidence-based policy-making, so too an increase can be found in the realisation that research and innovation can have unintended and sometimes detrimental consequences (Swierstra & Waelbers, 2012). For example, X-rays can be used for diagnostic purposes, but also cause cancer; combustion engines can be used in vehicles to transport people and goods rapidly and cheaply, but at the cost of a negative impact on both the environment and human health; smartphones help connect people to each other and to the virtually endless resources the internet has on offer, but bring with them possible threats to individuals' privacy and risks for mental health (Kawabe et al., 2016).

Arguably, this combination of trends calls for investments to ensure that R&I are imbued with societal values. Insofar as the general public is affected by the (in)direct consequences of research and innovation, it arguably should have a say in how the public interests should be respected (Dewey, 1927). In societies characterised by a plurality of values, the challenge is to incorporate the values of different stakeholders in the development of R&I, so as to work towards ethically acceptable, environmentally sustainable and socially desirable outcomes, products and impacts. This challenge is far from straightforward due to the often fundamental differences between stakeholders in their interests and values and therefore their framings of both problems and solutions (Blok & Lemmens, 2015). RRI provides a framework for the governance of research and innovation that helps facilitate the development of practices in which this is addressed.

Then, what does that framework consist of? This question has received various answers. One important source in the field, the European Commission (EC), states that RRI is "an inclusive approach [that] aims to better align both the process and outcomes of [research and innovation], with the values, needs and expectations of European society".[11] In the UK, the EPSRC has instituted a framework for responsible innovation built around the process elements of anticipation, reflection, engagement and action (AREA[12]), and along similar lines, Wickson and Carew proposed the following four core constituents of RRI: "(1) a focus on addressing significant socio-ecological needs and challenges; (2) a commitment to actively engaging a range of stakeholders for the purpose of substantively better decision-making and mutual learning; (3) a dedicated attempt to anticipate potential problems, assess available alternatives and reflect on underlying values, assumptions and beliefs; and (4) a willingness amongst all participants to act and adapt according to these ideas" (2014, p. 255). In this chapter, we propose a heuristic and analytical model of RRI that shows clear affinity with such understandings, with an additional specification of more fine-grained criteria and a methodology that can be used in applying these criteria in enacting RRI. The combination of criteria and methodology can help bring together the simultaneous practice and assessment of the rich, yet fairly abstract, prevailing understanding of RRI.

7.3 Formulating criteria for RRI — and putting them to good use

With developments towards the conceptualisation and enactment of RRI, there is a need to establish whether RRI leads to the assumed outcomes. Looking beyond its good intentions, can the environmental and societal merit of RRI be proven? If so, can these proven outcomes be linked to RRI processes (i.e., can we draw lessons on what determines 'quality' in RRI practice)? In other words, can we validate the value of RRI? To help answer these questions, there have been several attempts to further specify what it entails to engage in RRI (sometimes dedicated to RRI in a specific context), and to assess RRI's impact (Davis & Laas, 2014; Heras & Ruiz-Mallén, 2017; Kupper et al., 2015; Peter et al., 2018; Wickson & Carew, 2014). In the context of the EC-funded project, RRI Tools, two of the authors (A1 and A3) contributed to the formulation of criteria for RRI.

[11] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/science-and-society
[12] See https://epsrc.ukri.org/research/framework/area/

This attempt commenced with the deceptively simple question: 'What is RRI?'. Deceptively simple, because we wanted to avoid pure theory-based conceptualisations, and to steer clear of the implicit suggestion that it would be possible to 'discover' a single fixed and context-independent answer to this question. Thus, even if hypothetically it would be possible to define rules for behaving responsibly in research and innovation, we would still see these rules in themselves as being of little or no value. In our view, such a 'sterile' theory-driven code-book would be self-contradictory, as it would absolve those applying the model from the very kind of responsibility it seeks to instil. Rather, the burden of deciding what are responsible decisions and actions always remains with those directly or indirectly taking part in research and innovation processes. These actors themselves need to reflect upon what is responsible in the specific contexts of their work, and this requires careful and systematic reflection on a case-by-case basis. To give expression to this view, the criteria that emerged as our answer to the question 'What is RRI?' were eventually translated into questions inviting reflection and deliberation, rather than assertions.

To arrive at a comprehensive model of RRI and its criteria, we engaged in a process of iterative conceptual modelling (see Figure 1, and see Klaassen et al. 2017 for a more extensive description). Central to this methodology for concept development are different and disparate forms of expertise, confronted in a series of iterative steps which, in this case, sought to answer our question "What is RRI?". Figure 1 presents the six different steps of (1) literature review on RRI and the many disciplinary and conceptual resources RRI is built up upon; (2) expert consultation; (3) stakeholder workshops; (4) cataloguing promising RRI practices; (5) deriving RRI criteria; and (6) in-depth case-study analyses, as well as the iterations to arrive at and subsequently test and adapt our conceptualisation of RRI and its associated criteria.

The starting point for any specification of ideals is to ask *why* the ideals are needed. Specifically, why would responsibility need to be introduced in the research and innovation environment? Or, to paraphrase a famous super-villain: *Why so responsible?* The answers to this question can be linked to two perspectives from which the need for responsibility can be approached. First, as seen by those *inside* the research and innovation environment, responsibility is an obligation that follows from the (social, financial, and moral) power granted to those in a position to carry out or regulate research and innovation processes. Writing from this perspective, Owen et al. (2013) refer to responsibility as an 'imperative'. Second, from the perspective of those actors traditionally located *outside* research and innovation, gaining responsibility is a *right*, derived from the breadth of impacts they (may) experience as result of those processes. This second view asserts that scientists should not be isolated from society for the rather straightforward reason that the impacts of their work are not.

In a European context, these ideas have been integrated in research policy, and resonate with what are known as the seven grand social challenges (Klaassen et al., 2018; Lund Declaration, 2009) — although it is now more usual to refer to the Sustainable Development Goals (SDGs) adopted by the United Nations. As we interpret them, the grand challenges or SDGs are not so much research topics, but rather points of intersection between science and society (Cummings et al., 2018). Indeed, these challenges can *also* function as the abstract, but palpable, purposes for responsible research and innovation.[13]

Taking challenges or the SDGs as long-term purposes, in order to achieve them, we need to think of more concrete products that can function as intermediate stepping stones on our journey. Through our discussions with experts and laypersons alike, we have discovered that, in pursuing responsibility, the associated learning is as important as the results of R&I. We thus make a distinction between R&I products (i.e., the artefacts created through research and innovation) and learning products (i.e., the skills acquired and practices established in creating R&I products). Responsibility, we maintain, is to be equally sought in both types of products: it is simultaneously a dimension of the products of R&I, and of the hard and soft institutions thereby brought to life. If solving the seven grand challenges can be compared to crossing a river, then R&I products are stepping stones and learning products are the skills needed to discover and jump from one stepping stone to the other. If we have the former and not the latter, we might end up stranded in the middle of the river with little idea of what to do next. In brief, if responsibility is to be 'fostered', the cultural by-products of R&I activities are as important as the activities themselves.

To take an example: if transgenic crops are an R&I-driven solution supporting the second grand challenge (food security, sustainable agriculture and forestry), then the practices institutionalised in the process of researching transgenic crops responsibly are as important as the actual crops. The two might affect technological progress in different timespans, R&I products having a direct and immediate impact while learning products have an indirect and a delayed one — but the two are of equal importance. The questions are 'Do GMOs constitute a step towards the second grand challenge?', and 'How can research into GMOs be carried out responsibly and how can we instil this responsibility in the future?'.

[13] Perhaps needless to add: such purposes will always remain up for discussion, and their concrete interpretation should be left to those taking responsibility for pursuing one purpose or the other.

Having delineated the destinations of RRI, we can reason backwards to clarify the processes needed to reach them. The conceptualisation of RRI in terms of four process dimensions was arrived at, much in line with much-cited sources on RRI such as Owen et al., (2012) and Stilgoe et al. (2013). Thus, *Anticipation and reflection*, *Openness and transparency*, *Diversity and inclusion*, and *Responsiveness and adaptive change* were identified as core elements in the process of doing research and innovation responsibly. These dimensions were then further specified in terms of criteria and sub-criteria, and ultimately one or more inviting questions. To emphasise: we refer to these questions as *inviting questions* to highlight their intended status as points of departure for various forms of reflection and deliberation, rather than as instrument of accounting. To illustrate what this has reaped, Table 7.2 presents criteria, sub-criteria and inviting questions for the process dimension of Anticipation and reflection (and for the other three dimensions, see Kupper et al. (2015)).

Table 7.2 Criteria, sub-criteria and inviting questions for the RRI process dimension of Reflection and anticipation (adapted from Kupper et al. 2015).

Criteria	Specification	
	Sub-criteria	*Inviting questions*
Analysis of the background, current situation and context of the (planned) research or innovation. (Nordmann, 2014)	Up-to-date information	*Has content research been done on relevant background knowledge and up-to-date information?*
	Influence other R&I	*Has the influence of other innovations/research on the course of this practice been taken into consideration (e.g., alternative R&I and complementary R&I)?*
	Actor analysis	*Did an actor analysis take place, identifying on whom the practice might have an impact or who might have an interest in, and might have relevant expertise for the practice — and identifying how these actors relate to each other?*
	Diverging problem definitions	*Have efforts been made to address potentially diverging definitions of the problem at stake in the practice?*
	Societal role in problem definition and course of practice	*Have efforts been made to give a role to societal values, perceptions and interests in defining the problem addressed in the practice and the further course of the practice?*
Envisioning of plausible futures (Nordmann, 2014)	Variety of future parameters and impacts	*Is there active identification and consideration of immediate, mid-term and are long-term social, environmental and economic impacts and consequences of the practice — intended and unintended — identified?*
	Variety of established methods	*Did a well-considered selection and implementation of the methods for anticipation take place (based on previous experience)? (e.g., scenario development, real-time technology assessment)*
	Variety of R&I trajectories	*Have alternative R&I trajectories been considered? (process of R&I)*

Variety of impacts	Ethics	*Are ethical aspects and impacts of the practice sufficiently addressed? (e.g., are research ethics honoured, by protecting the subjects of research, approval from an ethical committee and documented compliance with research ethics and voluntary codes of conduct — in which for example fraud and plagiarism are prohibited? (Wickson & Carew, 2014)*
	Legislation	*Are legal aspects and impacts of the practice sufficiently addressed? (e.g., is there documented compliance with highest-level governance requirements; (Wickson & Carew, 2014)*
	Society	*Are social aspects and impacts of the practice sufficiently addressed?*
	Environment	*Are environmental aspects and impacts of the practice sufficiently addressed?*
	Grand Challenges	*Are one or more of the Grand Challenges set by the European Commission addressed in the practice?*
Facilitating deliberation on values, perceptions, needs, interests, choices and definition of the problem at issue in the practice	Integrated reflection and deliberation	*Has room for reflection and deliberation on e.g., impacts, alternatives, possibly changing social values, perceptions and needs/ interests and choices made during the practice, been built in? (Stilgoe et al., 2013)*
	Deliberating values	*Do the actors involved regularly engage in a critical analysis of the values, perceptions, needs, interests, choices and definition of the problem at issue underlying their practice?*
Addressing roles in RI trajectories	Awareness of differences	*Do the actors involved develop an awareness of their own assumptions, values and purposes in relation to the perspectives of others?*
	Awareness of responsibilities	*Are actors involved aware of and open to reflection on their role responsibilities and accountability? (Stilgoe et al., 2013)*

7.4 Approaches to assessing RRI

Criteria for RRI can be used beforehand both by researchers and innovators in designing R&I responsibly or by funding organisations for assessing proposals, as well as afterwards for assessing to what degree R&I have been undertaken responsibly, and whether desired outcomes and impacts have been achieved. Such criteria can also be used during the execution of R&I projects or programmes, as a means to monitor and continuously improve them. Arguably, the latter constitutes a way of assessing (attempts at) RRI that is most in line with the earlier description of RRI, especially in relation to RRI's intrinsic openness and connection with experimental learning. A complex new field of its own might appear to open up with this entrance into the realm of evaluation of R&I, but we in fact enter a field in which there is significantly more scholarly experience than in the world of RRI alone.

Depending on the prevailing needs and intentions, evaluators of R&I projects or programmes can choose a method from a wide variety of available (theoretical) approaches. Traditionally, evaluations function as a type of disciplined inquiry to assess whether a programme or project has contributed to its designated purposes. From an instrumental perspective, this approach primarily serves the needs of project funders in terms of accountability (also understood as upwards accountability; Ebrahim (2005). Especially, for instance, large-scale innovation and research projects that depend on private-sector investments are prone to evaluation focused on accountability based on criteria set by investors. Such criteria tend to spring from a technical–economic perspective, narrowing down the set of interpretations of a project's success to a matter of pre-defined deliverables. Academic evaluation based on output criteria is also a familiar practice, for instance through quantitative metrics meant to capture the (academic) impact of publications. For example, the Centre for Science and Technology Studies (CSTS) at Leiden University is invested in the development of bibliometrics, and evaluates academic impact through the analysis of citation patterns of different research groups. Through this approach they also attempt to determine the use and value of publications for marketable applications, public policy development or other ways of attaining public value. However, although they offer some insights, the metrics produced have limited use in informing action to improve research practices. This is a known argument against evaluation approaches that focus solely on accountability. Although a legitimate approach, it tends to assume rather linear input–output relations, and the evaluation findings are seldom suitable to feed back into the practice of those actors who implemented the project to help them improve their work. The recognition that *learning* by those actors is required to improve a project's conduct and increase the likelihood of achieving the designated purposes, contributed to the emergence of so-called 'learning-oriented' evaluation approaches. In such approaches, evaluations are designed to focus on the understandings, concerns and learning needs of actors and stakeholders involved in or surrounding the project or programme in question.[14] Arguably, this type of perspective also inspired the 'Leiden manifesto for research metrics',[15] which proclaims that the scope of research metrics should be extended beyond output criteria alone, by developing more 'meaningful metrics' in order to inform research performance (Hicks et al., 2015). These two functions of evaluation (i.e., accountability

[14] This type of evaluation can be understood as part of the 'fourth-generation evaluation' paradigm proposed by Guba & Lincoln (1989).
[15] See http://www.leidenmanifesto.org.

and learning) and their supposed irreconcilability have received considerable attention in the recent scholarly literature in evaluation studies (Adelle & Weiland, 2012; Guijt, 2010; Kunseler & Vasileiadou, 2016; Owens et al., 2004; Regeer et al., 2016).

For, while learning through evaluation is generally professed to be important, evaluation for accountability still predominates. This is especially problematic for long-term R&I projects that search to address complex societal problems that are of a more goal-seeking and emergent nature, and for which the societal impacts and technological advancements weigh heavily but are difficult to predict. Such research projects very rarely take the shape of a linear process of problem formulation, project design and implementation, as they tend to unfold as dynamic and experimental interactions between multiple actors from multiple levels or domains — academic, industrial, societal, policy — during which mutual learning is essential in order to produce the types of responsible outcomes that are sought (i.e., outcomes contributing to the SDGs or grand challenges, that are for instance ethically and socially acceptable and environmentally sustainable) (Regeer et al. 2009). As such projects progress, greater insights and developments in their institutional and societal context may require the reformulation of the project's goals and, subsequently, the research design. Adaptive capacity is required for the project team to be able to anticipate and accommodate such changes. From this perspective, evaluating for accountability based on pre-defined criteria becomes meaningless. Simultaneously, evaluation that focuses solely on learning without considering the developments in institutional and societal contexts might improve specific procedures, but would lose sight of a project's greater ambitions of achieving societal impact. For complex R&I projects, there is a call for evaluation approaches that reconcile the dual purpose of accountability and learning (Lehtonen, 2014; Regeer et al., 2016). More specifically, there is a call for evaluation approaches that allow to assess R&I projects' progress towards their greater ambitions, *while* learning — about goals, strategies, actions and contexts — is promoted (Klerkx et al., 2010). By creating space for learning processes, the project team and other stakeholders may build their adaptive capacities and abilities to inform adaptive change, in light of new insights and a changing environment. The evaluation approach that is adopted should be as flexible as the project itself.

From different fields — including development studies, system innovation, transdisciplinary research, transition management and RRI — scholars have argued that the answer lies in evaluation approaches geared towards enhancing *reflexivity* (Arkesteijn et al. 2015; Botha et al. 2016). Reflexivity is a capacity that allows projects' participants to act in greater accordance with and are responsive to their institutional and societal context, in line with their understanding(s) of a project's success (Elzen et

al., 2017). Indeed, reflexive approaches seek to encourage the participants' adaptive capacities. Reflexive evaluation occurs during a project's course and is ideally integral to it, as part of iterative cycles of defining, implementing and adjusting its design in order to move towards responsible research outcomes (Regeer et al. 2009). As such, reflexive evaluation methods simultaneously function as promotors for and assessors of learning processes to achieve environmentally sustainable and socially acceptable goals. **Table 7.2** shows the three conceptualisations of evaluation approaches and their characteristics we have discussed here.

Table 7.2 Three ideal–typical conceptualisations of evaluation approaches (adapted from Kunseler & Vasileiadou (2016))

	Evaluation approaches for accountability	Learning-oriented approaches	Reflexive evaluation approaches
Evaluation purpose	Evaluation as mechanism for accountability assurance, assessing a project's performance	Evaluation as mechanism to enhance learning processes for improved (research) practice (procedures, processes)	Evaluation as mechanism to enhance reflexivity through monitoring of and reflection on goals, strategies, actions, institutional and social contexts
Characteristics of evaluand	Simple project: mostly linear causal chain of inputs, outputs and outcomes	Complicated project: multi-actor interactions with complicated multi-directional causal mechanisms	Complex project: multi-actor, multi-level governance constellations with unclear/unknowable and uncertain causal mechanisms
Evaluation criteria	Pre-defined deliverables (technical–economic; (cost)efficiency, effectiveness)	Learning needs of project team in light of their own practice	Learning needs in light of jointly determined goals and principles (e.g., ethics, inclusion, responsiveness)
Evaluation design	Technical–analytical evaluation design by evaluators	Deliberative evaluation design in concert between evaluators and project team	Deliberative–analytical, emergent and responsive evaluation design in concert between evaluators, project team and other relevant stakeholders
Role of evaluator	Distant observer, judge	Facilitator of learning processes	Facilitator, mediator, observer: 'critical friend'

Time aspect	Evaluation takes part prior or after the project	Evaluation occurs parallel to the project	Evaluation is integral part of the project

7.5 The potential of Reflexive Monitoring in Action

There is one specific reflexive approach that has proliferated in various domains and that we especially consider to have potential for promoting and assessing RRI projects: Reflexive Monitoring in Action. RMA is an interactive, action-oriented monitoring and evaluation (M&E) method, originally developed to support projects with ambitions to make system innovations and which require major institutional and social change (Van Mierlo et al. 2010). As for most reflexive approaches, RMA is intended to increase a project's reflexivity — its ability to affect and interact with the context within which it operates — by encouraging its participants' collective learning processes, through which institutional and societal barriers to system innovation are identified and overcome. As Arkesteijn et al. (2015) explain, RMA builds on the premise that, while the contribution of a single system innovation project to the overarching, complex system innovation processes cannot be assessed, it is possible to understand and work (and document!) towards a project's design and outcomes in terms of relevance to long-term ambitions of system innovation. Key to the methodology is recurrent reflection on the institutional and societal context of a project in relation to its long-term ambitions, and its concrete project design and effects (i.e., on the Theory of Change on which a project's design is founded). The actual monitoring activities of RMA are integral to the project, and are usually guided by an appointed reflexive monitor. The monitor observes moments of interaction (like team meetings) and focuses on whether learning and actions towards the intended purposes take place, assisting such processes by making use of a range of reflexive monitoring tools activities (see Van Mierlo et al. 2010). As such, the reflexive monitor is not merely an observer, but rather a facilitator or 'critical friend'; close enough to fully understand the issues encountered, but with sufficient distance to legitimately and critically reflect on the process. What it is that needs to be learned, or by whom, is not a given in this method, but rather something that itself needs to be recurrently assessed and reviewed. RMA does not provide a blueprint or a strictly structured action plan, but should rather be seen as a flexible methodology that can be customised to support any endeavour that aspires to contribute to sustainable development and societal change.

We argue that this holds merit for the evaluation of RRI. While a project's societal impact cannot be foreseen or anticipated in full due to internal and external

developments and contingencies, what *is* possible is to guide action along the RRI criteria framework as described above and support the coordinating team with dealing with unexpected happenings. RMA methodology may help to enhance the reflexivity of a project team and build its capacity to enact RRI criteria. As a result, one has worked towards responsible R&I outcomes, while simultaneously the process towards these becomes more responsible.

7.6 Discussion and conclusion

Recent years have seen much conceptual and practical work geared towards opening up research to society and aimed at redirecting science and innovation towards offering societally beneficial, ethically acceptable and environmentally sustainable outcomes. Under the label of RRI, an increasingly comprehensive framework is emerging to guide the governance and execution of responsible R&I. In this chapter we have shared our understanding of what such a framework should look like. Comprised of process dimensions that can be operationalised in the form criteria, sub-criteria and inviting questions, the framework is rooted in a philosophy that simultaneously underlines the importance of producing responsible R&I outcomes, as well as of a responsible research process towards these. In our view, the framework is most useful when adopted by actors who take part in R&I processes to carefully and systematically reflect on what *responsible* means in their given situation. In doing so, we emphasise, researchers and innovators learn to act in a more responsible manner. We have made the case for the potential role that a specific methodology for reflexive monitoring, RMA, can play to promote such learning. At the heart of this is RMA methodology's encouragement of the responsibility of actors involved with R&I, through seeking to increase these actors' reflexivity.

Most experience with RMA so far has primarily been geared towards supporting action-oriented, transdisciplinary projects that aspire to contribute to system innovations. In these contexts, RMA has become a proven methodology for reflexive evaluation that promotes learning and reflexivity through precisely those values that are also central in the process dimensions of RRI: reflection, anticipation, openness and responsiveness. It will be interesting to see whether RMA can make a similarly significant contribution to learning for responsibility in the context of other types of research and innovation. The outlooks for this are bright, given the close alignment of RMA with general features of RRI and the potential to use it simultaneously to assess the degree to which R&I projects or programmes can be seen as RRI, to contribute to learning in or on RRI, and, hence, to promote RRI in these projects or programmes.

8 | COMBINING ROLES OF FACILITATOR AND ASSESSOR TO SUPPORT THE QUALITY AND IMPACT OF TRANSDISCIPLINARY RESEARCH

Abstract

This chapter builds on the previous chapter in that it now *empirically* explores the potential of Reflexive Monitoring in Action (RMA) for simultaneously promoting and assessing knowledge co-production (sub-question 4). Instead of positioned in the literature on Responsible Research and Innovation (RRI) (as the preceding chapter), this chapter is located in the field of transdisciplinary research. As for RRI, in this field a knowledge gap is identified as regards the assessment of transdisciplinary research's societal impacts and the quality of its process. To address this gap, this chapter reflects upon our – myself and colleagues from the Athena Institute – roles during the Natuurpact research project's first cycle (2014-2017) (reflecting on the fourth and final study in this thesis).

During the project, we were involved as reflexive monitors and, as such, were tasked with supporting the project team's learning processes as regards the selected knowledge co-production approach to advance its quality, and with assessing the project's impact. We thus involved both as facilitators of learning processes, as impact assessors. With the assessment, there was specific attention for reviewing the value of the selected knowledge co-production approach. Combining the roles of facilitators and assessors is shown to allow for better access to the policy researchers and policy actors who had participated in the project. Furthermore, taking up this combination of roles enhanced our sensitivity to the perspectives of policy researchers and policy actors on the project's impact. It also increased our understanding of complex internal and external project dynamics, as well as of how these dynamics shaped the project's outcomes. This resulted in a meaningful assessment of the Natuurpact research project's policy impacts and enabled us to attribute these to specific elements of the co-production process. Additionally, the approach supported the project team's capacities for developing co-production's key features, albeit less so than was originally expected. Four capacities are identified: *building co-production ownership, openness and transparency for integrating divergent knowledge needs, purposeful responsiveness* and *navigating institutional realities and co-production ambitions.* It is concluded that combining tasks of facilitating and assessing as reflexive monitor works complementary and is recommendable to accommodate transdisciplinary research's inherent contingencies, but requires a balancing act between these tasks.

This chapter presents the final empirical chapter of this thesis. Next is **Chapter 9**, in which I will discuss the findings of the current and previous chapters to answer the research sub-questions, and, finally, the main research question. This is followed by a general discussion in which I position the conclusions in wider academic and societal discussions, a reflection on the research approach and ways forward for both research and practice.

Published as:
Verwoerd, L., Klaassen, P., Van Veen, S. C., De Wildt-Liesveld, R., & Regeer, B. J. (2020). Combining the roles of evaluator and facilitator: Assessing societal impacts of transdisciplinary research while building capacities to improve its quality. Environmental Science & Policy, 103, 32-40.

8.1 Introduction

Accompanying today's many complex and critical societal issues – such as climate change, depletion of natural resources, global food security – is a rise in demand for transdisciplinary research (TDR). In TDR, academics and societal actors collaborate to integrate knowledge and develop socially robust answers to real-world issues (Pohl, 2011). It is assumed that TDR is better equipped to contribute to solving the complex problems facing society than mono- or interdisciplinary research (Hansson & Polk, 2018). The growing practice of TDR makes it more pressing to look beyond its intentions, and to ask how its impacts can be shown and its quality assessed, and ultimately tied to these impacts.

Four key TDR features include participation of relevant stakeholders, knowledge integration, responsive and emergent designs that allow for the research to develop as insights unfold, and managing boundary dynamics (Hoffmann et al., 2017; Regeer & Bunders, 2009; Scholz & Steiner, 2015a). These features, however, also make it particularly challenging to evaluate TDR (Walter et al., 2007). As TDR, by definition, crosses disciplinary boundaries and aims to achieve impacts beyond the scientific realm, conventional academic quality or impact criteria are inadequate (Belcher et al., 2016). Even if standardised criteria were readily available, they would have to be sufficiently flexible to accommodate every project's contextual specificity (Carew & Wickson, 2010). Furthermore, rather than a linear process of formulating a problem, research, and achieving impact, TDR is more like iterative and experimental interactions between actors from different domains (Regeer et al., 2009). TDR's many contingent internal and external project dynamics are hard to accommodate in evaluations that build upon predefined quality criteria (Hansson & Polk, 2018).

The literature has proliferated with analytical and methodological frameworks to measure TDR impacts (Walter et al., 2007) or to assess the quality of TDR conduct (Wickson et al., 2006). Others have focused on developing principles and criteria for TDR processes, emphasising on-going reflection and learning by researchers and practitioners to ensure quality (Lang et al., 2012). Here, another purpose of evaluation comes to the fore: next to assessing impact and quality, scholars have argued that evaluation may also support researchers and practitioners in dealing with TDR complexities and contingencies, and support learning-by-doing (Bergmann et al., 2005; Zscheischler et al., 2018). This article aims to reconcile these two distinct foci – transdisciplinary capacity building and impact evaluation – of TDR evaluation. We argue that TDR's complex nature indeed warrants an evaluation approach that is simultaneously supportive of this complexity and hypothesise that combining these evaluation foci meets this purpose. While others have previously reported on similar

endeavours (e.g., Gaziulusoy et al., 2016; Roux et al., 2010; Zscheischler et al., 2018), we explicitly examine how simultaneously supporting and assessing TDR might be complementary and allow for more meaningful impact evaluation and enhanced TDR quality.

To make this case, the next section discusses challenges of TDR evaluation in greater depth. **Section 8.3** presents our case description: the evaluation of a TDR project on Dutch nature policy during which we combined the roles of facilitators and evaluators. **Section 8.4** shows that our approach simultaneously supporting transdisciplinary learning by the coordinating TDR team regarding the operationalisation of four key TDR features, thereby improving the quality of the TDR process, and for meaningful assessment of the project's societal impacts and attributing these to specific process features. **Section 8.5** critically reflects on these findings and the approach.

8.2 Evaluation of TDR projects

As TDR becomes increasingly common it becomes necessary to demonstrate its societal effects and account for the resources invested. Various attempts have been made to empirically capture TDR impacts and TDR has been linked to more usable research products, denser stakeholder networks, enhanced decision-making capacities and policy change (Walter et al., 2007). Wiek et al. (2014) differentiate between TDR *outputs* (usable products) and *outcomes* (network effects and enhanced capacities), both considered intermediate effects which, indirectly and in complex interplay, contribute to societal *impacts* (structural changes and action). The latter tends to occur with significant delay and is found harder to attribute to the specific TDR project (Hansson & Polk, 2018).

To explain – and ultimately advance – these various effects, scholars have been seeking for ways to measure the quality of TDR processes. For instance, Belcher et al. (2016) suggest the perceived *credibility, legitimacy* and *relevance* of TDR research by stakeholders as determinant for impact. However, the complex (political) contexts in which TDR projects take place may make for highly diverse stakeholder views on a project's credibility, legitimacy and relevance, complicating the use of these concepts in guiding the TDR process (Hansson and Polk, 2018). Others have focused on differentiating between types of 'productive interactions' between researchers and stakeholders as the key to quality and societal impact (De Jong et al., 2016), for which Wiek et al. (2014) distinguish between the *nature* (number, type and sequence of interactions) and the *quality* (representation of perspectives, addressing conflict) of participatory processes.

Evaluators, however, are often faced with a lack of high-quality data to sufficiently test these conceptualisations, due to low participation rates and time-lag, which affects stakeholders' memories (Wiek et al., 2014). To fully grasp the complexity that surrounds a project and make informed judgements on the value of its effects and the quality of its process, it is vital to have access to practitioners' perspectives. One reason why TDR evaluation has yet to fully address these challenges is that evaluations run the risk of becoming *decoupled* from the project in question. The evaluator tasked to assess societal impact does so at a relative distance. From this position it is almost impossible to comprehend how a project developed in response to internal and external dynamics, and what, through this lens, constitutes impact, or what might be the appropriate criteria to assess the quality of the project's process (Regeer et al., 2009). To address this, we suggest that an embedded approach geared towards monitoring is better suited for assessing impacts and attributing these to the TDR process.

This is consistent with the observation that the complex character of TDR requires an evaluation approach that is supportive of this complexity (Carew and Wickson, 2010; Klaassen et al., in press). This is pertinent, because it is unlikely that the features the literature suggests as key to successful TDR are all in place when a project commences. Think, for instance, of stakeholders' commitment to collaborate, openness to other worldviews and capacities for bridging epistemic cultures. A novice TDR team can hardly be expected to meet such conditions immediately, or even to know how to develop these from the outset. Because of each project's unique nature this even holds for experienced teams, as each project differs regarding, for example, the relevant stakeholders, their interests and cultures. There is no one-size-fits-all recipe – or 'blueprint' – for inducing effective collaboration in relation to diverse contextual factors (Bracken et al., 2014:5).

In light of this, various scholars have argued that researchers who 'do' TDR require additional capacities (Pohl et al., 2010), such as critical awareness of stakeholders' diverse assumptions, values and worldviews and how these shape participatory research processes (Popa & Guillermin, 2017). Such capacities may be promoted through participatory action–research approaches (Gaziulusoy et al., 2016; Roux et al., 2010), in which insights into the experiences with the on-going TDR project function as direct feedback mechanism to improve its quality. Pohl et al. (2010) identify the role of facilitators: those researchers who are tasked with promoting joint reflection and transdisciplinary learning. We would argue that these facilitators may also be the most appropriate to evaluate the TDR project. Indeed, a facilitator who helps a TDR team build the required capacities to develop key TDR features may also have the best access to researchers' and practitioners' perspectives on its impacts. Being a facilitator

allows for a 'view from the trenches' – a comprehensive understanding of the project's intricacies and how these relate to its outcomes, which we argue is necessary for meaningful impact evaluation.

8.3 Methodology

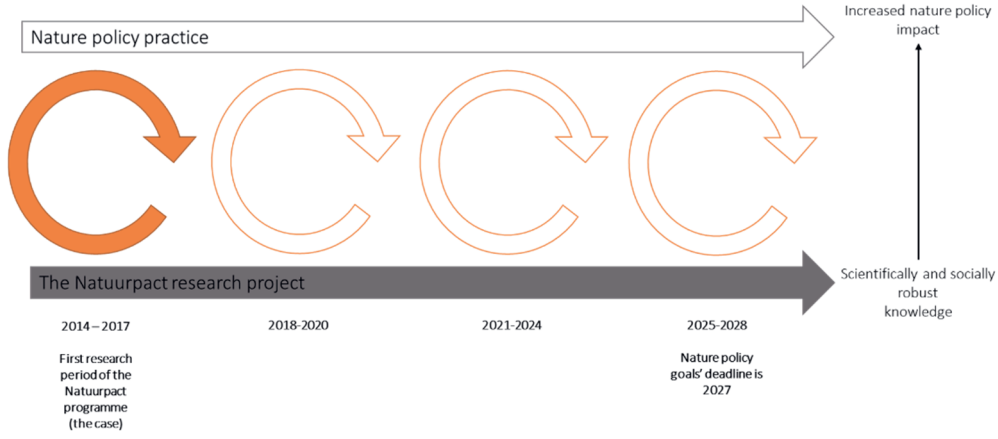

Figure 8.1 Timeline of the entire Natuurpact research program (2014–2028). The case presented in this paper concerns the program's first cycle (2014-2017).

Case study

We present a case study of the first period of the Natuurpact research program: a large-scale, long-term TDR program in the Netherlands. Conflicting agricultural interest and nature conservation goals have resulted in Dutch nature policy becoming increasingly polarised. Uncoordinated attempts on the part of national and provincial governments to address this polarisation were unsuccessful and generated a level of conflict between these governmental bodies. As a step forwards, national and provincial governments, and a number of societal organisations, signed the Natuurpact agreement (2013). In this agreement they finalised the decentralisation of nature policy to the provinces and agreed on ambitions to halt the decline biodiversity and increase social engagement with nature. Part of the agreement was a transdisciplinary policy research program geared at mutual learning and increased nature policy impact: the Natuurpact program. The program is conducted by a team of researchers from the PBL Netherlands Environmental Assessment Agency, (*Planbureau voor de Leefomgeving,* PBL), a government expert organisation, and Wageningen Environmental Research (WER), a university research department, with support of the authors (Athena Institute) as both facilitators and evaluators. It is planned to run until 2028 and comprises sequential and generative three-year research periods. The findings we present are derived from its first research period

(2014–2017), which we refer to as 'project' or 'case' for the remainder of this paper (**Figure 8.1**).

An interdisciplinary core team of six researchers from the PBL and the WER conducted the project (including two project leaders). Few had prior experience with TDR. The project team met twice-weekly to discuss progress and plan research activities. The participants were primarily provincial policy actors who are responsible for the development and execution of nature policy since the decentralisation, and national policy actors who are responsible for international obligatory biodiversity goals. Societal actors such as nature organisations were consulted, but not intensively engaged.

The project comprised three phases: 1) developing a joint research design, 2) conducting research and shared sense-making of findings, and 3) joint formulation of action and dissemination.

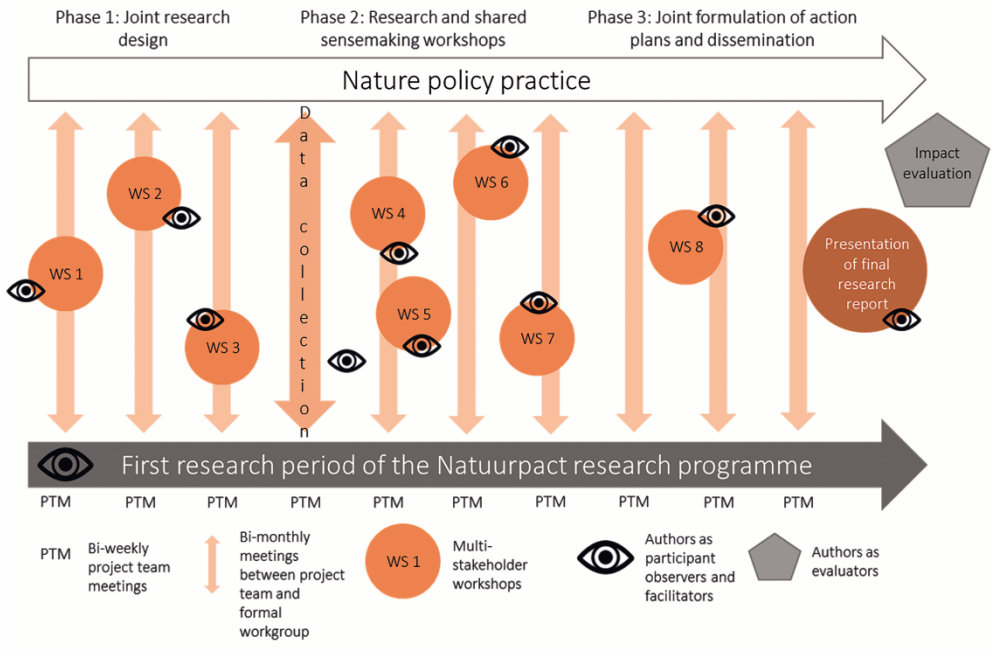

Figure 8.2 Schematic overview of our case, the first period (2014–2017) of the Natuurpact research program.

The main interaction between the team and participants occurred through bi-monthly meetings with a formal working group of 12 representatives from provincial government and eight multi-stakeholder workshops that took place throughout the project and whose purpose differed in accordance with the respective research phase.

Interaction also took place with the program commissioners (administrators from national and provincial government) twice a year, to check that the project was still on track. **Figure 8.2** depicts a schema of the project's design.

Roles, material and methods

In this section we describe our research design according to the two roles we combined: facilitator and evaluator.

Facilitators

Authors A1, A3 and A5 were (variably) part of the project team and supported the team with their first TDR endeavour. As facilitators, they encouraged joint reflection on the challenges posed by developing and practicing the four key TDR features. Due to TDR's contextual specificity, the coordinating team is required to operationalise these more generic TDR features into a bespoke design that adequately corresponds to the issue at hand. Challenges surrounding this operationalisation are therefore considered inherent part of the conduct of TDR; the right conditions, knowledge, competences, circumstances are never in place from the start. The four TDR key features that were to be operationalised are: 1) participation by relevant stakeholders; 2) knowledge integration for change; and 3) responsive and emergent design; and 4) effective boundary management (Hoffmann et al., 2017; Regeer & Bunders, 2009; Scholz & Steiner, 2015a).

The facilitators recorded the developments in the team regarding the challenges associated with these features and other internal and external project developments in field notes, as well as participant observations made during the multi-stakeholder workshops. The facilitators also supported the design (and facilitation) of these workshops, and the operationalisation of the TDR features featured heavily in the discussions around their design. Interviews were held with individual team members at the start and halfway through the project on their views on the project's successes or failings, which were audio-recorded and transcribed. The insights from these data on the project's progress were used to inform future sessions of joint reflection and served as feedback mechanism.

To identify the development of the team's TDR capacities, we retrospectively analysed these data using content analysis (Hsieh & Shannon, 2005) and focused on the team's encountered challenges regarding the TDR features, their actions to overcome these, as well changes in the team's overall narrative regarding these challenges.

Evaluators

A1, A2, A4 and A5 evaluated the project after its conclusion through an exclusively qualitative approach (including semi-structured interviews followed by a focus group discussion (FGD). Here, the facilitators' earlier involvement provided important contextual background and familiarisation with the project and its participants that informed the evaluation design, such as the respondent selection and the design of the interview and FGD guidelines. The interviewees included seven provincial and three national policy actors who were selected on the basis of diversity in their degrees of participation (ranging from two workshops attended, to being a member of the working group) and levels of enthusiasm for the TDR approach (to ensure critical appraisal). As we were interested in the participants' views on what constituted impact and attributable process elements, we adopted Hellström's action-value attribution framework (2015; Hansson and Polk, 2018). During the interviews, the participants were asked in which ways the project had been of value to their policy practice and to what particular elements of the process they attributed this value. A printed timeline of the project's main events and development was produced based on the facilitators' knowledge combined with project publications, meeting minutes and a guided team self-evaluation (similar to Figure 2). This timeline worked to structure the interviews (which were otherwise flexible and open-ended) and aided potential memory distortion. All interviews were audio-recorded and transcribed. All data were analysed using content analysis, focused on categorising different types of effects and attributed process elements.

The preliminary findings from the interviews were presented and discussed during an FDG with the entire working group, for validation through member checks and joint sense-making. We frequently reminded the participants of our dual role as facilitators and evaluators of the TDR project for full disclosure of our research purposes.

8.4 Results

In this section, we first present the findings obtained through our role as facilitators and discuss the four key TDR features and highlight actions of the team to develop these and the corresponding capacities they built.

Key TDR features and corresponding capacities

From the outset, it was evident to the team that they would conduct a long-term transdisciplinary study aimed at mutual learning and improving nature policy impact. Much, however, remained unclear: how to develop particular TDR features, such as stakeholder participation, were matters the team had to learn along the way.

Participation by relevant stakeholders

The first TDR key feature is the participation of relevant stakeholders to address real-world problems and to access their knowledges for socially robust solutions to these issues. During the first project phase, the team approached provincial policy actors to develop a joint research design. They were faced with limited willingness to participate in the research, despite the provincial agreement given by signing the Natuurpact. The team explored the concerns the provinces had for participation through informal conversations, through which they learned about the levels of mistrust between national and provincial government, which had intensified during the decentralisation. Some provinces suspected the research was a strategic move by national government to retain control over nature policy, despite the recent decentralisation. This made the provinces hesitant to open up their policy processes to the team. Stakeholder participation was the first TDR feature the team had to develop.

The team decided to focus their efforts on encouraging the provinces' buy-in to the project and its transdisciplinary ambitions. Three multi-stakeholder workshops were organised to come to a joint research design. During the first two, primarily provincial representatives were invited (approximately 50 versus five from national government) to stimulate their ownership over the project. Nevertheless, the provinces' limited willingness to participate continued to create difficulties in the second phase, during the actual research and shared sense-making of the findings. When the team approached the provinces to collect data on policy plans some withheld information and questioned the legitimacy of the team. The team decided to visit each province individually to explain face to face their intended co-partnership, by which they learned that some provinces strongly felt nature policy was their prerogative and experienced the research as invasive, illustrating their strong sense of ownership over nature policy. By visiting personally, the team came to understand the provinces' point of view, through which rapport was built and access to provincial policy plans for analysis permitted.

In parallel, the team also used other strategies to further incentivise participation. For instance, the team sought provincial government officials from a high strategic level who functioned as ambassadors of the TDR approach. These officials underlined the importance and urgency for the Natuurpact program and their endorsement also encouraged provincial participation.

These diverse strategies to encourage stakeholder participation eventually proved successful: the provinces started to share their policy processes. We observed that by the time the workshops for shared sense-making were organised near the end of phase two, provincial participation had dropped off the team's list of challenges. We

identify the respective capacity that the team built was *developing TDR ownership* for stakeholder participation.

Knowledge integration for change

The second key TDR feature is knowledge integration for change. In light of this, the team made an inventory of the participants' TDR needs during the first three workshops (phase one) for the joint research design. These needs, however, proved difficult for the team to translate into feasible research questions because of their largely operational character. Furthermore, the needs were highly diverse; not just amongst provinces, but also between government tiers there was large diversity. In addition, some team members had their own 'expert view' on what would be relevant research questions. For the final design the team decided to build on the inventoried needs, but also communicated that feasibility of the research design would be a leading criterion.

The workshops in phase two focused on knowledge integration by means of shared sense-making of the research findings. This time it was the team who had to share details of their research process just as the participants had to share their policy plans. As facilitators, we observed how unnerving this was to the team: what if the participants did not recognise the findings? Would their expertise be questioned? Despite feeling vulnerable, the team decided that for mutual learning and knowledge integration, equal footing between them and the participants was paramount. They decided to open up the 'black box' of their analyses and explain their work in way that would allow for deliberation and joint interpretation, by using visualisations and steering clear of jargon. While the team was initially nervous, this approach proved an important success for the entire project, as we will discuss in **Section 8.4.2.** The mode of working initiated by this approach was continued during the project, and led to the development of three shared principles for collaboration: openness, transparency and being upfront about decisions, referred to as 'working without surprises' (Dutch: *verrassingsvrij werken*).

However, the workshops in phase two also highlighted a mismatch between the scale at which some findings were presented (national) and the scale at which provinces sought policy recommendations (provincial). We observed that to the team it had been self-evident that their models would not produce scientifically sound findings at such a local scale, while the provinces felt the national scale held little relevance to their practice. It appeared that the research design had not been communicated with this level of detail because of a failure to understand the need for it. While the mismatch was addressed (which is discussed next), this was an important lesson in expectation management for the team.

We observed the team gradually became skilled in integrating knowledge in a way that ensured the research addressed the participants' divergent needs. Relevant to this was their mode of 'working without surprises' that embodied the equal footing between researchers and participants, as well as expectation management. The capacity that we saw built to develop the feature of knowledge integration for change, was that of *openness and transparency for integrating divergent TDR needs.*

Responsive and emergent design

The third key TDR feature is its responsive and emergent design, which allows the research to develop as insights increase and TDR needs develop. As discussed previously, a mismatch was identified regarding the scale at which the findings had relevance. With intent of being responsive to the participants' needs, the team allocated resources to resolve the mismatch, which resulted in a significantly greater workload that in turn compromised research feasibility. The team experienced a tension between a responsive and emergent design on the one hand, and institutional realities of available time and budget on the other. The lessons that were drawn concerned the need to build in budgetary space and capacity to allow for contingencies, and for critical consideration of which needs the project should be responsive to, namely those that enhance the usability of the findings to contribute to real-world problems.

In line with TDR ideology, as the conclusion of the project approached, the team discussed options for co-authoring the final report with the participants to underline their mutual investments and co-partnership. To the team's surprise, the participants were opposed to this idea. They made clear that they preferred an independent research report with policy recommendations, as these would be more effective for public legitimisation of their policy decisions. As a compromise, the final report was published by the PBL and WER, but featured text boxes with stories from the participants on their experiences with the project.

We observed how familiarisation with the participants' points of view allowed for the team to develop the research in a way that optimised its relevance and usability, while taking the project's feasibility into account. The capacity the team built corresponding to a responsive and emergent design was that of *purposeful responsiveness to emergent TDR needs.*

Effective boundary management

The final key TDR feature is effective boundary management. This concerns the boundaries between the different worlds and institutional backgrounds – with often different rules and expectations – of the actors that are brought together in TDR. In

the project these boundaries were, for example, especially tangible when the team was confronted with the realities from their home organisations. For both the PBL and the WER, TDR was a novel approach and, particularly for the PBL, technocratic conventions on sound policy research prevailed. Consequently, the team frequently had to account for the scientific rigor of the TDR approach. The team managed this boundary, and ensured institutional support for the project, by framing the TDR approach in terms that adhered to PBL's mission statement: as 'a method to enhance policy impact'. Here, we were also mobilised; as TDR 'experts' we were presented by the team as tasked with guarding the project's scientific quality, appealing to the organisations' technocratic rationales.

As facilitators we perceived how the team navigated the different institutional realities of their home organisations by reframing TDR in a way that stroked with the dominant frames within these realities, without compromising TDR's purpose. For effective boundary management, we identified *navigating institutional realities and TDR ambitions* as the final built capacity of the team. **Figure 8.3** shows all four capacities.

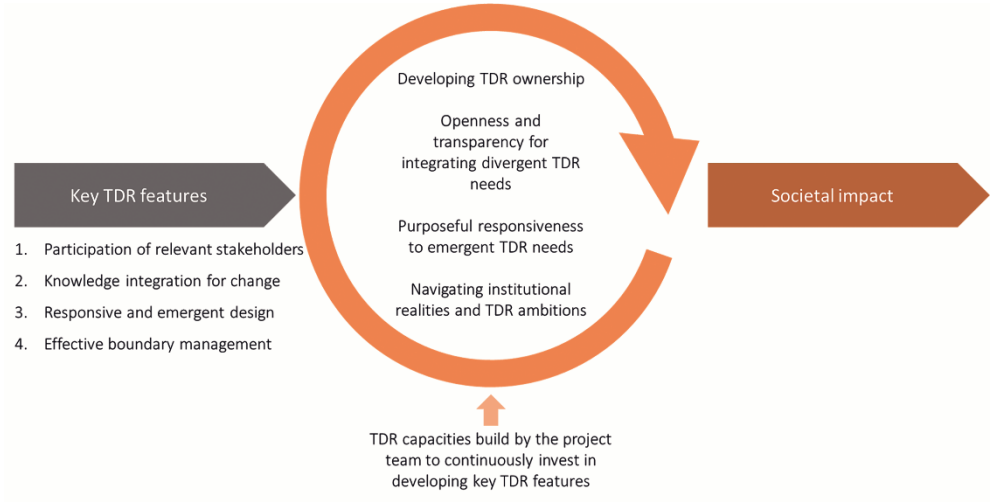

Figure 8.3 Four key TDR features and the respective TDR capacities that the project team built.

Each capacity that was built proved relevant throughout the entire project: developing TDR features was not just a task at the start of the project, but required constant awareness and anticipation. As the team became more skilled, we observed their demeanour and corresponding narrative developed in concert. For instance, rather than expressing concern when provinces were uncooperative, the team's response became more relaxed as they learned to recognise it as strategic play. It became increasingly second nature to the team to consider the participants' points of view in

any research decision, and their confidence with the TDR approach grew as the project progressed.

Impact assessment

We now turn to the project's effects in terms of outputs, outcomes and impact (as suggested by Wiek et al., 2014) (**Figure 8.4**), followed by process elements that were found to be attributable to these effects. Where applicable, we highlight how the role of facilitator complemented our role of evaluator.

Societal effects

Outputs

The outputs of the Natuurpact project comprise its final products: reports and a multitude of presentations, including shared action plans. The participants perceived these as an important prerequisite for societal effects and considered the 'deliverables' useful for attaining or contributing to other outcomes.

Outcomes

Wiek et al. (2014) distinguish *network effects* and *enhanced capacities* (which we termed *cognitive effects*), to which we added two effect categories: *affective effects* and *legitimising effects*.

Network effects

The participants agreed that the increased frequency of their interactions during the project helped expand professional networks and strengthened existing relationships. We could corroborate this with the facilitators' observations from the multi-stakeholder workshops; for example, provinces were seen to talk enthusiastically and change information for future contact. Also illustrative of this effect is the initiation of a provincial 'learning policy network', a platform that focuses on mutual learning on nature policy topics that are outside the scope of the Natuurpact program.

Affective effects

For the following category our role as facilitators was of value in two ways: first, we knew to ask about the participants' relations because we were aware that their tense history had affected their participation. Second, as we had become familiar faces to the respondents, they seemed to be at ease with sharing their concerns and reliefs. It was through this that we found that, more so than network effects, the participants valued the project's *affective effects*: interacting with interprovincial colleagues instilled a sense of relief and reassurance through learning that they face similar issues with

nature policy, thereby validating their own experiences. The subsequent sense of belonging corresponds with Wiek et al.'s community identity (as part of *network effects*). We also consider the increased levels of trust between national and provincial government and the project team an affective effect, the significance of which we were able to grasp through our knowledge of their history. We single out affective effects as separate outcome category as our findings show it was vital for sustained stakeholder participation.

Cognitive effects

Most participants said they had learned from the project: *cognitive effects*. With regard to what they had learned, we identify two categories. The first comprises the newly acquired knowledge that was produced by the project. This knowledge has been formalised in the project's reports and encompassed system, goal and transformation knowledge (e.g., Walter et al., 2007). Part of this new knowledge was also the shared language that we as facilitators observed had developed between participants and the team. In a similar vein, we also observed conceptual alignment and alignment of purposes with nature policy amongst the participants. We consider this first category *new knowledge and enhanced understanding*.

The second category concerns knowledge of a more implicit guise. The participants discussed the value of the project for instilling deeper understanding with respect to their own and each other's perspectives and worldviews. They highlighted how this affected their interactions regarding nature policy in a manner congruent with what we understand as 'anticipatory competence' in relation to stakeholder perspectives – we consider this second category *transdisciplinary competence*.

Legitimising effects

As a final outcome we add the research's *legitimising effects* to the framework. Knowing the participants had declined co-publication, we inquired whether the anticipated legitimising effect of the independent report had been a success, and how. They confirmed that it politically and publicly legitimised and justified their policy agendas, while still enjoying the TDR benefits of enhanced understanding and usability of the findings.

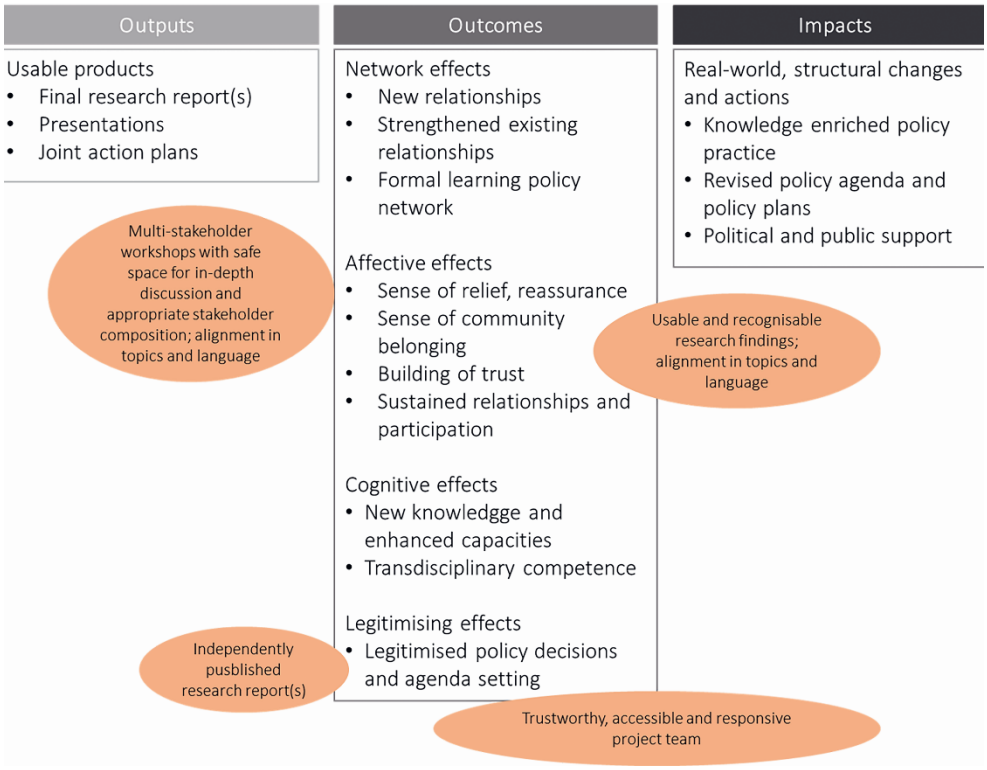

Figure 8.4 Societal effects of the Natuurpact project presented as outputs, outcomes and impacts. Process elements that were identified as attributable by the participants are indicated in orange ovals.

Impacts

Structural changes and actions

Finally, we turn to *impacts* – effects that are the ultimate goal of TDR, namely *structural changes and actions* (Wiek et al. 2014). The participants said that the project's outputs and legitimising effects had been instrumental to their practice: the provinces had used the knowledge to inform nature policy programs and set their policy agenda, for which the legitimising effects issued political and public support. Together with the other effects, which encompass a more social dimension, the development of a *knowledge enriched policy practice* was identified.

Attributed process elements

In order to understand how the TDR project helped to create these effects, we asked the participants which specific elements of the TDR process were attributable. It was here we experienced our preceding role of facilitators as especially pertinent: knowing

the team's actions to develop the four key TDR features, how the project in itself had developed and how workshops and other interactions had passed, guided our inquiry.

The participants identified four main elements that were attributable to the project's impacts. To start with, they attributed networks and affective effects, and parts of the cognitive effects, to the various multi-stakeholder workshops. The provinces mentioned the 'safe space' the workshops had provided, within which they could discuss more sensitive matters, such as their assumptions and opinions on nature policy. We then inquired what specifically had produced this safe space, which was ascribed to the stakeholder composition of workshops: the provinces felt more at ease to discuss potentially sensitive topics when the stakeholder ratio gravitated towards them, as a consequence of the initial distrust between the governments.

Second, the participants attributed cognitive effects and policy actions to the usability and recognisability of the project's findings. Despite the initial mismatch regarding the scale at which the findings were applicable, the research findings had predominantly met the participants' research needs. We also asked about the role of the workshops in light of usability, to which most participants reflected that the alignment of the content of the workshops to the participants' frames of reference – both regarding the workshops' topics and the language used by the team – as crucial element.

The third factor was that the final report was published independently, and therefore attributable to the legitimising effects of the project. The fact that the participants could say *'this is what science advises'* helped them argue for certain policy decisions. The provinces reflected that their own institutional realities did not allow them to diverge too far from a traditional science–policy relationship, and that the responsiveness of the team by providing a compromise had effectively navigated this tension.

Finally, the participants attributed the project's overall success and quality to the project team. Most spoke highly of the team, in particular regarding their transparency, their accessibility to answer questions and their responsiveness to concerns. Participants explained that their trust in the team grew throughout the project, and that this was an important factor in their motivation for active participation in the project (**Figure 8.4**).

8.5 Discussion and conclusion

It has been argued that general agreement on how to evaluate TDR societal impact and quality is viewed as the final phase of TDR development as a research discipline (Carew & Wickson, 2010). This article has sought to address the difficulties that have

been identified in literature for conducting meaningful evaluation of TDR impact and quality. We have argued that some of these difficulties derive from decoupled evaluation, in which access to participants and comprehensive understanding of a project's intricacies are nearly impossible to attain. We also argued that the complexity of TDR warrants evaluation approaches that support teams in terms developing key TDR features from a project's outset in a way that improves its quality. We have shared the results from a combined approach in which we acted both as facilitators and as evaluators of a TDR project, to conduct an embedded, meaningful evaluation.

We did so by outlining this dual role during the evaluation of the first period of the Natuurpact program, which allowed us to identify several transdisciplinary capacities that the team built to enhance the quality of their research project. The capacities we found are consistent with previous work on TDR researchers' skills and associated challenges. To explain researchers' success in addressing these challenges Sarkki et al. (2013) use the metaphor of sensitivity: researchers' ability to be open to the needs and problem framings of stakeholders, to respect different worldviews, perspectives and forms of knowledge, and to understand biases and power relations. To this we add detail on the various ways in which researchers may subsequently act on the knowledge gathered through their sensitivity. Indeed, the capacities we identified go beyond greater insights and suggest a developed know-how and confidence in dealing with stakeholders and their diverse knowledges and perspectives, demonstrating how personal interactions are vital for initiating contact and encouraging sustained engagement (Woltersdorf et al., 2019). Our findings confirm that developing key TDR features requires a constant effort throughout the project (Di Iacovo et al., 2016; Ribeiro et al., 2019). To better understand the link between transdisciplinary capacities, the research process and its impacts, it would be interesting to explore how these capacities relate to the concepts of credibility, legitimacy and relevance as suggested by Belcher et al. (2016), which we intend to do in the following Natuurpact research program period.

In the results we have highlighted moments where we experienced that our involvement with the project as facilitators of transdisciplinary learning processes proved especially complementary to our work as evaluators. We identify four (interrelated) benefits. First, it increased our access to participants, an issue previously identified by others (Wiek et al., 2014). We enjoyed the support of the project leadership and had become familiar faces to the participants over the course of the project. For example, we were allowed to use an entire meeting of the formal working group for our FDG, a quite exceptional occurrence due to their demanding schedules. Second, we had obtained a 'lived through' perspective of the project's internal and

external context and how this, in complex interplay, had shaped the research. For instance, the tense relationship between national and provincial governments influenced provincial willingness to participate, which in turn affected how the workshops were designed. It also influenced which outcomes were attained: the preceding power dynamics increased the importance of the legitimising effects of the research report for the provinces. Although legitimising effects of TDR output in itself are not new, the need for an independent report in our case was a direct consequence of the tense relationships between the parties involved. It is a perfect example of why evaluating TDR according to pre-determined criteria (e.g., 'co-created output') is a poor measure for meaningful evaluation and stresses the importance for evaluators to understand complex contextual factors, such as the political arena in which participants are acting, in order to conduct an adequate impact assessment (Rau et al., 2018). The third benefit we experienced was that our involvement with the project allowed us to conduct better interviews. It made us more sensitive to topics about which to inquire and to carefully probe. In addition, the established rapport with the participants allowed us to discuss impacts that went beyond the direct use of findings and also brought to light impacts that '*are far more intangible but considered just as important by participants'* (Bracken et al., 2014:10). These included affective effects such as a sense of belonging and increased trust between governments, topics that might not have been explored in such depth had we not been aware of the preceding tense relations or if we had not built rapport. Such insights have been suggested as vital for a meaningful and comprehensive evaluation (Hansson & Polk, 2018). Finally, and in line with the previous point, the fourth benefit was that our combined role allowed us to corroborate – or triangulate (Creswell & Miller, 2000) – our interpretations of the project from our facilitator perspective during our evaluation work. This allowed for a more rigorous analysis of different 'impact pathways' that linked the team's capacities to process elements and, finally, to societal effects.

This is not to say the approach was beyond reproach. One shortcoming is its limited potential for capturing unintended effects that go beyond the more obvious 'outcome spaces' of research (Mitchell et al., 2015). Our 'insiders'' view on impacts may have narrowed down our perspective on possible effects and attributable process features. We observe that awareness of this risk and openness to alternative signals are fundamental characteristics of facilitators/evaluators to guard against research bias. In light of this, additional researchers who have not been involved as facilitators (such as A2 and A4 in our case) play a crucial role. Another strategy may be to expand the selection of interviewees to non-participant actors, although then similar difficulties as with decoupled evaluation may jeopardise their commitment to participation. To this, there is no obvious solution.

Furthermore, as facilitators we sought to encourage joint reflection on challenges with the team to formulate collective action. However, it was often difficult to reserve time for reflection, as other matters were perceived as more urgent, an issue also identified by De Wildt-Liesveld et al. (2015). This may also be partly explained by our other role as evaluators. As mentioned previously, to ensure institutional support we were sometimes presented as TDR experts responsible for the scientific quality of the TDR approach. While this appealed to the technocratic culture of their organisations, it may have overshadowed our role as facilitators of transdisciplinary learning: implicitly, the team had outsourced the responsibility for the TDR quality of their project to us. Interestingly, we experienced similar challenges as the team members had faced themselves: just as they had to develop TDR ownership amongst the project participants, we had to develop the team's ownership over 'our' evaluation. As we could not always compel the team to pause and reflect, it appears we were not completely successful. Much has been written on evaluations that seek to reconcile purposes of learning and impact assessment (Botha et al., 2016; Regeer et al., 2016; Van der Meer & Edelenbos, 2006). It is challenging to combine the two foci because they serve different needs (practitioners want a learn-by-doing approach while managers seek insights into cost efficiency) and require a different evaluation approach. Although the difficulties we encountered did not concern tensions regarding our study design, it did lead to confusion about our role and the externalisation of ownership over the evaluation.

Although the issue of linking impact to transdisciplinary research processes is not definitively resolved, our approach implies that combining the roles of facilitator and evaluator results in an evaluation that is better matched to the project under scrutiny. We found these roles are complementary: they allow for in-depth understanding of a TDR project's intricacies and access to practitioners' experiences and their views on the project's impacts, while supporting the TDR team with developing key TDR features in the face of complex internal and external project dynamics. With national and international research funding for solution-oriented research on the rise (e.g., Mazzucato, 2018), TDR popularity is likely to increase. The call for assessing its impact will not decline, nor is it likely that TDR complexity will diminish. This underlines the need to reconcile TDR impact evaluation with promoting 'learning-by-doing' and transdisciplinary capacity building to accommodate TDR's inherent contingencies. The approach presented in this paper may serve as stepping stone for the TDR community to further the conversation on (the impact of) inclusive, reflexive and responsive research.

9 | CONCLUSION, DISCUSSION AND WAYS FORWARD

In the face of wicked environmental and sustainability issues, more reflexive modes of knowledge production are called for, in which academic and non-academic actors collaborate in a process of knowledge co-production for aimed at societal transformation and sustainable development. Much of the literature on such reflexive research is concerned with explaining the discrepancy between its theoretical outcomes and the actual results that are achieved in real-life settings, providing suggestions on how to remedy this 'gap' (e.g., Felt et al., 2016; Flinders et al., 2016; Jagannathan et al., 2020; Mach et al., 2020). It has been argued that reflexive research, despite its ideals for new collaborations between science and policy, in practice deviates little from its technocratic counterparts and more classical ideas on how science and policy should interact (Reinecke, 2015; Turnhout et al., 2013; Van der Hel, 2016). Consequently, rather than the transformation of science-policy interfaces that is called for, the status quo for classical interactions is arguably sustained, while policy researchers appear to talk the talk of co-production, but appear incapable of walking the walk. In this light, scholars have recurrently called for institutional change and the institutionalisation of reflexive research (e.g., Flinders et al., 2016; Kueffer et al., 2012; Schneidewind & Augenstein, 2012; Yarime et al., 2012). Yet, the process by which institutionalisation might be achieved and what its outcomes might look like has so far received relatively little empirical and theoretical attention. This thesis aims to contribute to this through an exploration of the process by which knowledge co-production in policy evaluation – as an enactment of reflexive research – becomes normalised at the PBL Netherlands Environmental Policy Assessment Agency.

A practice-based perspective was adopted and Normalisation Process Theory (NPT) (May & Finch, 2009) was used to study the experiences of PBL's policy researchers who aspire knowledge co-production within the confines of conventional structures that largely shape the science-policy interface. NPT provided a framework to study the work that is done by involved actors to implement, embed and integrate a new practice in contexts marked by complexity and emergence, to enhance understanding of why some practices become routinely embedded while others do not. With normalisation I explicitly do not refer to a situation where a knowledge co-production has become the new normal, subsuming prevailing practices; rather I refer to a situation where it has become *just as normal* to select knowledge co-production approaches as other approaches to policy research.

The research in this thesis was guided by the following main research question:

> *How does the process of normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency take shape?*

So, what do the findings and discussions presented so far imply? What lessons may be drawn from the normalisation process of knowledge co-production to advance reflexive research? I will reflect on this thesis' contributions along the lines of the research sub-questions that were formulated in **Chapter 3**, positioning these in wider academic and societal discussions. This is followed by reflections on the validity of this thesis and its strengths and limitations (**Section 9.3**). **Section 9.4** provides some considerations for ways forward for research and practice, and **Section 9.5** draws this thesis to a close with a concluding reflection.

9.1 Balancing legitimacy and integrity

Developing legitimacy as license-to-operate

The first sub-research question that was formulated was:

> *1: What challenges do policy researchers experience with the normalisation of knowledge co-production and how may these challenges be understood?*

From **Chapters 4** and **5** (and parts of **Chapter 8**) I conclude that the primary challenge policy researchers face to normalise knowledge co-production originates from the need to develop legitimacy for the selected co-production approach within its contexts of application, without compromising the integrity of its theoretical ideals. Such ideals include, for example, the inclusion of a diversity of stakeholders in all phases of the research process. **Chapter 4** and **5** showed how research teams of different policy evaluation projects with a co-production approach – the Natuurpact (NP), Inter-Administrative Programme Vibrant Rural Areas (VRA) and Regional Deals (RD) projects – were faced with critical scrutiny and disciplining of (parts of) their approach by the PBL community. The collaborative character of co-production evoked a strong response regarding its perceived risk for objectivity and independence, and the organisation's credibility in general. Furthermore, **Chapters 4** and **5** (and parts of **8**) showed that various policy actors were apprehensive to become actively involved in the co-production projects, due to perceived risks regarding accountability. Knowledge co-production in policy evaluation was not immediately understood nor supported by actors from both the organisational and policy context: it was only limitedly considered a legitimate and valid approach to policy research.

Chapter 4 introduced the ideal-typical modernist and reflexive institutional logics that provide the scripts upon which actors may draw when engaging with policy research. It was shown that the modernist logic – associated with technocracy and a linear view on policy processes – predominates within the PBL and policy practice, and prescribes many of the norms, quality standards and generally what is found conventional regarding policy research. This corroborates other studies that conclude the

privileging of modernist logic in science-policy systems of the Global North (Chouinard, 2013; Kunseler & Vasileiadou, 2016; Nieminen & Hyytinen, 2015). The reflexive logic – revolving around complexity, system perspectives and interactive learning – was less commonly drawn upon. As some of knowledge co-production's features (manifesting the reflexive logic) may appear highly problematic from within the modernist logic (e.g., stakeholder inclusion may be viewed as risk for researcher objectivity, one of the entrenched golden standards for policy research), this provided depth to understanding why knowledge co-production in policy evaluation lacked initial legitimacy.

Table 9.1 NPT's core mechanisms operationalised for the purpose of studying the normalisation of knowledge co-production (also see chapter 2). The four generative mechanisms should not be understood linearly. Rather, they are interdependent and may occur simultaneously and/or intermittently, depending on what actors who seek to normalise a novel practice feel is required (after May & Finch, 2009).

NPT's core mechanisms	Operationalisation
Sensemaking	Coherent and shared understanding and invested meaning in knowledge co-production that allows actors to act in concert as they share a basic view on the approach's necessity and merit
Engagement	The legitimacy of co-production required for commitment, buy-in and (material) support by actors beyond its original instigators
Enactment	The *doing* of knowledge co-production, including operationalisation of its key features to establish workability, divisions of tasks, resources and responsibility and power
Appraisal	Reflection on and evaluation of knowledge co-production to advance understanding of how it works and to judge its value and impact, thereby informing sense-making, engagement and enactment

In terms of NPT's core mechanisms (**Table 9.1**), the main challenge is rooted in the mechanism of *engagement,* meaning the involvement, commitment and support for co-production of actors from the organisational and policy context required as a 'license to operate'. Partially, this can be explained by a lack of *sensemaking,* meaning a shared understanding of what co-production is, does and requires, amongst these actors. Other challenges were also encountered, such as the limited or contradictory understanding about co-production's purpose and how its 'should' be done, or what the appropriate role for the PBL is. In other words, a lack of *sensemaking* was also seen

on the team-level, which frustrated the interdisciplinary collaboration within the teams (**Chapters 4** and **5**).

Activities to develop legitimacy while guarding integrity

The research project teams were observed to invest significant time and effort to address the lack of legitimacy for co-production, to ensure organisational support and involvement of policy actors required for their projects' success (**Chapters 4, 5** and parts of **8**). The second sub-research question concerned:

> *2: What type of activities do policy researchers undertake to address these challenges and what does this mean for the normalisation of knowledge co-production?*

The actions the team undertook were directed at establishing alignment between embedded norms and customs (rooted in the modernist logic) for policy research, and co-production's features (embodying reflexive logic). Other scholars have previously underscored the importance of aligning newly introduced reflexive practices to prevailing structures for their uptake and institutionalisation (Hoes & Regeer, 2015; Regeer et al., 2016; Schuitmaker-Warnaar et al., 2021). I identified two types of alignment activities: the negotiation of prevailing structures, and the modification of co-production's theoretical-ideal features in practice (**Chapter 5**).

Negotiating contextual structures

The first type of activities was aimed at *negotiating* standing norms and customs that were unconducive to co-production (**Chapter 4** and **5**). I found that the NP and VRA teams sought to accentuate co-production's scientific rigor by appealing to the organisation's golden standards of objectivity and independence, and explaining how these standards were maintained using the same terminology. They engaged in academic collaborations (including the Athena Institute, of which I was part), and extra attention was given to showcase the organised extended peer reviews, and to explain how roles were divided over team members to ensure objectivity. Additionally, the NP team commissioned the Athena Institute to review the value of the co-production approach for the project's 'quality, usability and policy impact', thereby aligning to the organisation's understanding of what policy research impact entails. Implicitly, our review played an important role for legitimising co-production by demonstrating its impact. In NPT terms, to address the lack of legitimacy the teams sought alignment between unconducive norms and co-production features by investing in *sensemaking* (enhancing the mutual understanding of co-production) and *appraisal* (investing in reflection and evaluation of its process and outcomes).

As regards policy practice, **Chapter 5** showed the NP and RD research project teams undertake activities to work unconducive – more traditional – conventions amongst

policy actors concerning science-policy interactions, which led to apprehensiveness to actively take part in the co-production. These activities were directed at building relationships and mutual trust, and included more frequent meetings, time to discuss co-production and mutual expectations more in-depth, and mobilising top-managerial policy actors for their endorsement. These activities belong to *sensemaking* and *engagement* and, as a result, contributed to policy actors taking on an increasingly active role as their familiarisation with the policy researchers and the co-production approaches grew.

I observed that, when successfully negotiated, unconducive contextual norms and customs were broadened to encompass key knowledge co-production features. In such situations, the legitimacy for knowledge co-production increased, leading to enhanced organisational support and involvement of policy actors.

Modifying knowledge co-production features

Chapters 4, 5 and **8** also showed instances when the encountered structures appeared too rigid to negotiate and allow features of co-production.[16] Such instances mostly occurred when the roles of policy researchers and policy actors were perceived to become too intertwined (by both parties alike) or when relational structures – such as inter-professional relationships and pre-existing infrastructures – were absent. In these situations, policy researchers were observed to *modify* theoretical-ideal features of co-production to align to these structures (or to resign to their absence). For instance, in all three projects the feature of stakeholder inclusion was modified to mean that actors were only included during particular research phases, as it was felt that deciding upon research methods and writing conclusions should remain the policy researchers' prerogative. Regarding the feature of stakeholder diversity, the factual involvement of various types of stakeholders remained limited at best in all three cases, due to strained relationships amongst policy actors (the NP project) or to mutual unfamiliarity and absence of previously established networks for effective science-policy interfacing (the VRA and RD projects). Both key features – stakeholder inclusion and stakeholder diversity – were operationalised in ways to accommodate to prevailing structures, but that diverged from what in the literature is proposed as theoretical ideal. In NPT terms, the teams modified the *enactment* of knowledge co-production. It is here that the second part of

[16] In Chapter 2, I introduced the concept of contextual elasticity to explain this phenomenon: the ease with which what is considered conventional and the norm can be negotiated – or stretched – to make space and include ideals for knowledge co-production (after May et al., 2016). It is also used in Chapter 5.

the primary challenge policy researchers faced with normalising knowledge co-production comes to the fore: the development of legitimacy for knowledge co-production, without compromising the integrity of its original purpose.

Already in 1982, Roitman & Mayer discussed how, when introducing a new practice, a certain degree of on-site modification is likely required for it to be workable and align to prevailing norms and customs that prescribe ways of working and interacting in a given 'host' context. They also warn against modification *"beyond the zone of drastic mutation"* (p:3), as this may cause the loss of purpose fidelity, or integrity, as it has been termed in this thesis.[17] I observed such over-modification occur as stakeholders were excluded from certain research phases and the delimitation of stakeholder diversity. Over-modification may explain why policy researchers are perceived to do *"more of the same but under a different name"* (Van der Hel, 2016:173). As they introduce co-production in contexts more partial to technocratic discourses on expertise, rigid unconducive norms and customs they encounter may 'push back', leading them to modify key features to the point that their integrity is lost. It may also (unwittingly) cause tokenism (Paylor & McKevitt, 2019; Wynne, 2007): stakeholder inclusion may become limited to who policy researchers consider relevant and whose involvement is considered feasible (e.g., due to problem framings that align closely to that of the researchers), thereby perpetuating power imbalances (Wynne, 2007), while at the same time they may boast the inclusivity and deliberative character of their approach. As features are over-modified, policy researchers may (unintentionally) default into modernist logic routines, thereby perpetuating more conservative knowledge production practices and technocratising participation (Chilvers, 2008). Together, these findings contribute to deepened understanding of how the observed theory-practice gap (Flinders et al., 2016; Jagannathan et al., 2020; Mach et al., 2020) becomes manifest.

The findings suggest that certain (aspects of) features of knowledge co-production are 'non-negotiable': over-modification of these features leads to disappointing outcomes as integrity is lost. I found that what exactly is non-negotiable was obscure to the policy researchers initially: it was not immediately clear to them when a line was crossed and integrity was compromised. Policy researchers who were less familiar

[17] In **Chapter 2**, I introduced the concept of *plasticity of co-production features* in this regard*:* the discretion researchers have to operationalise co-production's features to accommodate to conventional normative and relational structures, without losing co-production integrity (after May et al., 2016). It is also used in **Chapter 5**.

with co-production more easily made modifications, unwittingly potentially putting the integrity of their approach at risk. Insufficient knowledge and understanding of knowledge co-production – e.g., the mechanism of *sensemaking* – thus seems an important precursor for integrity loss. It also presents an opportunity: **Chapters 4** and **5** also showed that in some situations integrity loss was (partially) remedied as experience and understanding of knowledge co-production grew. I turn to this next.

Learning-by-doing to develop knowledge co-production capacity

Chapters 4 and **5** showed several instances during which the research project teams were confronted with the effects of over-modification. Delimiting stakeholder inclusion to certain phases had reduced the relevance and usability of the research findings to some policy actors (NP project). Similarly, a narrow view on stakeholder diversity (i.e., a focus on national policy actors) was found to also negatively affect usability and societal support for the research (all three projects). As the value of these features became materially tangible (by investing in *appraisal*), policy researchers were observed to re-modify features and re-negotiate structures to enhance co-production integrity. Concretely, this led to policy actors becoming more actively informed and consulted regarding research methods in the second NP cycle (the final decision, however, remained the researchers' responsibility), and efforts to involve a broader range of actors were increased in all three projects.

The findings demonstrate that policy researchers developed knowledge co-production capacity: awareness of key features and the skills to practise these with integrity, while simultaneously aligning to conventional norms and customs. Specific skills included sensitivity to disparate problem framings, interests and forms of knowledge (including their own), understanding of asymmetrical power relations and institutional realities, and the ability to act upon these insights (**Chapter 8**). Co-production capacity grew through experience and learning-by-doing, by which policy researchers became better able at developing conducive conditions and, as their awareness of co-production's non-negotiables grew, they became more advanced in guarding its original purpose.

This stresses the importance of reflexive learning for normalising knowledge co-production (Egeland et al., 2019; Lang et al., 2012; Pohl et al., 2010). However, the learning about co-production I observed in the studies to large extent appeared superficial. Often, learning occurred ad hoc, and only insofar as was required to resolve immediate (legitimacy) issues or insofar as was necessary to continue collaboration within the teams (without truly resolving interdisciplinary or epistemological confusion). Indeed, as the lack of *engagement* presented a more direct risk for the projects' continuation and success, combined with time pressure, this caused the challenges of *sensemaking* and *enactment* to be addressed only insofar as

was required to continue with the projects' execution. Deeper reflection on prevailing norms and customs guiding their research practice and on knowledge co-production's underlying values remained largely absent. This may have important implications for co-production normalisation. In their book on the institutionalisation (or rather, lack thereof) of public engagement with the governance of science and technology in the UK, Wilsdon and colleagues (2005) make a similar observation. They argue that to achieve a situation where reflection and wider engagement of science with social and ethical aspects is inherently understood as part of scientific excellence, social researchers who advocate such change should more engage with the values and norms that govern scientific practice. For reasons understandable, the authors write, social researchers have been inclined to focus on what they frame as the 'hardware' of co-production, meaning methods like focus groups and workshops, and other more procedural aspects of co-production. However, to ensure normalisation, attention for hardware needs to be accompanied by attention for 'software', i.e., co-production's underlying norms, values and purposes and how these may conflict with other norms and values in surrounding research practices (also pointed out by Stilgoe, 2007, and Stirling, 2008). This is pertinent, it is argued, as without the attention for software the required cultural change may lag behind, resulting in *"little more than the scientific equivalent of corporate social responsibility: a well-meaning, professionalised and busy field, propelled along by its own conferences and reports, but never quite impinging on fundamental practices, assumptions and cultures"* (Wilsdon et al., 2005:19). Attention for software by reflecting on the norms and values – both novel and prevailing – that guide the research practice hence appears elemental for the normalisation of an integer version of knowledge co-production that holds true to its original ideals. Here lies a potentially opportune role for reflexive monitoring to promote learning processes, which I shall discuss shortly. First, I turn to developments in the organisational and policy contexts in which the projects transpired.

Contextual readiness for knowledge co-production

The third sub-question that was formulated was:

> *3. How do developments in the organisational and policy contexts in which knowledge co-production occurs affect its normalisation?*

In addition to providing detail to the process of normalisation, this thesis contributed to the understanding of the interactions between co-production practices in individual research projects and contextual developments. Specifically, I found that contextual developments affect the rigidity of contextual structures that policy researchers may encounter and therefore the space they experience for adopting and normalising a diverging approach to policy research. From within their research projects, policy

researchers may seek to negotiate these structures, but normalisation arguably also requires what has been referred to as contextual readiness (May, 2013). Such readiness comprises a shared belief and collective commitment to normalisation of actors in the contexts of application, and may be affected by contextual developments.

Organisational readiness

I found indications for such readiness, for instance, in PBL's Executive Board that actively endorsed knowledge co-production, in the hiring of departmental managers with affinity to deliberative and participatory methods and in the embedding of the approach in the organisation's 'vision on quality' and Vision2025 (**Chapter 5**). Together with the increased occurrence of co-production projects, these indications point towards the emergence of a wider organisational discourse in which reflexive modes of research are embedded. In support of normalisation, scholars underline the importance of reflexive learning not just on the level of individual researchers or teams (as discussed previously), also organisational learning is pointed out as elemental for institutional change (Pallett & Chilvers, 2013; 2015). My findings corroborate previous studies that suggest that learning on the organisational level is crucial for normalisation to be successful. Such organisational learning may be understood as to pertain to the managing of co-production's 'orgware', meaning the organisational set-up and mechanisms to support the implementation and development of the aforementioned hardware and software (Schuijer, 2020; Taibi et al., 2016). In their study on the transition of the Pacific energy sector, Taibi et al. (2016) argue that such a transition cannot be attained without investing in orgware, hardware and software simultaneously. An organisation may play a significant (and intentional) role in advancing normalisation by providing the required orgware and the *"responsive institutional environment"* (Mommaas & Eweg, 2011:56, on stimulating agro innovations) necessary for organisational innovation.

After the research in this thesis had concluded, the emergence of a reflexive discourse and the embedding of co-production approaches had become further apparent in the organisationally supported initiation of an internal research programme (in which I am also involved) explicitly aimed at advancing organisational professionalism as regards co-production and furthering its normalisation, and additional requests for knowledge co-production in the evaluation of large-scale national policy programmes. At the same time, however, discussions within the organisation on the co-production's validity and legitimacy and on the role and position of PBL researchers as advocates for reflexive governance continue to endure. To some, *"it reeks too much of action research"* (personal communications, 18-6-2021), where 'too much' refers to the observed risk of being accused of normativity as the ultimate threat to objectivity and impartiality. The golden standards of objective research that is impartial to policy

processes thus continue to be mobilised, which attests to how profoundly institutionalised such standards are (Chouinard, 2013; Flinders et al., 2016; Lahsen & Turnhout, 2021; Turnhout et al., 2020).

Readiness of policy practice

As regards the readiness of policy practice for knowledge co-production, **Chapter 6** showed that extrinsic political developments affected how knowledge co-production was valued by policy actors, and thereby determined the space for its enactment. Four ideal-typical impact rationales were presented, which comprise actors' take on the perceived function of knowledge co-production, the appropriate pathway to impact and, underlying these notions, how they identified nature policy issues. These were the *accountability, instrumental, network* and *transformative* rationale. Notably, it was found that while the rationales appeared largely incompatible in theory, individual actors in practice were shown to mobilise parts of multiple impact rationales interchangeably over time. Political developments, such as the approaching policy goal deadlines, and the decision of the Council of State that led to the so-called 'nitrogen crisis', were identified as an important factor. As political stakes became more pressing, national policy actors in particular were observed to increasingly mobilise impact rationales that correlated with modernist views on science-policy interfaces. More classical functions of policy evaluation – i.e., serving accountability purposes and providing strategic and instrumental knowledge – became favoured over functions such as joint policy learning. This has also been addressed by Kowalczewska & Behagel (2019) who studied the influence of political context on policy actors' preference for conventional or co-production approaches to knowledge production regarding agricultural and environmental policies in Poland. They found that as political tension grew, policy actors more and more defaulted into conventional science-policy relationships to distance themselves from public accountability for policy decisions. Other scholars have also drawn attention to how political circumstances influence what type of science-policy interfacing is considered appropriate, showing that when political stakes are high, policy actors are inclined to favour undisputedly independent knowledge production processes (Dunn & Laing, 2017; Flinders & Buller, 2006). This raises important questions, such as: can the privileging of conventional approaches to policy research be absolved under political pressure? In other words, is it practically possible to normalise reflexive research in highly political settings in which policy actors request undisputedly independent and objective research (meaning: conducted at literal, physical distance) to distance themselves from public accountability? At the same time, I also found that political developments increased the prevalence of the transformative rationale, as the perceived urgency for system learning and societal change increased, mostly amongst

provincial policy actors and societal actors. Political developments hence may also enhance contextual readiness for knowledge co-production, as actors may experience increased urgency for system change and reflexive ways of working.

The increasing readiness of policy practice for reflexive research may also be observed in developments that were outside the scope of this thesis. For example, in response to a motion on drawing lessons from previous policy experiences to improve current policy and to guard against another potential economic crisis, filed by members of parliament Sneller and colleagues (Najaarsnota, 2018), the Ministry of Finance initiated a national government wide 'Operation Insight in Quality' (Dutch: *Operatie Inzicht in Kwaliteit*). This programme is directed at improving the quality and impact of public policy through instigating a cultural change from rational – modernist – outlooks on policymaking to more iterative, reflexive and deliberative views more suitable for contemporary policy issues (Houppermans, 2018). As part of this operation, some Dutch ministries have started to experiment with knowledge co-production in policy evaluation to enhance policy learning from evaluation (VWS, 2019). Other ministries have similarly been stressing the importance of more reflexive modes of governance in the face of wicked problems, and the elemental role of reflexive policy research as guide therein (BZK, 2018). Each of these developments points towards the emergence of a more reflexive view on both governance and knowledge production, and the readiness for normalisation within the policy context surrounding the PBL. While the questions posed above are outside of the scope of this thesis, from the perspective of normalisation, the importance of (political) developments in the wider contexts of reflexive research underline this process cannot be understood nor supported without appreciation for their interactions with the knowledge co-production practices (e.g., May et al., 2016).

In both the organisational and policy context, the coexistence of multiple logics or rationales (or, "sociotechnical imaginaries" (Jasanoff & Kim, 2009:120, in Felt et al., 2016) on knowledge production and respective science-policy interfaces is apparent. Although sometimes in uneasy relationships, in the messiness of daily practice, the PBL organisation appears to embrace their coexistence. By broadening the organisation's research repertoire to also include co-production approach, the capability of policy researchers to reflexively determine their appropriate role and tailor their research approach to the issue-governance combination at hand with discretion, is encouraged – as for instance is captured by 'public value-driven policy research' (Dutch: *opgavegericht evalueren*; Van der Steen et al., 2018). Such a disposition reflects an ultimately pragmatic view.

Reflexive monitoring to enhance understanding and assess impact

The findings so far lead me to reflect upon my role(s) during the NP research project and the potential of reflexive monitoring for supporting the normalisation of knowledge co-production, captured by the fourth and final sub-question:

> *4: How may reflexive monitoring support the normalisation of knowledge co-production?*

In **Chapter 7** I made the case for reflexive monitoring to adVance co-production integrity and its normalisation (Lang et al., 2012; Lux et al., 2019; Norström et al., 2020; Regeer et al., 2009). It was argued that the complexity and context-specificity of transdisciplinary knowledge co-production processes warrant a monitoring approach that supports the respective project team with implementing its key features with integrity. Reflexive Monitoring in Action (RMA) (Van Mierlo et al., 2010) was identified as potentially suitable methodology to do just that. Much alike the knowledge co-production approach it may seek to support, reflexive monitoring is a methodology aimed at enhancing the reflexivity of the project team to allow them to act in greater accordance with their ambitions and ideals in light of the institutional, societal and political context. Furthermore, it was posited that, as reflexive monitors (who join a research project team) develop a deep understanding of the project's progress, they have unique position to assess the project's impact in a *meaningful* manner. This means they can assess the progress that has been made from the perspective of what the project team and other involved actors would regard as success, as opposed to narrowed down and pre-defined metrics that might say little on whether progress has been made on ambitions due to unexpected internal and external developments. As such, reflexive monitors simultaneously function as promotors for and assessors of knowledge co-production processes. Based on **Chapter 7**, I expected reflexive monitoring to give input and structure to the mechanism of *appraisal*, to from thereon inform the other mechanisms. As such, I presumed reflexive monitoring would support the normalisation of knowledge co-production by supporting the integrity with which its features are practised. On the basis of **Chapter 8** I conclude that reflexive monitoring indeed has potential for supporting normalisation, albeit in somewhat different ways than I originally anticipated.

Reflexive monitoring to enhance understanding

Chapter 8 reflected on our involvement as reflexive monitors and our combined roles as facilitator of reflexive learning processes and impact assessor in the NP research project. As facilitators, our tasks included encouraging joint reflection by the project team on the challenges they encountered with putting key features of knowledge co-

production to practice. As Dewey argued – a prominent advocate for reflexive learning – said: *"We do not learn from experience. We learn from reflecting on experience"* (cited in Di Stefano et al., 2014). However, we experienced difficulties with reserving sufficient time for such reflection, as other project matters were often found more pressing. Consequently, reflection often remained superficial. Unwittingly, the NP project team had outsourced the responsibility for maintaining the integrity of their approach to us: they were more occupied with the hardware and the 'doing' of co-production (*engagement* and *enactment*) than with its software and 'thinking' about it (*sensemaking* and *appraisal*).

As we could not always compel the team to pause and reflect on the values, presumptions and normative orientations that guided their practice, it appears our work as facilitator of learning processes was not completely successful in encouraging *sensemaking* via *appraisal*. For reflexive monitoring to successfully contribute to normalisation, I found the respective project team should feel ownership over their reflexive learning process and structural reflection should be sufficiently embedded within the knowledge co-production process. This resonates with studies on reflexive monitoring (Fielke et al., 2017; Rijswijk et al., 2015) and comparable roles for supporting co-production (Klerkx et al., 2017), which underline the importance of embedding such support roles in the very process they seek to support.

Reflexive monitoring to assess impact

The assignment we were given to review the NP projects first cycle (and later, also its second) was formulated as assessing knowledge co-production's merit for 'research quality, usability and policy impact', concepts that view the impact of knowledge production processes as end-result. While 'end-use' was indeed demonstrated (i.e., the findings were instrumental for informing policy-decisions and could be used strategically for setting policy agendas), also important network and affective effects were found, as were more cognitive effects (**Table 9.4**), each a type of impact of co-production that became manifest during its process. Understanding impact as embedded in the process of co-production corresponds better to its dynamic and interactive character, and attests to the value (and quality) of its process (Walter et al., 2007; Wiek et al., 2014). Our assessment also advanced policy researchers' understanding of the process of co-production and broadened institutional ideas on what impact entails (informing *sensemaking*). The corresponding framing of impact (as shown in **Table 9.4**) found some uptake within the organisation beyond the NP project in discussions on the value of knowledge co-production for other policy research projects.

While the formal aim of our impact assessment was to enhance learning on knowledge co-production and improve its quality in the following cycles, as mentioned previously, our assessment also implicitly (and unexpectedly) served an additional purpose of legitimising co-production by demonstrating its impact and validity as policy research approach. This second purpose proved paramount to the project team, who strategically used our involvement as 'independent experts' and our reports in discussions within and outside the organisation to attest to the scientific validity of the approach (of itself, an action that is illustrative of the technocratic discourse on expertise that predominates within the organisation). Demonstrating the impact of knowledge co-production (*appraisal*) thus appeared an important instrument for legitimisation (*engagement*).

Table 9.4 The effects and policy impacts of the NP project, based on Chapter 8.

Effects of knowledge co-production embedded in the process of the NP research project	Policy impacts (structural changes and actions) to which these effects contributed
• *Network effects* (new or strengthened relationships) • *Affective effects* (sense of community belonging, building of trust, sustained relationships and participation) • *Instrumental effects* (insights to directly inform policy decision making) • *Cognitive effects* (new knowledge and enhanced understanding, transdisciplinary competence) • *Strategic effects* (political and public support for policy decisions and agenda setting)	• Knowledge enriched policy practice • Development of formal learning policy network • Revised (provincial) policy agendas • Increased political and public support

Chapter 6 added some nuance to the observation that demonstrating impact unequivocally leads to legitimisation. The presence of multiple impact rationales not only made it challenging for the NP team to align to the diverse knowledge needs of actors who constituted the policy field, it also affected how these actors perceived impact and the legitimacy of knowledge co-production. While some actors considered the approach elemental for supporting the progress on the nature policy goals as it encouraged social and policy learning, others thought the approach undermined accountability and diverted attention from making faster progress.

In reflection on our combined roles as facilitators of reflexive learning and impact assessors, these were found to largely work complementary. However, the precedence given by the team to our impact assessment toned down our facilitator role become (we found our time was increasingly used for our review, instead of

organising reflection sessions). Fielke et al. (2017) point out that reflexive monitors who perform multiple roles within a project may experience challenges with the degree of influence they have on guiding the project. In their study, they reflect upon the experiences of five novice reflexive monitors with six different agricultural co-innovation projects in New Zealand. They argue that a dual role may cause tension between how the reflexive monitor sees her role and how this role is seen by the project leader and project team, which may compromise the monitor's influence on the project. This accurately reflects my own experience – policy researchers seemed more occupied with our role of impact assessor than our role of facilitator of learning processes, due to the former's legitimising value – and seems explanatory for why we had trouble with organising (sufficient time for) joint reflection. While our combined roles also had evident advantages, attention is required for how both roles may be more adequately balanced.

In terms of NPT's mechanisms, I expected reflexive monitoring to mostly promote normalisation by informing the mechanism of *appraisal* and thereby, *sensemaking*. However, as facilitators we were somewhat less successful than we had originally anticipated. As assessors of impact, via *appraisal* reflexive monitoring contributed to *engagement.* Additionally, in this role we informed *sensemaking* (and potentially more successfully so than as facilitators).

9.2 Normalisation as contested, emergent and dynamic process

As regards this thesis' main research question, I found that the normalisation of knowledge co-production takes shape as a (sometimes heavily) negotiated and contested process that emerges over time. It involves the on-going negotiation and stretching of contextual normative and social structures (thereby slowly transforming these to more conducive ones) and modification of co-production features to establish a level of contextual 'fit' that allows policy researchers to practise co-production with legitimacy and integrity.

To study the uptake of knowledge co-production in science-policy systems, I followed May & Finch (2009) who favour the term normalisation over concepts such as institutionalisation, structuralisation or habitualisation. In their view, normalisation places more emphasis on the dynamic and emergent character of the process by which practices become routinely embedded and sustained in everyday life. Indeed, I found that the knowledge co-production's normalisation process was dynamic, not necessarily leading to a final, concluded state. Rather, contextual structures were re-negotiated and key features were re-modified as co-production capacity and contextual readiness grew. This is in line with the contributions of other social practice theorists who propose that practitioners' agency and the structures they encounter

continuously shape and are being shaped by one another (Arts et al., 2013; Spaargaren et al., 2019). The process of normalisation may be understood as a unchoreographed *dance* between agency and contextual structures, between developing legitimacy on the one hand and guarding integrity to co-production's theoretical ideals on the other. Normalisation constitutes the continuous mutual influencing and therefore co-evolution of contextual structures and the actions of policy researchers to practise knowledge co-production – normalisation is successful only if the context in which it is introduced changes alongside the new practice itself (e.g., May et al., 2016; Schuitmaker-Warnaar et al., 2021).

Small steps

Without seeking to disregard the importance of critique on current reflexive research practices and their contested transformative capacity (e.g., Oliver et al., 2019; Turnhout et al., 2020), nor the scholarly work on principles, procedures and frameworks for its advancement (e.g., Belcher et al., 2016; Hegger & Dieperink, 2014; Lang et al., 2012), a more dynamic understanding of normalisation may provide an alternative outlook on the observed discrepancy between reflexive research in theory and real-life settings (Flinders et al., 2016; Jagannathan et al., 2020). I found that temporal less-rigorous application of co-production features may be permissible to accommodate to unconducive structures, and orchestrate commitment and buy-in. As legitimacy was developed, policy researchers were able to re-modify features later on as to enhance their integrity. Rather than focusing on a 'perfect' practice, which may be counter-productive, focusing on what steps *can* be taken to establish a foothold may be more fruitful (Kuzma & Roberts, 2018 in Schuijer, 2020). For example, ensuring active involvement from a limited set of stakeholders might be favourable over involving a wider range of stakeholders who are unwilling to engage in open, constructive dialogue and do not share ownership over the knowledge production process. From thereon, policy researchers may seek to increase the integrity of their practice by, in this example, seeking how stakeholder diversity may be broadened over time. This, however, arguably requires advanced understanding of integrity and non-negotiables to know what features may temporarily be more or less rigorously applied. Moreover, due to reflexive research's high context-specific and situatedness, what is non-negotiable and what determines integrity may likely be equally context- and situation-dependent.

While observations on the risks of co-production (e.g., tokenism, greenwashing, disempowerment of stakeholders) and arguments that co-production demands caution (Flinders et al., 2016; Lemos et al., 2018; Oliver et al., 2019; Paylor & McKevitt, 2019) should be taken seriously, and the scholarly work on the difficulties with its institutionalisation should not be disregarded (Schuitmaker-Warnaar et al., 2021;

Zweekhorst et al., 2002), the current experimentation with its practice that takes place could also be appreciated as one of the many steps that are required for its embedding into the daily practice of science-policy systems. Here, I draw inspiration from scholars in political ecology, who, sceptical about the likelihood of instant, full-scale transformation of prevailing power and political systems in their field, argue for attention to the mundane and everyday: "*the everyday provides a pragmatic site through which to actively participate in the (re-)production of a different order of things, a different culture and society"* (Lawhon et al., 2014:511). Pieterse (2008) suggests that the appreciation of daily and local behaviours may bring to light myriad ways actors may impair, mitigate or navigate engrained norms, conventions and power dynamics. For such efforts to have effect, he continues, it is important they are part of a larger movement that works towards their institutionalisation (corresponding with other scholars' contributions on the need for a systemic view on normalisation of reflexive modes of research, e.g., Braun & Könninger, 2018). This situated and emergent process is defined by Pieterse (2008) as *radical incrementalism*. Rather than a relentless focus on the end-goal, radical incrementalism keeps the large objectives in mind while appreciating the importance of the processes that get us there, placing emphasis on learning-by-doing (Hajer, 2011). This resonates with this thesis' contributions to the understanding of how normalisation takes shape, namely as a dynamic and emergent change process that is continuously negotiated and contested in the daily practices of policy researchers (and policy actors for that matter) and informed by reflexive learning. Such change occurs not in revolutionary steps, but rather knows a more evolutionary and incremental character (as has been shown for other innovation processes as well; e.g., Fischer et al., 2012).

The value of NPT for understanding normalisation

Before attending to the strengths and limitations of this research and suggestions for ways forward in research and practice, I reflect upon on the value of the NPT framework for studying normalisation.

Alternative frameworks?

Scholars have called for the institutionalisation of reflexive practices in various fields (also see **Chapter 2**) (e.g., Fischer et al., 2015; Jahn et al., 2012; Schneider et al., 2019). For example, in the field of health care scholars have argued for the institutionalisation of reflexive approaches and knowledge co-production to more adequately address sustainability issues within health care systems (i.e., diminishing returns and the misalignment between health care services and societal needs) (Ribeiro et al., 2019; Oliver et al., 2019; Paylor & McKevitt, 2019). Various studies adopt system innovation theory to study and promote the institutionalisation of reflexive practices in health care systems (Essink, 2012; Schuitmaker-Warnaar et al.,

2021). What might my research approach and conclusions have looked like should I have adopted system innovation theory instead of the NPT framework to study the normalisation of co-production? To understand how transitions to a new system take shape, system innovation theorists have developed the multi-level perspective (MLP) (Geels, 2006; Geels & Schot, 2007; Kemp et al., 2005; Schot & Geels, 2008). With MLP, three nested and interrelated levels are distinguished which together and in complex interplay comprise a socio-technical system. The level of the socio-technical regime provides the rules along the lines of which actors' (and actor groups') actions are coordinated, thereby providing systemic stability. I would argue that in my research, the organisational and policy contexts in which the policy evaluation projects geared towards knowledge co-production (my selected cases) took place, formed this regime level. It follows then that the projects transpire at what is referred to as the niche-level (De Wildt-Liesveld, 2015; Geels, 2014; Regeer et al., 2016; Schot & Geels, 2008): "*the locus for radical innovations"* (Geels, 2006:171). Niches are proposed to provide a relatively 'protected space' to shield innovations against early processes of market selection and unconducive systemic barriers. Within these niches it is possible to elude rules and norms that prevail within the system, providing the necessary space to learn about the technical and social aspects of an innovation and their interaction with the regime to promote its mainstreaming (Raven et al., 2007; Schot & Geels, 2008). From this theoretical perspective, the cases selected in my research would identify as 'niche experiments' (De Wildt-Liesveld et al., 2015). Finally, the wider external environment – the socio-technical landscape – presents the backdrop against which socio-technical systems exist, and includes wider developments such as globalisation or environmental changes. Actors in the regime have no direct influence on this landscape. In my research, this landscape arguably is formed by wider developments in the organisational and policy context that are outside the direct sphere of influence of the projects, but that (together with developments on the regime level) play an important role for whether institutionalisation may occur.[18] Highly similar to the NPT framework, scholars argue that innovations may have a mismatch with the existing regime causing difficulties for their breakthrough – a level of 'fit' is required between prevailing rules and the values and ideals proposed by the innovation (Geels, 2006). This level of fit may more easily occur when conditions in the relevant regimes and

[18] There is on-going academic debate on what counts as landscape or regime, or niche for that matter. Scholars have criticised the operationalisation and specification of regimes and how the analytical levels of MLP should be applied empirically (Berkhout et al., 2004; Kok et al., 2021). However, these debates are outside the scope of this thesis.

landscapes are simultaneously favourable: a window of opportunity arises for the innovation's mainstreaming and wider system transition.

From the above I derive that MLP and system innovation would also have provided a suitable theoretical lens through which to interpret the research presented in this thesis, and on a general level might have led to similar conclusions: that normalisation of knowledge co-production is occurring within the PBL and that this process is dynamic and emerging. However, an important difference is that from MLP it follows that the regime level is of primary interest, as transitions are understood as the shift from one regime to the next. There has been some criticism in this regard that MLP underplays the role of agency in transitions and therefore provides limited analytical tools for studying the interactions between actors' activities at the niche level, and the regime and landscape levels (Geels, 2011). NPT is a practice-based theory, meaning that the actions of actors at the niche level are the main object of study – but with high regard for their nested position and relations with the regime and landscape. NPT's core mechanisms provided a more detailed framework along the lines of which the work practitioners do for normalisation may be guided, evaluated and understood (May et al., 2011). As such, the way normalisation is conceptualised in NPT holds more analytical space for the intricate negotiations and debates that affect (the direction of) its process, on which I (as participant-observer) had a direct view. Consequently, the research presented in this thesis takes a mostly 'zoomed in' perspective on the niche level; on practices. I studied regime and landscape developments, and windows of opportunity (which I referred to as contextual readiness), only from the perspective of the selected cases. I have not (or limitedly) 'zoomed out' to study developments on the regime and landscape levels and how these promoted or obstructed normalisation of co-production on the niche-level. As I discuss shortly in **Section 9.4** as ways forward for research, scholars argue for complementing zooming in with zooming out to fully grasp normalisation processes within socio-technical (science-policy) systems (Braun & Könninger, 2018).

Additionally, my selected cases arguably do not meet the characteristics to be defined as niche experiments. Each project was initiated and took place in the heart of the organisation's work: each presented a high profile, potentially politically-sensitive and large-scale project, for which the supposed protective space was largely absent. These endeavours were not shielded from regime rules to allow for experimentation, rather they were directly confronted with such rules, demanding immediate action by their initiators ensure their viability, continuation and success. This led to messy situations, debate and negotiation, resulting in over- and re-modification of features and the co-evolvement of regime rules, which provided the empirical context to study the work policy researchers undertook for implementing and normalisation co-production.

Recently, Grin (2020) suggested the emergence of 'second generation' innovation initiatives that occur from within the heart of the regime as opposed to at the fringes of the prevailing system, which is where niche experiments are supposedly located. The challenges these second generation initiatives face differ, as they seek to simultaneously benefit from their nearness to the incumbent system (e.g., access to resources and competence, opportunities to influence views of regime actors) while orchestrating sufficient space for developing solutions that defy current rules and structures. This conceptualisation seems a more fitting description of the selected cases in my research, and it would be worthwhile for future research to take the propositions Grin suggests for the guidance and study of second generation experiments into consideration in addition to a practice-based view.

Sights on core mechanisms for normalisation

I found that policy researchers' challenges with practising co-production and their actions to address these could be understood in terms of NPT's *sensemaking, engagement, enactment* and *appraisal,* providing further action perspectives for its normalisation. While the project teams heavily invested in *engagement* and *enactment*, the mechanisms of *sensemaking* and *appraisal* gained less attention (or at least only insofar as was necessary to encourage *engagement* and *enactment*), while this might have enhanced the integrity with which co-production features were implemented in earlier stages. Other scholars have drawn similar conclusions, pointing out that explicit attention should be given to the normalisation process early on (e.g., Clarke et al., 2013; Jones et al., 2016). This corroborates my earlier observation that learning occurred ad hoc, and that policy researchers were inclined to focus on hardware as opposed to software.

Additionally, I found that impact assessment as input for *appraisal* was deemed a more important instrument for policy researchers to enhance legitimacy as I was led to expect on the basis of NPT literature (e.g., May et al., 2015; Murray et al., 2011), as it helped to 'prove' co-production's merit and validity. Indeed, impact assessment supported normalisation as policy researchers could use the results to attest to the approach's value, resulting in increased organisational support (*engagement*). This observation adds some nuance to other literature that uses the NPT framework, as it suggests that different types of investments in different mechanisms at different moments in time may be required, depending on the level of controversy that surrounds a new practice, rather than overall 'equal' investment in all mechanisms at once. A practice that is considered more controversial in its intended contexts may require more initial investment in *engagement* and may benefit from highly formal external impact assessment, while practices that are highly complex but less contested may require earlier on investment in *sensemaking*. This has implications for scholars

who use the NPT framework either prospectively to guide normalisation (e.g., De Brún et al., 2016) or retrospectively to gain understanding for why a practice became embedded or not (e.g., May et al., 2011; Murray et al., 2010). In both types of uses, high degrees of contextual sensitivity is required to understand what type of investments in which mechanisms and when are necessary to advance normalisation.

In its most elaborate conceptualisations, NPT comprises numerous (sub)concepts pertaining to each core mechanism (May & Finch, 2009) (for example, the core mechanisms of *enactment* is conceptualised to comprise *interactional workability, relational integration, contextual integration, skill set workability*).[19] While in theory this may make a framework versatile in use, I found – as did many other users, as pointed out in a review of 29 articles using NPT by McEvoy et al. (2014) – these concepts not instantly intuitive. To make them applicable, significant translation efforts are required, leaving too much room for interpretation and thereby potentially compromising their empirical use (a matter also addressed by Finch et al., 2012). Hence, I followed the example of McNaughton et al. (2019) and chose to focus my analysis around NPT's four core mechanisms, and re-labelled them to make them more accessible and to operationalise the concepts in a relevant way with more ease. This way, 'coherency' was re-labelled as *sensemaking,* 'cognitive participation' as *engagement,* 'collective action' as *enactment* and 'reflexive monitoring' became *appraisal*.

Lastly, I have used NPT retrospectively in this thesis: I have used it to look back to understand the doings and sayings of policy researchers to gain insight into the process of normalisation and how it may be further fostered. In literature, NPT is also proposed as a framework to use formatively and as guide to support normalisation *in situ* (e.g., de Brún et al., 2016). As this may support the normalisation of co-production in practice, I reflect upon this potential in **Section 9.4** as practical way forward.

9.3 Reflections on the research approach

Before I continue with some suggestions for ways forward for research and practice, I reflect on the research approach that was adopted in this thesis. I attend to my own role as researcher-practitioner and the strategies undertaken to establish credibility, dependability, transferability and confirmability (together understood as the trustworthiness of the research, or as the research's internal and external validity;

[19] Also see www.normalisationprocess.org/what-is-npt

Lincoln & Guba, 1985). In so doing, I also attend to the strengths and limitations of this thesis.

During the studies I performed a role of researcher-practitioner at the PBL. I took active part in the research projects I presented as cases, and workshops and meetings with the PBL community and other involved actors. This provided me with a 'view from the trenches' to experience the practical concerns of policy researchers with implementing knowledge co-production in evaluation first hand. Apart from its obvious assets, with this role also specific focal points emerge as regards the trustworthiness of the research.

To start with, working in such close proximity to the object of study presents possible risks for researcher bias by becoming too involved and identify too much with the practitioners. Several strategies were employed to remedy such bias. Firstly, triangulation of research methods was applied to establish a rich and comprehensive account. We drew from multiple data sources obtained through various methods, for instance by combining participant observations with semi-structured interviews (**Chapter 4**), document analyses and focus groups (**Chapters 5, 6** and **8**) and literature studies (**Chapter 7**) in both single case (**Chapters 4, 6** and **8**) and comparative case studies (**Chapter 5**). By triangulating data obtained through various research methods, interpretations could be challenged, confirmed or enriched as they emerged over the course of the research. Secondly, qualitative data obtained through semi-structured interviews and focus groups were audio-recorded and transcribed verbatim. Analyses occurred either by qualitative data analysis software or by Microsoft Excel to gain a structured and systematic overview of relevant concepts and their underlying patterns. To enhance credibility, we send interview and focus group respondents summaries that presented our interpretations of the conversation or discussion, to test these for accuracy from their perspective (i.e., member checking) (**Chapters 4**, **5, 6** and **8**). Also, (parts of) study findings were presented during meetings and workshops with commissioners and/or respondents to gain input and check for recognisability (**all Chapters**). In both situations, we took respondents' and commissioners' feedback seriously, especially with regards to factual inaccuracies. We, however, held prerogative over the studies' findings and conclusions. Thirdly, researcher triangulation was applied: by conducting each study in a team of three or more researchers, we sought to take advantage of each other's individual qualities and perspectives to challenge one another's interpretations and underlying assumptions, to check on biased perceptions and to identify possible blind spots.

Furthermore, I pursued a reflexive stance towards my own perspective and position regarding the PBL and the normalisation of co-production by organising feedback on

my interpretations and conclusions by both academic and practice experts. The research teams were expanded with a (changing) researcher who had not been previously involved with the respective projects nor their context to function as "critical friend". They were tasked with suggesting alternative explanations to the findings, bringing in fresh theoretical and societal orientations and overall critically challenging interpretations and conclusions. Also, I engaged in reflexive dialogues with PBL colleagues to test for complementary and divergent understandings of the findings, to enhance awareness of my own disposition and assumptions and advance understanding of their views and the contexts in which these manifest. To further enhance this reflexive stance, for several studies external scientific review committees were installed to review the research process and outcomes (**Chapters 6** and **8**).

As regards transferability, as a researcher-practitioner I was dependent on the projects and activities regarding knowledge co-production that presented themselves and which I selected for data collection. I have particularly focused on knowledge co-production in policy *evaluation* studies. This means that I have analysed a limited part of PBL's research practice, namely that which focuses on policy evaluation as opposed to, for instance, outlook studies, computational modelling or other types of analyses in which co-production approaches may be adopted. Other occurrences regarding the (the normalisation of) co-production within the organisation have therefore remained largely unexplored. However, the policy evaluation projects that constituted the empirical context of the studies (the NP project in particular) presented three high-profile projects that had the attention of the PBL community due to their innovative approach to evaluation. This augmented the expression of organisational norms for what is conventional and acceptable regarding science-policy interactions during discussions on the new approach and presented a unique opportunity to explore the interactions between these structures and policy researchers who aspired co-production. As regards policy practice, the fact that the cases were policy evaluations had a similarly amplifying effect on the rigidity with which norms and customs were maintained, due to evaluations' inherently political nature. I feel confident that the challenges and activities to pursue the implementation and normalisation of knowledge co-production that were found are similar – if somewhat less amplified – in other knowledge co-production projects. This is further supported by, for example, the challenges of other PBL projects that seek to address complex multi-actor and multi-level (MLMA) policy programmes that are discussed in the organisation's formal Community of Practice MLMA Governance.

As the transferability to other research organisations is concerned, I would argue that the PBL presents a representative case for environmental boundary organisations who are confronted with demands for more reflexive research while experiencing

limitations to the discretion they have to practise alternative modes of knowledge production due to political, societal and institutional expectations on what policy research 'ought to do'. In the literature, institutions that compare to the PBL are frequently used as cases to reflect upon challenges with institutionalisation of co-production. See, for instance, Beck et al. (2014) for a reflection on reflexive research in the IPCC and IPBES, or Van der Hel & Biermann (2016) for a comparison of six science institutions engaged with the Sustainable Development Goals and how these seek to acquire epistemic authority in the face of wicked problems. I content that the difficulties with practising reflexive research that are identified in these cases may also be understood through the lens of normalisation, and expect that loss of integrity due to modification of theoretical-ideal features is explanatory for why reflexive research in these cases sometimes appears a rather conservative operationalisation of how it is presented in theory.

Universities are also frequented in studies on knowledge co-production. While difficulties with practising and normalising reflexive research are also identified for universities (for instance, co-production is considered a risky endeavour, especially for young academics, as it requires them to diverge from traditionally practices and how these are rewarded in academic settings; Oliver et al., 2019), their distance to the science-policy interface as compared to boundary organisations such as the PBL also suggests that they may be less bounded by societal and political expectations for policy researchers. As such, they may experience more discretion in the approaches to knowledge production they select. This, however, would require further inquiry.

In this thesis I have strived for analytical depth to understand the process of normalisation of knowledge co-production. In addition to adopting the NPT framework, I have drawn from multiple bodies of literature from a range of fields to derive a broad range of sensitising concepts to help guide my analysis of the doings and sayings of policy researchers. These bodies included research and innovation studies, sustainability science, and environmental policy research. In each of these fields, different commonly used theories or methodologies for knowledge production aimed at responsible and sustainable development and societal transformation can be identified (also see **Chapter 2**). For instance, in research and innovation studies, Responsible Research and Innovation (RRI) has found wide uptake (e.g., Owen et al., 2012), whereas transdisciplinary research is more commonly used in sustainability science (e.g., Scholz & Steiner, 2015a, 2015b). As regards environmental policy research, collaborative approaches to policy evaluation, such as reflexive evaluation, prevail (e.g., Nieminen & Hyytinen, 2015; Regeer et al., 2009). While the division between these theories and methodologies across bodies of literature is not necessarily strict – for instance, the term transdisciplinarity is also frequented in

studies on RRI (Wickson & Carew, 2014) – in general, they appear to remain mostly separated in research practice. I identified common ground in their issue-driven and actionable character, their deliberative and interactive view on science-policy relationships and their aim to advance the reflexivity of those involved with the knowledge production process, policy researchers and actors alike. Importantly, each of these fields identifies problems with moving from theory and practice, accompanied by scholarly discussion on the importance of and difficulties with institutionalisation of knowledge co-production (e.g., Chouinard, 2013; Polk, 2014; Van Hove & Wickson, 2017). By bringing these fields together in this thesis, I was able to draw on these discussions and mobilise their particular strengths to derive sensitising concepts. For example, in the literature on reflexive evaluation, the difficulties for practising and normalising reflexive evaluation have been explained through the exploration of differing institutional logics regarding policy evaluation practice (Chouinard, 2013; Dahler-Larsen & Schwandt, 2012; Kunseler & Vasileiadou, 2016), while this appears less the case for sustainability science and transdisciplinary research (but see Felt et al., 2016, and Van der Hel, 2016). Similarly, the body of knowledge on the intricacies of participatory learning and reflexive monitoring to support transdisciplinary research endeavours is more advanced than is the case for RRI or reflexive evaluation (Gaziulusoy et al., 2016; Lux et al., 2019). Mobilising different elements of different theories and methods regarding knowledge co-production during my studies allowed me to explore the normalisation of co-production from different angles and provided complementary theoretical explanations for the challenges I observed policy researchers encounter with its normalisation, as well as recommendations to support this process. I feel that, should I have focused on a single theory or methodology, the richness of my exploration of the normalisation process of knowledge co-production as presented in this thesis would have been reduced and less in-depth. The communal search for how to move from reflexive research in theory towards its institutionalisation in real-life settings in all these fields (and the proliferation of concepts and theories that accompanies it) in and of itself suggests its on-going normalisation.

9.4 Ways forward

For research

The research in this thesis brought forward several insights regarding the process of normalisation of knowledge co-production that require further inquiry. To start, I found the temporal less-rigorous application of certain (aspects of) co-production key features may be permissible to develop the legitimacy and commitment required for its initiation. This arguably requires advanced understanding of matters of integrity

and the non-negotiables of co-production to prevent over-modification and unrepairable loss of integrity. However, I expect that due to co-production's high context specific and situatedness, what is non-negotiable may likely be equally context- and situation-dependent. Further study into non-negotiables and the temporal aspect of integrity of knowledge co-production is required to advance understanding on how integrity may be managed and how its normalisation may be fostered over time.

Furthermore, I found that normalisation is dependent on developments in its wider contexts and the readiness of these contexts for reflexive modes of research. In particular, political pressure was observed to be pertinent for policy actors' appreciation for knowledge co-production processes and for what they thought were appropriate science-policy relationships, with implications for how policy researchers may design and implement co-production, as also found by other scholars (Kowalczewska & Behagel, 2019). In **Section 9.1** I posed a number of questions regarding the possibility of normalising co-production in highly political settings. Further inquiry into such instances and the subsequent sayings and doings of policy actors is necessary to deepen understanding of how political developments determine the space for practising reflexive modes of research.

This also brings me to the following consideration regarding contexts. The studies presented in this thesis have mostly considered contexts from the perspective of individual research projects and the experiences from involved policy researchers. I have thereby strived to go beyond studying individual endeavours of knowledge co-production (as is observed to be common practice; Chilvers & Kearnes, 2015; Irwin et al., 2013) and to appreciate these in their wider contexts of application. As social practice theorist suggest, practices are never standalone, but rather are part of a bundle (Schatzki, 2016) or a network of practices that are all practised simultaneously. While outside of the scope of this thesis, this advocates studying the relations and interconnectedness of other practices regarding knowledge use and production in the policy contexts that surrounds the PBL. Nicolini's (2012) suggestion for alternating between 'zooming in' (as I have mostly done) and 'zooming out' to study practices appears especially appt in this regard. As Braun & Könninger (2018) also address, a more holistic or systemic (zoomed out) view on the practice of reflexive research is required to fully appreciate the limitations and opportunities for its normalisation within science-policy systems. While they suggest a number of frameworks to study co-production more systemically, they conclude that further scholarly discussion is required on how best to apply such a zoomed out view. Here, potentially, the Multi-level Perspective (MLP) originating from the system innovation paradigm might

provide useful vantage points (Geels, 2006) as would Grin's (2020) conceptualisation of innovation initiatives from within the heart of the regime.

For practice

The research in this thesis also points towards several considerations for knowledge co-production practice, specifically as regards how normalisation may be fostered. Policy researchers were inclined to focus more on the doing of knowledge co-production that on the thinking about it – in NPT terms, there was less (or, later) invested in *sensemaking* and *appraisal* than might be considered preferable given not only the literature on normalisation (e.g., May & Finch, 2009) but also this thesis' findings. Using the NPT framework as guide during the normalisation process may make policy researchers better able to identify the challenges they encounter earlier in the process, and provide inspiration for strategies to address these. For instance, awareness that a lack of mutual understanding within the research project team frustrates collaboration might lead to active pursuit of shared sensemaking and giving meaning to co-production. In **Chapter 5**, a heuristic tool based on NPT and Kemmis & McTaggart's (1982) action-reflection cycle to support policy researchers with normalising knowledge co-production, by helping them identify unconducive contextual structures and plan relevant activities to address these. During the implementation of knowledge co-production, the heuristic guides them through successive cycles of action-reflection to support the integrity their approach and its normalisation. In **Figure 9.1** I have adapted this heuristic to also include NPT's core mechanisms.

Such prospective or formative use of NPT has been recommended by other scholars (e.g., Murray et al., 2010). Notably, de Brún et al. (2016) recommend to use the NPT framework in combination with participatory learning and action, an action-oriented research methodology that shows close resemblance to reflexive monitoring. While reflexive monitoring in this thesis' studies was not combined with NPT (the NPT framework was applied retrospectively in some to understand the process of normalisation), taking these together has obvious potential for supporting normalisation. However, it was our experience that for reflexive monitoring to have effect, it needs to be sufficiently embedded in the primary process and requires shared ownership for the respective research project team to take the reflexive monitor (and her reflections) seriously (Fielke et al., 2017; Rijswijk et al., 2015). Formalising reflexive monitoring as a compulsory component of knowledge co-production might be of aid in this regard, although a more intrinsic motivation for reflexive learning of policy researchers seems a more important condition for shared ownership.

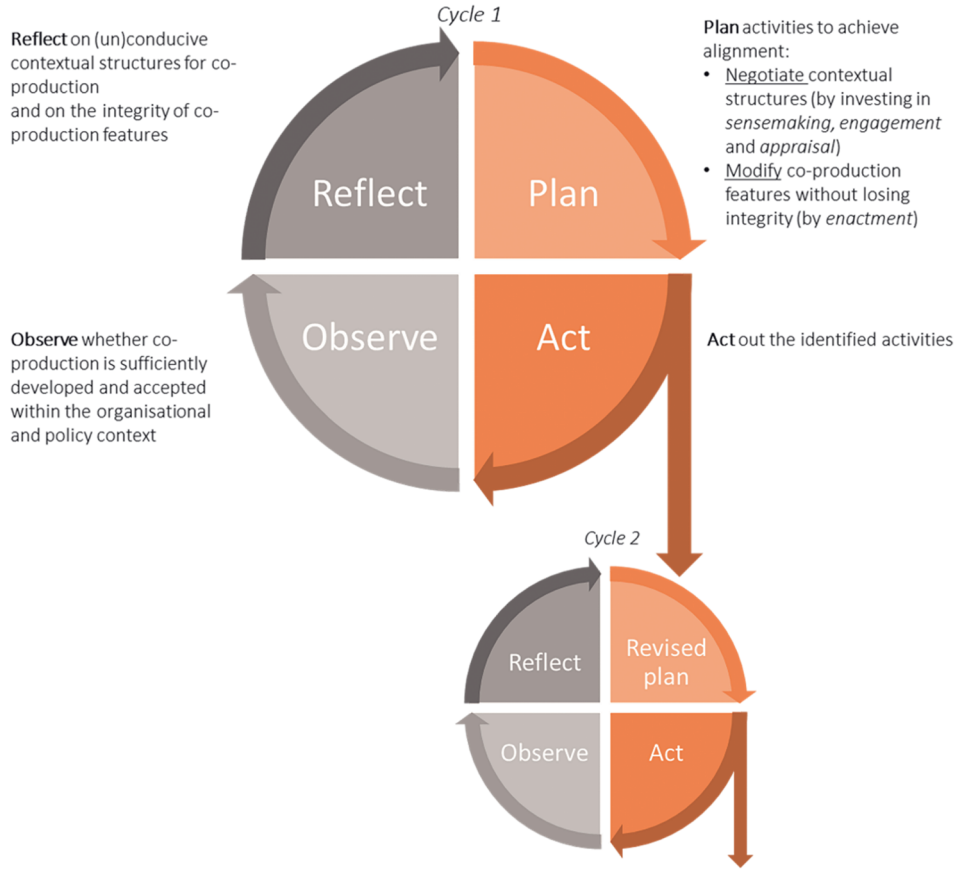

Figure 9.1 A heuristic tool proposed to support policy researchers with the normalisation of knowledge co-production during its implementation (based on Chapter 5).

Finally, I posited that organisations may encourage contextual readiness for the normalisation of knowledge co-production, by supplying the 'orgware' required for supporting the implementation and development of hardware and software (Taibi et al., 2016). Orgware, for instance, concerns the availability of sufficient expertise within the organisation regarding knowledge co-production. This may imply the reconsideration of an organisation's hiring policy. Expertise may also be nurtured by organising platforms for discussing methodological innovation (i.e., hardware) and underlying values and ideals on the roles of policy researchers in science-policy systems (i.e., software), for instance through (informal) Communities of Practice (Galarneau, 2002; Wenger, 2000) or via trainings, workshops and guided peer consultations. Additionally, orgware also concerns the consideration of organisational process and procedures regarding how policy research is conducted. For instance, the emergent character of knowledge co-production processes generally fit poorly with

strictly maintained research proposals and planning, and rather require a type of management style that allows for emergence and iteration, and has adaptive capacity. Finally, in literature on organisational innovation there is argued for appointing (more or less formally) innovation ambassadors: employees who enjoy endorsement from executive and managerial levels and general credibility within the organisation, who may inspire collective pursuit of organisational innovation (Karlsson & Björk, 2017).

As regards the readiness of the policy context, for successful science-policy knowledge co-production degrees of normalisation are arguably also required in these settings. While outside the scope of this thesis, the core mechanisms of the NPT framework and the notions of hardware, software and orgware may also provide useful analytical and practical frameworks to study and promote co-production's normalisation.

9.5 Concluding remarks

This thesis started out with a quote from PBL's chief scientist dr.prof. Bas Arts, illustrating the normalisation of co-production in policy evaluation since it was first introduced in the Natuurpact research project in 2014. While hardly the organisation's first undertaking with knowledge co-production, its scale and duration and the fact that it concerned a *policy evaluation* stirred up quite some dust. As I have shown in this thesis, the process of normalisation that followed was a rocky road, full of contestation and debate, taking one step forwards and, sometimes, two steps back. Notably, it continuous to take shape. Rather than a revolutionary change, the innovation towards a more reflexive research practice has an evolutionary and incremental character (Fischer et al., 2012, in reference to system innovation in the agri-food sector). Nevertheless, relatively small steps may, together, lead to a significant result. In his essay, Hajer (former executive director of the PBL; 2011) quotes Zhang (2011): "*you feel for stones to cross a river",* illustrating how the path towards transformative change comprises a series of – more modest and simpler – changes, and cannot be known beforehand; rather, one has to learn one's way towards it.

Such small steps are visible throughout the organisation. For instance, in explicating the organisation's ambitions for a reflexive practice in its Vision 2025, the recently formulated organisation-wide visions on quality and impact and the initiation of various research projects in which co-production approaches are adopted. Or, in the reflexive dialogues with practice experts on the appropriate role of the PBL in light of contemporary, complex environmental policy issues that are orchestrated by the organisation. It is by taking these small steps, the organisation and its policy researchers may gradually build the required capacity to walk the walk of co-production and move towards a practice of reflexivity.

References

Aarts, N., & Leeuwis, C. (2010). Participation and Power: Reflections on the Role of Government in Land Use Planning and Rural Development. *Journal of Agricultural Education and Extension*, *16*(2), 131–145. https://doi.org/10.1080/13892241003651381

Acs, Z. J., Groot, H. L. F., & Nijkamp, P. (Eds.). (2013). *The Emergence of the Knowledge Economie: A regional perspective*. Springer Science & Business Media. https://doi.org/10.1007/978-3-540-24823-1_1

Adelle, C., & Weiland, S. (2012). Policy assessment: state of the art. *Impact Assessment and Project Appraisal*, *30*(1), 25–33.

Arkesteijn, M., Van Mierlo, B., & Leeuwis, C. (2015). The need for reflexive evaluation approaches in development cooperation. *Evaluation*, *21*(1), 99–115.

Arnott, J. C., Kirchhoff, C. J., Meyer, R. M., Meadow, A. M., & Bednarek, A. T. (2020a). Sponsoring actionable science: what public science funders can do to advance sustainability and the social contract for science. *Current Opinion in Environmental Sustainability*, *42*, 38–44. https://doi.org/10.1016/j.cosust.2020.01.006

Arnott, J. C., Mach, K. J., & Wong-Parodi, G. (2020b). Editorial overview: The science of actionable knowledge. *Current Opinion in Environmental Sustainability*, *42*, A1–A5. https://doi.org/10.1016/j.cosust.2020.03.007

Arts, B., Behagel, J., Van Bommel, S., de Koning, J., & Turnhout, E. (Eds.). (2013). *Forest and Nature Governance - A Practice Based Approach*. Springer. https://www-springer-com.vu-nl.idm.oclc.org/gp/book/9789400751125

Bammer, G., O'Rourke, M., O'Connell, D., Neuhauser, L., Midgley, G., Klein, J. T., Grigg, N. J., Gadlin, H., Elsum, I. R., Bursztyn, M., Fulton, E. A., Pohl, C., Smithson, M., Vilsmaier, U., Bergmann, M., Jaeger, J., Merkx, F., Vienni Baptista, B., Burgman, M. A., … Richardson, G. P. (2020). Expertise in research integration and implementation for tackling complex problems: when is it needed, where can it be found and how can it be strengthened? *Palgrave Communications*, *6*(1), 1–16. https://doi.org/10.1057/s41599-019-0380-0

Beck, S., Borie, M., Chilvers, J., Esguerra, A., Heubach, K., Hulme, M., Lidskog, R., Lövbrand, E., Marquard, E., Miller, C., Nadim, T., Neßhöver, C., Settele, J., Turnhout, E., Vasileiadou, E., & Görg, C. (2014). Towards a reflexive turn in the governance of global environmental expertise the cases of the IPCC and the IPBES. *Gaia*, *23*(2), 80–87. https://doi.org/10.14512/gaia.23.2.4

Beck, S., & Mahony, M. (2018). The IPCC and the new map of science and politics. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(6), e547.

https://doi.org/10.1002/wcc.547

Beck, U., Bonss, W., & Lau, C. (2003). The Theory of Reflexive Modernization. *Theory, Culture & Society*, *20*(2), 1–33. https://doi.org/10.1177/0263276403020002001

Beck, U., Giddens, A., & Lash, S. (1994). *Reflexive modernisation. Politics, Tradition and Aestetics in the Modern Social Order*. Standford University Press.

Beers, P. J., & Van Mierlo, B. (2017). Reflexivity, reflection and learning in the context of system innovation: Prying loose entangled concepts. In B. Elzen, A. M. Augustyn, M. Barbier, & B. Van Mierlo (Eds.), *AgroEcological Transitions: Changes and Breakthroughs in the Making* (pp. 243–256). Wageningen University and Research, Applied Arable and Vegetable Research. https://doi.org/http://dx.doi.org/10.18174/407609

Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R., & Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, *25*(1), 1–17. https://doi.org/10.1093/reseval/rvv025

Berg Johansen, C. (2017). What are Institutional Logics - and where is the perspective taking us? In G. Krücken, C. Mazza, R. E. Meyer, & P. Walgenbach (Eds.), *New Themes in Institutional Analysis - Topics and Issues from European Research*. Edward Elgar Publishing.

Bergmann, M., Brohmann, B., Hoffmann, E., Loibl, M. C., Rehaag, R., Schramm, E., & Voss, J.-P. (2005). Quality Criteria of Transdisciplinary Research. A Guide for the Formative Evaluation of Research Projects. *Institut Fuer Sozial-Oekologische Forschung (ISOE)*, 5–75. http://www.isoe.de

Bergmann, M., Schäpke, N., Marg, O., Stelzer, F., Lang, D. J., Bossert, M., Gantert, M., Häußler, E., Marquardt, E., Piontek, F. M., Potthast, T., Rhodius, R., Rudolph, M., Ruddat, M., Seebacher, A., & Sußmann, N. (2021). Transdisciplinary sustainability research in real-world labs: success factors and methods for change. *Sustainability Science*, *16*(2), 541–564. https://doi.org/10.1007/s11625-020-00886-8

Berkhout, F., Smith, A., & Stirling, A. (2004). Socio-Technological Regimes and Transition Contexts. In B. Elzen, F. W. Geels, & K. Green (Eds.), *System Innovation and the Transition to Sustainability: Theory, Evidence and Policy*. (pp. 48–75). Edward Elgar.

Bijker, W. E. ., Bal, R. ., & Hendriks, R. . (2009). *The Paradox of Scientific Authority - The Role of Scientific Advice in Democracies*. The MIT Press. https://mitpress-mit-edu.vu-nl.idm.oclc.org/books/paradox-scientific-authority

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research*, *26*(13), 1802–1811. https://doi.org/10.1177/1049732316654870

Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation. In B. J. Koops, I. Oosterlaken, H. Romijn, T. Swierstra, & J. Van den Hoven (Eds.), *Responsible Innovation 2* (pp. 19–35). Springer International Publishing.

Boon, W. P. C., Hessels, L. K., & Horlings, E. (2019). Knowledge co-production in protective spaces: case studies of two climate adaptation projects. *Regional Environmental Change*, *19*(7), 1935–1947. https://doi.org/10.1007/s10113-019-01517-4

Borquez, R., Aldunce, P., & Adler, C. (2017). Resilience to climate change: from theory to practice through co-production of knowledge in Chile. *Sustainability Science*, *12*(1), 163–176. https://doi.org/10.1007/s11625-016-0400-6

Borrás, S., & Højlund, S. (2015). Evaluation and policy learning: The learners' perspective. *European Journal of Political Research*, *54*(1), 99–120. https://doi.org/10.1111/1475-6765.12076

Borrás, S., & Laatsit, M. (2019). Towards system oriented innovation policy evaluation? Evidence from EU28 member states. *Research Policy*, *48*(1), 312–321. https://doi.org/10.1016/j.respol.2018.08.020

Bosenman, M. F. (1988). Serendipity and scientific discovery. *The Journal of Creative Behavior*, *22*(2), 132–138.

Botha, N., Coutts, J., Turner, J. A., White, T., & Williams, T. (2016). *Evaluating for learning and accountability in system innovation : Incorporating reflexivity in a logical framework*. *August*.

Bourdieu, P. (1977). Outline of a Theory of Practice. In *Outline of a Theory of Practice*. Cambridge University Press. https://doi.org/10.1017/cbo9780511812507

Bovens, M., 't Hart, P., & Kuipers, S. (2009). The Politics of Policy Evaluation. In R. E. Goodin, M. Moran, & M. Rein (Eds.), *The Oxford Handbook of Public Policy* (pp. 319–335). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199548453.003.0015

Bracken, L. J., Bulkeley, H. A., & Whitman, G. (15 C.E.). Transdisciplinary research: understanding the stakeholder perspectives. *Journal of Environmental Planning and Management*, *58*(7), 1291–1308. https://doi.org/10.1080/09640568.2014.921596

Braun, K., & Könninger, S. (2018). From experiments to ecosystems? Reviewing public participation, scientific governance and the systemic turn. *Public Understanding of Science*, *27*(6), 674–689. https://doi.org/10.1177/0963662517717375

Broerse, J. E. W. (1998). *Towards a new development strategy: how to include small-scale farmers in the biotechnological innovation process.* Eburon Publishers.

Brouwers, H. J. H., Verwoerd, L., & Klaassen, P. (2021). *Op weg met verantwoorden en waardevol leren - Impactstudie van de Lerende Evaluatie Natuurpact 2020.* Athena Instituut, VU University Amsterdam.

Brugnach, M., & Ingram, H. (2012). Ambiguity: the challenge of knowing and deciding together. *Environmental Science & Policy*, *15*(1), 60–71.

Bunders, J. F. G., Broerse, J. E. W., Keil, F., Pohl, C., Scholz, R. W., & Zweekhorst, M. B. M. (2010). How can transdisciplinary research contribute to knowledge democracy? In *Knowledge Democracy: Consequences for Science, Politics, and Media* (pp. 125–152). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11381-9_11

BZK. (2018). *Adaptief Bestuur. Essays over adaptiviteit en openbaar bestuur*. 125. https://kennisopenbaarbestuur.nl/media/255279/adaptief-bestuur.pdf

Candy, L. (2006). Practice Based Research: A Guide. *CCS Report*, *1*(November 2006), 19. http://www.creativityandcognition.com/resources/PBR Guide-1.1-2006.pdf

Carew, A. L., & Wickson, F. (2010). The TD Wheel: A heuristic to shape, support and evaluate transdisciplinary research. *Futures*, *42*(10), 1146–1155. https://doi.org/10.1016/j.futures.2010.04.025

Carter, N., Bryant-Lukosius, D., Dicenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum*, *41*(5), 545–547. https://doi.org/10.1188/14.ONF.545-547

Cash, D., Clark, W., Alcock, F., Dickson, N., Eckley, N., & Jäger, J. (2002). Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making. *KSG Working Paper Series*, *RWP02-046*. https://doi.org/http://dx.doi.org/10.2139/ssrn.372280

Chambers, J. M., Wyborn, C., Ryan, M. E., Reid, R. S., Riechers, M., Serban, A., Bennett, N. J., Cvitanovic, C., Fernández-Giménez, M. E., & Galvin, K. A. (2021). Six modes of co-production for sustainability. *Nature Sustainability*, *4*(11), 983–996.

Chilvers, J. (2008). Deliberating Competence: Theoretical and Practitioner Perspectives on Effective Participatory Appraisal Practice. *Science Technology Human Values*, *33*(155–185).

Choi, B. C. K., Pang, T., Lin, V., Puska, P., Sherman, G., Goddard, M., Ackland, M. J., Sainsbury, P., Stachenko, S., Morrison, H., & Clottey, C. (2005). Can scientists and policy makers work together? In *Journal of Epidemiology and Community Health* (Vol. 59, Issue 8, pp. 632–637). https://doi.org/10.1136/jech.2004.031765

Chouinard, J. A. (2013). The Case for Participatory Evaluation in an Era of Accountability. *American Journal of Evaluation*, *34*(2), 237–253. https://doi.org/10.1177/1098214013478142

Clarke, D. J., Godfrey, M., Hawkins, R., Sadler, E., Harding, G., Forster, A., Mckevitt, C., Dickerson, J., & Farrin, A. (2013). Implementing a training intervention to support caregivers after stroke: a process evaluation examining the initiation and

embedding of programme change. *Implementation Science*, *8*(96), 1. https://doi.org/10.1186/1748-5908-8-96

Craft, J., & Howlett, M. (2013). The dual dynamics of policy advisory systems: The impact of externalization and politicization on policy advice. *Policy and Society*, *32*(3), 187–197. https://doi.org/10.1016/j.polsoc.2013.07.001

Creswell, J. W., & Miller, D. L. (2000). Determining Validity in Qualitative Inquiry. *Theory Into Practice*, *39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

Cummings, S., Regeer, B. J., de Haan, L., Zweekhorst, M., & Bunders, J. (2018). Critical discourse analysis of perspectives on knowledge and the knowledge society within the Sustainable Development Goals. *Development Policy Review*, *36*(6), 727–742.

Cuppen, E. (2012a). A quasi-experimental evaluation of learning in a stakeholder dialogue on bio-energy. *Research Policy*, *41*(3), 624–637. https://doi.org/10.1016/j.respol.2011.12.006

Cuppen, E. (2012b). Diversity ad constructive conflict in stakeholder dialogue: considerations for design and methods. *Policy Sciences*, *45*, 23–46.

Cuppen, E., Breukers, S., Hisschemöller, M., & Bergsma, E. (2010). Q methodology to select participants for a stakeholder dialogue on energy options from biomass in the Netherlands. *Ecological Economics*, *69*(3), 579–591. https://doi.org/10.1016/j.ecolecon.2009.09.005

Cuppen, E., Hisschemöller, M., & Midden, C. (2009). Bias in the exchange of arguments: The case of scientists' evaluation of lay viewpoints on GM food. *Public Understanding of Science*, *18*(5), 591–606. https://doi.org/10.1177/0963662508091021

Dahler-Larsen, P. (2012). *The Evaluation Society*. Standford University Press.

Dahler-Larsen, P., & Schwandt, T. A. (2012). Political Culture as Context for Evaluation. *New Directions for Evaluation*, *135*, 75–87. https://doi.org/10.1002/ev.20028

Davis, M., & Laas, K. (2014). "Broader Impacts" or "Responsible Research and Innovation"? A Comparison of Two Criteria for Funding Research in Science and Engineering. *Science and Engineering Ethics*, *20*(4), 963–983. https://doi.org/10.1007/s11948-013-9480-1

De Brún, T., O'Reilly-de Brún, M., O'Donnell, C. A., & MacFarlane, A. (2016). Learning from doing: the case for combining normalisation process theory and participatory learning and action research methodology for primary healthcare implementation research. *BMC Health Services Research*, *16*(1), 346. https://doi.org/10.1186/s12913-016-1587-z

De Jaegher, H., Pieper, B., Clénin, D., & Fuchs, T. (2017). Grasping intersubjectivity:

an invitation to embody social interaction research. *Phenomenology and the Cognitive Sciences*, *16*(3), 491–523. https://doi.org/10.1007/s11097-016-9469-8

De Jong, S. P. L., Wardenaar, T., & Horlings, E. (2016). Exploring the promises of transdisciplinary research: A quantitative study of two climate research programmes. *Research Policy*, *45*, 1397–1409. https://doi.org/10.1016/j.respol.2016.04.008

De Wildt-Liesveld, R. (2015). *Understanding the governance of niche experiments towards sustainability - Strategies to enhance the adaptive capacity of niche experiments* (Issue March). VU University Amsterdam.

De Wildt-Liesveld, R., Bunders, J. F. G., & Regeer, B. J. (2015). Governance strategies to enhance the adaptive capacity of niche experiments. *Environmental Innovation and Societal Transitions*, *16*, 154–172. https://doi.org/10.1016/j.eist.2015.04.001

Dewey, J. (1927). *The Public and Its Problems: An Essay in Political Inquiry* (2012th ed.). Penn State Press. https://books.google.nl/books?hl=nl&lr=&id=M16E5ORLJqIC&oi=fnd&pg=PP1&dq=dewey+public+and+its+problems&ots=Xzb6UiiIt-&sig=xtKrVKm9ZwaODzTLka0Q0l2m254#v=onepage&q=dewey public and its problems&f=false

Dewulf, A., Klenk, N., Wyborn, C., & Lemos, M. C. (2020). Usable environmental knowledge from the perspective of decision-making: the logics of consequentiality, appropriateness, and meaningfulness. *Current Opinion in Environmental Sustainability*, *42*(October 2019), 1–6. https://doi.org/10.1016/j.cosust.2019.10.003

Di Iacovo, F., Moruzzo, R., Rossignoli, C. M., & Scarpellini, P. (2016). Measuring the effects of transdisciplinary research: the case of a social farming project. *Futures*, *75*, 24–35. https://doi.org/10.1016/j.futures.2015.10.009

Di Stefano, G., Gino, F., Pisano, G. P., & Staats, B. R. (2014). Learning by Thinking: How Reflection Aids Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2414478

Dos S. Ribeiro, C., Van de Burgwal, L. H. M., & Regeer, B. J. (2019). Overcoming challenges for designing and implementing the One Health approach: A systematic review of the literature. In *One Health* (Vol. 7). https://doi.org/10.1016/j.onehlt.2019.100085

Duijnhoven, H., & Neef, M. (2016). Disentangling wicked problems: A reflexive approach towards resilience governance. In *Advanced Sciences and Technologies for Security Applications* (pp. 91–106). Springer. https://doi.org/10.1007/978-3-319-21106-0_5

Dunn, G., & Laing, M. (2017). Policymakers perspectives on credibility, relevance and legitimacy (CRELE). *Environmental Science and Policy*, *76*, 146–152. https://doi.org/10.1016/j.envsci.2017.07.005

Ebrahim, A. (2005). Accountability Myopia: Losing Sight of Organizational Learning. In *Nonprofit and Voluntary Sector Quarterly* (Vol. 34, Issue 1). https://doi.org/10.1177/0899764004269430

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

Elzen, B., Augustyn, A., Barbier, M., & Van Mierlo, B. (2017). *AgroEcological Transitions: Changes and Breakthroughs in the Making*. 336. https://doi.org/http://dx.doi.org/10.18174/407609

Engel, P. G. H., & Carlsson, C. (2002). Enhancing learning through evaluation: Approaches, dilemmas and some possible ways forward. In *Annual Conference of the European Evaluation Society Three movements in Contemporary Evaluation: Learning, Theory and Evidence*. www.ecdpm.org

Essink, D. R. (2012). *Sustainable Health Systems: the role of change agents in health system innovation*. VU University Amsterdam.

Evers, D., Dignum, M., & Hamers, D. (2020). *Illuminating the Black Box of the Government to Governance Transformation Thesis: The Case of Dutch City Deals*. https://doi.org/10.1080/01900692.2020.1759629

Felt, U., Igelsböck, J., Schikowitz, A., & Völker, T. (2012). Challenging Participation in Sustainability Research. *International Journal of Deliberative Mechanisms in Science*, *1*(1), 4–34. https://doi.org/10.4471/demesci.2012.01

Felt, U., Igelsböck, J., Schikowitz, A., & Völker, T. (2013). Growing into what? The (un-) disciplined socialisation of early stage researchers in transdisciplinary research. *Higher Education*, *65*(4), 511–524.

Felt, U., Igelsböck, J., Schikowitz, A., & Völker, T. (2016). Transdisciplinary Sustainability Research in Practice: Between Imaginaries of Collective Experimentation and Entrenched Academic Value Orders. *Science Technology and Human Values*, *41*(4), 732–761. https://doi.org/10.1177/0162243915626989

Felt, U., Wynne, B., Gonçalves, M. E., Jasanoff, S., Callon, M., Jepsen, M., Joly, P.-B., Konopasek, Z., May, S., Neubauer, C., Rip, A., Siune, K., Stirling, A., & Tallacchini, M. (2007). Taking European knowledge society seriously - Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission. In *Science in Society: Dialogues and Scientific Responsibility*.

Fetterman, D. M., Wandersman, A., & Kaftarian, S. (2015). Empowerment evaluation is a systematic way of thinking: A response to Michael Patton. *Evaluation and Program Planning*, *52*, 10–14. https://doi.org/10.1016/j.evalprogplan.2015.03.004

Fielke, S., Nelson, T., Blackett, P., Bewsell, D., Bayne, K., Park, N., Rijswijk, K., &

Small, B. (2017). Hitting the bullseye: Learning to become a reflexive monitor in New Zealand. *Outlook on Agriculture*, *46*(2), 117–124. https://doi.org/10.1177/0030727017708490

Finch, T. L., Mair, F. S., O'donnell, C., Murray, E., & May, C. R. (2012). From theory to "measurement" in complex interventions: Methodological lessons from the development of an e-health normalisation instrument. *BMC Medical Research Methodology*, *12*(69), 16.

Fiorino, D. J. (1990). Citizen Participation and Environmental Risk: A Survey of Institutional Mechanisms. *Science, Technology & Human Values*, *15*(2), 226–243. https://doi.org/10.1177/016224399001500204

Fischer, A. R. H., Beers, P. J., Latesteijn, H. Van, Andeweg, K., Jacobsen, E., Mommaas, H., Van Trijp, H. C. M., & Veldkamp, A. (2012). Transforum system innovation towards sustainable food. A review. *Agronomy for Sustainable Development*, *32*(2), 595–608. https://doi.org/10.1007/s13593-011-0067-4

Fischer, F. (2009). *Democracy and expertise: Reorienting policy inquiry*. Oxford University Press . http://8.242.217.84:8080/jspui/handle/123456789/31709

Fischer, J., Gardner, T. A. T. A., Bennet, E. M., Balvanera, P., Biggs, R., Carpenter, R., Daw, T., Folke, C., Hill, R., Hughes, T. P. T. P., Bennett, E. M., Balvanera, P., Biggs, R., Carpenter, S., Daw, T., Folke, C., Hill, R., Hughes, T. P. T. P., Luthe, T., … Tenhunen, J. (2015). Advancing sustainability through mainstreaming a social-ecological systems perspective. *Current Opinion in Environmental Sustainability*, *14*, 144–149. https://doi.org/10.1016/j.cosust.2015.06.002

Fitzpatrick, J., Christie, C., & Mark, M. M. (2008). *Evaluation in Action: Interviews With Expert Evaluators*. SAGE Publications.

Flinders, M., & Buller, J. (2006). Depoliticisation: Principles, tactics and tools. *British Politics*, *1*(3), 293–318. https://doi.org/10.1057/palgrave.bp.4200016

Flinders, M., Wood, M., & Cunningham, M. (2016). The politics of co-production: risks, limits and polution. *Evidence and Policy*, *12*(2), 261–279.

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

Frame, B., & Brown, J. (2008). Developing post-normal technologies for sustainability. *Ecological Economics*, *65*(2), 225–241. https://doi.org/10.1016/J.ECOLECON.2007.11.010

Frantzeskaki, N., & Kabisch, N. (2016). Designing a knowledge co-production operating space for urban environmental governance—Lessons from Rotterdam, Netherlands and Berlin, Germany. *Environmental Science & Policy*, *62*, 90–98.

Friedland, R., & Alford, R. R. (1991). Bringing Society Back In: Symbols, Practices and Institutional Contradictions. In W. W. Powell & P. J. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis* (pp. 232–263).

Fullan, M., & Pomfret, A. (1977). Research on Curriculum and Instruction Implementation. *Review of Educational Research*, *47*(1), 335–397.

Funtowicz, S. O., & Ravetz, J. R. (1993). SCIENCE FOR THE POST-NORMAL AGE. *Futures*, *25*, 739–755.

Galarneau, J. (2002). Cultivating Communities of Practice: A Guide to Managing Knowledge. In *Project Management Journal*. https://doi.org/10.1177/875697280203300312

Gaziulusoy, A. I., Mcgrail, S., Chandler, P., Twomey, P., Ryan, C., Mcgrail, S., Chandler, P., & Twomey, P. (2016). Identifying and addressing challenges faced by transdisciplinary research teams in climate change research. *Journal of Cleaner Production*, *123*, 55–64. https://doi.org/10.1016/j.jclepro.2015.08.049

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems Insights about dynamics and change from sociology and institutional theory. *Research Policy*, *33*, 897–920. https://doi.org/10.1016/j.respol.2004.01.015

Geels, F. W. (2006). Multi-level perspective on system innovation: relevance for industrial transformation. In X. Olshoorn & A. J. Wieczorek (Eds.), *Understanding Industrial Transformation: Views from Different Disciplines* (pp. 163–186). Springer. http://dx.doi.org/10.1007/1-4020-4418-6

Geels, F. W. (2011). The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental Innovation and Societal Transitions*, *1*(1), 24–40. https://doi.org/10.1016/j.eist.2011.02.002

Geels, F. W. (2014). Regime Resistance against Low-Carbon Transitions: Introducing Politics and Power into the Multi-Level Perspective. *Theory, Culture & Society*, *31*(5), 21–40. https://doi.org/10.1177/0263276414531627

Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research Policy*, *36*(3), 399–417. https://doi.org/10.1016/j.respol.2007.01.003

Gianni, R., Pearson, J., & Reber, B. (Eds.). (2018). *Responsible Research and Innovation: from Concepts to Practices*. Routledge.

Gibbert, M., Välikangas, L., & Luistro-Jonsson, M. (2021). *The Sustainability Grand Challenge: A Wicked Learning Workbook - Google Boeken*. Routledge Taylor & Francis Group.

Gibbons, M, Limoges, C., Nowotny, H. S., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. Sage.

Gibbons, Michael. (1999). Science's new social contract with society. *Nature*, *402*(6761 SUPPL. 1). https://doi.org/10.1038/35011576

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California press.

Green, J., Roberts, H., Petticrew, M., Steinbach, R., Goodman, A., Jones, A., & Edwards, P. (2015). Integrating quasi-experimental and inductive designs in evaluation: A case study of the impact of free bus travel on public health. *Evaluation*, *21*(4), 391–406. https://doi.org/10.1177/1356389015605205

Gregory, A. (2000). Problematizing Participation: A Critical Review of Approaches To Participation in Evaluation Theory . *Evaluation*, *6*(2), 179–199. https://doi.org/10.1177/13563890022209208

Grin, J. (2020). 'Doing' system innovations from within the heart of the regime. *Journal of Environmental Policy and Planning*, *22*(5), 682–694. https://doi.org/10.1080/1523908X.2020.1776099

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: Managing the Responsible Emergence of Science and Innovation in Society* (pp. 119–142). John Wiley and Sons. https://doi.org/10.1002/9781118551424.ch7

Guba, E.G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation* (1st ed.). SAGE Publications Inc.

Guba, Egon G, & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Sage.

Guijt, I. (2010). Accountability and Learning: Exploding the Myth of Incompatibility between Accountability and Learning. In J. Ubels, N.-A. Acquaye-Baddoo, & A. Folwer (Eds.), *Capacity Development in Practice* (pp. 277–291). Earthscan.

Guston, D. H. D. H. (2001). Boundary organizations in environmental policy and science: an introduction. *Science, Technology & Human Values*, *26*(4), 399–408. https://atmos.washington.edu/~breth/PCC/guston_2001_BoundaryOrg.pdf

Hage, M, & Leroy, P. (2007). *Leidraad Stake- holderparticipatie voor het Milieu- en Natuurplanbureau Praktijkwijzer*. 20. http://www.rivm.nl/bibliotheek/rapporten/550032004.pdf

Hage, Maria, Leroy, P., & Petersen, A. C. (2010). Stakeholder participation in environmental knowledge production. *Futures*, *42*(3), 254–264. https://doi.org/10.1016/j.futures.2009.11.011

Hagendijk, R., & Irwin, A. (2006). Public deliberation and governance: Engaging with science and technology in contemporary Europe. *Minerva*, *44*(2), 167–184. https://doi.org/10.1007/s11024-006-0012-x

Hajer, M. (2003). Policy without Polity ? Policy Analysis and the Institutional Void. *Policy Sciences*, *36*, 175–195.

Hajer, M. (2011). The Energetic Society. In *Search of a Governance Philosophy for a Clean …*. http://www.zeeli.pro.br/wp-content/uploads/2012/06/PBL-Netherlands-Energetic_society-2011.pdf

Hajer, M. A., & Wagenaar, H. (2003). *Deliberative Policy Analysis: Understanding Governance in the Network Society*. Cambridge University Press.

Halffman, W. (2005). Science-policy boundaries: national styles? *Science and Public Policy*, *32*(6), 457–467.

Halffman, W., & Hoppe, R. (2005). Science/policy boundaries: A changing division of labour in Dutch expert policy advice. In S. Maasen & F. Weingart (Eds.), *Democratization of Expertise? Exploring novel forms of scientific advice in political decision-making*. Springer.

Halffman, Willem. (2009). Measuring the Stakes: The Dutch Planning Bureaus. In J. Lentsch & P. Weingart (Eds.), *Scientific Advice to Policy Making: International Comparison* (pp. 1–25). Barbara Budrich Publishers.

Halligan, J. (1995). Policy advice and the public service. In B. Guy Peters & D. T. Savoie (Eds.), *Governance in a changing environment* (pp. 138–172). McGill-Queen's University Press.

Hansen, H. F. (2005). Choosing Evaluation Models: A Discussion on Evaluation Design. *Evaluation*, *11*(4), 447–462. https://doi.org/10.1177/1356389005060265

Hansson, S., & Polk, M. (2018). Assessing the impact of transdisciplinary research: The usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact. *Research Evaluation*, *27*(2), 132–144. https://doi.org/10.1093/reseval/rvy004

Havelock, R. G. (1970). *A Guide to Innovation in Education*. University of Michigan, Center for Research on Utilization of Scientific Knowledge.

Havelock, R. G. (1971). *Planning for innovation: through dissemination and utilization of knowledge*. University of Michigan, Center for Research on Utilization of Scientific Knowledge.

Havelock, Ronald G. (1979). *Planning for innovation through dissemination and utilization of knowledge*. Center for Research on Utilization of Scientific Knowledge, Institute for ….

Hawe, P., Shiell, A., & Riley, T. (2009). Theorising interventions as events in systems. *American Journal of Community Psychology*, *43*(3–4), 267–276. https://doi.org/10.1007/s10464-009-9229-9

Hegger, D., & Dieperink, C. (2014). Toward successful joint knowledge production for climate change adaptation: Lessons from six regional projects in the Netherlands. *Ecology and Society*. https://doi.org/10.5751/ES-06453-190234

Hegger, D., Lamers, M., Van Zeijl-Rozema, A., & Dieperink, C. (2012). Conceptualising joint knowledge production in regional climate change adaptation projects: success conditions and levers for action. *Environmental Science & Policy*, *18*, 52–65. https://doi.org/10.1016/j.envsci.2012.01.002

Heinrichs, H. (2015). Advisory systems in pluralistic knowledge societies: a criteria-based typology to assess and optimize environmental policy advice. In S. Maasen & P. Weingart (Eds.), *Democratization of Expertise? Exploring Novel Forms of Scientific Advice in Political Decision-Making* (pp. 41–61). Springer.

Hellström, T. (2015). Formative evaluation at a transdisciplinary research center. In M. Polk (Ed.), *Co-producing Knowledge for Sustainable Cities - Joining Forces for Change* (1st ed., pp. 162–181). Routledge.

Hendriks, C. M., & Grin, J. (2007). Contextualizing Reflexive Governance: the Politics of Dutch Transitions to Sustainability. *Journal of Environmental Policy & Planning*, *9*(3–4), 333–350. https://doi.org/10.1080/15239080701622790

Heras, M., & Ruiz-Mallén, I. (2017). *International Journal of Science Education Responsible research and innovation indicators for science education assessment: how to measure the impact?* https://doi.org/10.1080/09500693.2017.1392643

Hermann, A. T., Pregernig, M., Hogl, K., & Bauer, A. (2015). Cultural Imprints on Scientific Policy Advice: Climate science-policy interactions within Austrian neo-corporatism. *Environmental Policy and Governance*, *25*(5), 343–355. https://doi.org/10.1002/eet.1674

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. In *Nature* (Vol. 520, Issue 7548, pp. 429–431). Nature Publishing Group. https://doi.org/10.1038/520429a

Hirsch Hadorn, G., Pohl, C., Hoffmann-Riem, H., Biber-Klemm, S., Wiesmann, U., Grossenbacher-Mansuy, W., Zemp, E., & Joye, D. (2008). Handbook of transdisciplinary research. In G. Hirsch Hadorn, C. Pohl, H. Hoffmann-Riem, S. Biber-Klemm, U. Wiesmann, W. Grossenbacher-Mansuy, E. Zemp, & D. Joye (Eds.), *Handbook of Transdisciplinary Research*. https://doi.org/10.1007/978-1-4020-6699-3

Hoes, A.-C. C., & Regeer, B. J. (2015). Adoption of Novelties in a Pluralist Society: Exploring an Agropark Case Study. *Journal of Environmental Policy and Planning*, *17*(1), 3–24. https://doi.org/10.1080/1523908X.2014.880635

Hoffmann, S., Pohl, C., & Hering, J. G. (2017). Exploring transdisciplinary integration within a large research program: Empirical lessons from four thematic synthesis processes. *Research Policy*, *46*(3), 678–692. https://doi.org/10.1016/j.respol.2017.01.004

Houppermans, M. (2018). Operatie Inzicht in Kwaliteit: dwing de cultuuromslag af met een vernieuwde RPE. *Beleidsonderzoek Online*, *0*(2). https://doi.org/10.5553/bo/221335502018000002001

Hovik, S., & Hanssen, G. S. (2015). The impact of network management and complexity on multi-level coordination. *Public Administration*, *93*(2), 506–523. https://doi.org/10.1111/padm.12135

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huberman, A. M., & Miles, M. B. (1984). *Innovation up Close: How School Improvement Works*. Springer US.

Huberman, A. M., & Miles, M. B. (2013). *Innovation up close: How school improvement works*. Springer Science & Business Media.

Huitema, D., & Turnhout, E. (2009a). Working at the science-policy interface: A discursive analysis of boundary work at the Netherlands environmental assessment agency. *Environmental Politics*, *18*(4), 576–594. https://doi.org/10.1080/09644010903007427

Huitema, D., & Turnhout, E. (2009b). Working at the science–policy interface: a discursive analysis of boundary work at the Netherlands Environmental Assessment Agency. *Environmental Politics*, *18*(4), 576–594.

Iatridis, K., & Schroeder, D. (2016). *Responsible Research and Innovation in Industry - The Case for Corporate Responsibility Tools*. Springer.

Jagannathan, K., Arnott, J. C., Wyborn, C., Klenk, N., Mach, K. J., Moss, R. H., & Sjostrom, K. D. (2020). Great expectations? Reconciling the aspiration, outcome, and possibility of co-production. *Current Opinion in Environmental Sustainability*, *42*, 22–29. https://doi.org/10.1016/j.cosust.2019.11.010

Jahn, T., Bergmann, M., & Keil, F. (2012). Transdisciplinarity: Between mainstreaming and marginalization. *Ecological Economics*, *79*, 1–10. https://www.sciencedirect.com/science/article/pii/S0921800912001681

Jahn, T., & Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. *Futures*, *65*, 195–208. https://doi.org/10.1016/j.futures.2014.10.015

James, T. E., & Jorgensen, P. D. (2009). Policy knowledge, policy formulation, and change: Revisiting a foundational question. *Policy Studies Journal*, *37*(1), 141–162. https://doi.org/10.1111/j.1541-0072.2008.00300.x

Jasanoff, S. (2003). TECHNOLOGIES OF HUMILITY: CITIZEN PARTICIPATION IN GOVERNING SCIENCE. *Minerva*, *41*, 223–244.

Jasanoff, S. (2004a). Science in Culture and Politics. In *States of Knowledge. The co-production of science and social order.* (pp. 25–98). Routledge.

Jasanoff, S. (2004b). *States of knowledge: the co-production of science and the social order*. Routledge.

Jasanoff, S. (2011). Quality control and peer review in advisory science. In *The Politics of Scientific Advice: Institutional Design for Quality Assurance* (pp. 19–35). Cambridge University Press. https://doi.org/10.1017/CBO9780511777141.002

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

Jones, C., Boxenbaum, E., & Anthony, C. (2015). The immateriality of material practices in institutional logics. In E. Boxenbaum & M. Lounsbury (Eds.), *Institutional Logics in Action, Part A (Research in the Sociology of Organizations)* (Vol. 39, Issue A, pp. 51–75). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2013)0039A

Jones, C. H. D., Glogowska, M., Locock, L., & Lasserson, D. S. (2016). *Embedding new technologies in practice-a normalization process theory study of point of care testing*. https://doi.org/10.1186/s12913-016-1834-3

Karlsson, A., & Björk, J. (2017). Establishing and managing a network for continuous innovation: Invoking organizational pressure. *Creativity and Innovation Management*, *26*(2), 128–141. https://doi.org/10.1111/caim.12215

Kawabe, K., Horiuchi, F., Ochi, M., Oka, Y., & Ueno, S. ichi. (2016). Internet addiction: Prevalence and relation with mental states in adolescents. *Psychiatry and Clinical Neurosciences*, *70*(9), 405–412. https://doi.org/10.1111/pcn.12402

Kemmis, S., & McTaggart, R. (1982). *The Action Research Reader*. Deakin University Press.

Kemmis, S., & McTaggart, R. (2000). Participatory action research: Handbook of qualitative research. *Thousand Oaks, CA, Sage Publications*, 567–605.

Kemp, R., Parto, S., & Gibson, R. B. (2005). Governance for sustainable development: moving from theory to practice. *International Journal of Sustainable Development*, *8*(1/2), 12. https://doi.org/10.1504/IJSD.2005.007372

Keune, M. (2016). Nog steeds een mirakel? De legitimiteit van het poldermodel in de eenentwintigste eeuw. In *Nog steeds een mirakel? De legitimiteit van het poldermodel in de eenentwintigste eeuw*. https://doi.org/10.5117/9789089647092

Kirkhart, K. E. (2000). Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation*, *88*, 5.

Klaassen, P., Rijnen, M., Vermeulen, S., Kupper, F., & Broerse, J. (2018). Technocracy versus experimental learning in RRI. In R. Gianni, I, J. Pearson, & B. Reber (Eds.), *Responsible Research and Innovation* (pp. 77–98). Routledge. https://doi.org/10.4324/9781315457291-5

Klaassen, P., Verwoerd, L., Kupper, F., & Regeer, B. J. (2021). Reflexive Monitoring in Action as a methodology for learning and enacting Responsible Research and Innovation. In E. Yaghmaei & I. Van de Poel (Eds.), *Assessment of Responsible Innovation: methods and practices* (pp. 222–243). Routledge.

Klerkx, L., Aarts, N., & Leeuwis, C. (2010). Adaptive management in agricultural innovation systems: The interactions between innovation networks and their

environment. *Agricultural Systems*, *103*(6), 390–400. https://doi.org/10.1016/J.AGSY.2010.03.012

Klerkx, L., Seuneke, P., de Wolf, P., & Rossing, W. A. H. H. (2017). Replication and translation of co-innovation: The influence of institutional context in large international participatory research projects. *Land Use Policy*, *61*, 276–292. https://doi.org/10.1016/j.landusepol.2016.11.027

Klijn, E. H., Steijn, B., & Edelenbos, J. (2010). The Impact Of Network Management On Outcomes In Governance Networks. *Public Administration*, *88*(4), 1063–1082. https://doi.org/10.1111/j.1467-9299.2010.01826.x

Knottnerus, A., Van Asselt, M., Berkhout, F., Dierikx, M., Dijkstra, L., Feind, P., Vet, L., & Van Oosteren, C. (2017). *Kennisontwikkeling voor een duurzame samenleving: Bevindingen visitatiecommissie Planbureau voor de Leefomgeving 2017*. External Inspection Committee PBL.

Kok, K. P. W., Loeber, A. M. C., & Grin, J. (2021). Politics of complexity: Conceptualizing agency, power and powering in the transitional dynamics of complex adaptive systems. *Research Policy*, *50*(3), 104183. https://doi.org/10.1016/j.respol.2020.104183

Kowalczewska, K., & Behagel, J. (2019). How policymakers' demands for usable knowledge shape science-policy relations in environmental policy in Poland. *Science and Public Policy*, *46*(3), 381–390. https://doi.org/10.1093/scipol/scy065

Kueffer, C., Underwood, E., Hadorn, G. H., Holderegger, R., Lehning, M., Pohl, C., Schirmer, M., Schwarzenbach, R., Stauffacher, M., Wuelser, G., & Edwards, P. (2012). Enabling effective problem-oriented research for sustainable development. *Ecology and Society*, *17*(4). https://doi.org/10.5751/ES-05045-170408

Kunseler, E.-M. (2017). *Government expert organisations in-between logics*. VU University Amsterdam.

Kunseler, E.-M., & Tuinstra, W. (2017). Navigating the authority paradox: practising objectivity in environmental experise. *Environmental Science & Policy*, *67*, 1–7. https://doi.org/10.1016/J.ENVSCI.2016.10.001

Kunseler, E.-M., & Vasileiadou, E. (2016). Practising environmental policy evaluation under co-existing evaluation imaginaries. *Evaluation*, *22*(4), 451–469. https://doi.org/10.1177/1356389016668099

Kunseler, E.-M., & Verwoerd, L. (2019). *Kennis met impact: Reflexief werken bij publieke kennisorganisaties*. PBL.

Kunseler, E.-M., Verwoerd, L., & Verwest, F. (2020). Inzet op omgevingsbewust en kwaliteitsbewust werken in beleidsonderzoek. *Beleidsonderzoek Online*, *0*(9). https://doi.org/10.5553/bo/221335502020000009001

Kupper, J. F. H., Klaassen, P., Rijnen, M. C. J. A., Vermeulen, S., & Broerse, J. E. W.

(2015). *Report on the quality criteria of Good Practice Standards in RRI* . https://research.vu.nl/en/publications/report-on-the-quality-cri-teria-of-good-practice-standards-in-rri

Lahsen, M., & Turnhout, E. (2021). How norms, needs, and power in science obstruct transformations towards sustainability. *Environmental Research Letters*, *16*(2). https://doi.org/10.1088/1748-9326/abdcf0

Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Swilling, M., & Thomas, C. J. (2012). Transdisciplinary research in sustainability science: Practice, principles, and challenges. *Sustainability Science*, *7*(SUPPL. 1), 25–43. https://doi.org/10.1007/s11625-011-0149-x

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society: LaTour, Bruno: Amazon.nl*. https://www.amazon.nl/Science-Action-Scientists-Engineers-Through/dp/0674792912

Lawhon, M., Ernstson, H., & Silver, J. (2014). Provincializing Urban Political Ecology: Towards a Situated UPE Through African Urbanism. *Antipode*, *46*(2), 497–516. https://doi.org/10.1111/anti.12051

Laws, D., Hogendoorn, D., & Karl, H. (2014). Hot adaptation: What conflict can contribute to collaborative natural resource management. *Ecology and Society*, *19*(2). https://doi.org/10.5751/ES-06375-190239

Leeuw, F. L., & Furubo, J. E. (2008). Evaluation systems: What are they and why study them? *Evaluation*, *14*(2). https://doi.org/10.1177/1356389007087537

Lehtonen, M. (2014). Evaluating megaprojects: From the “iron triangle” to network mapping. *Evaluation*, *20*(3), 278–295. https://doi.org/10.1177/1356389014539868

Lemos, M. C., Arnott, J. C., Ardoin, N. M., Baja, K., Bednarek, A. T., Dewulf, A., Fieseler, C., Goodrich, K. A., Jagannathan, K., Klenk, N., Mach, K. J., Meadow, A. M., Meyer, R., Moss, R., Nichols, L., Sjostrom, K. D., Stults, M., Turnhout, E., Vaughan, C., … Wyborn, C. (2018). To co-produce or not to co-produce. *Nature Sustainability*, *1*(12), 722–724. https://doi.org/10.1038/s41893-018-0191-0

Lentsch, Justus, & Weingart, P. (2011). The politics of scientific advice: Institutional design for quality assurance. In J. Lentsch & P. Weingart (Eds.), *The Politics of Scientific Advice: Institutional Design for Quality Assurance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511777141

Leventon, J., Fleskens, L., Claringbould, H., Schwilch, G., & Hessel, R. (2016). An applied methodology for stakeholder identification in transdisciplinary research. *Sustainability Science*. https://doi.org/10.1007/s11625-016-0385-1

Li, Y., & Wagenaar, H. (2019). Revisiting deliberative policy analysis. *Policy Studies*, *40*(5), 427–436. https://doi.org/10.1080/01442872.2019.1618813

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE Publications.

Lo, C. (2018). Between Government and Governance: Opening the Black Box of the Transformation Thesis. *International Journal of Public Administration*, *41*(8), 650–656. https://doi.org/10.1080/01900692.2017.1295261

Loeber, A., Van Mierlo, B., Grin, B., & Leeuwis, C. (2007). The practical value of theory: conceptualising learning in the pursuit of a sustainable development. In A. E. J. Wals (Ed.), *Social learning towards a sustainable world* (pp. 83–98). Wageningen Academic Publishers.

Loeber, Anne, Versteeg, W., & Griessler, E. (2011). Stop looking up the ladder: analyzing the impact of participatory technology assessment from a process perspective. *Science and Public Policy*, *38*(8), 599–608.

Lund Declaration. (2009). Conference: New worlds - New solutions. *Research and Innovation as a Basis for Develpoing Europe in a Global Context*.

Lux, A., Schäfer, M., Bergmann, M., Jahn, T., Marg, O., Nagy, E., Ransiek, A. C., & Theiler, L. (2019). Societal effects of transdisciplinary sustainability research—How can they be strengthened during the research process? *Environmental Science and Policy*, *101*, 183–191. https://doi.org/10.1016/j.envsci.2019.08.012

Mach, K. J., Lemos, M. C., Meadow, A. M., Wyborn, C., Klenk, N., Arnott, J. C., Ardoin, N. M., Fieseler, C., Moss, R. H., Nichols, L., Stults, M., Vaughan, C., & Wong-Parodi, G. (2020). Actionable knowledge and the art of engagement. *Current Opinion in Environmental Sustainability*, *42*(June 2019), 30–37. https://doi.org/10.1016/j.cosust.2020.01.002

Mair, F. S., May, C. R., O'donnell, C., Finch, T., Sullivan, F., & Murray, E. (2012). Systematic reviews Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. *Bull World Health Organ*, *90*, 357–364. https://doi.org/10.2471/BLT.11.099424

Marjanovic, S., Cochrane, G., Robin, E., Sewankambo, N., Ezeh, A., Nyirenda, M., Bonfoh, B., Rweyemamu, M., & Chataway, J. (2017). Evaluating a complex research capacity-building intervention: Reflections on an evaluation of the African Institutions Initiative. *Evaluation*, *23*(1), 80–101. https://doi.org/10.1177/1356389016682759

Mauser, W., Klepper, G., Rice, M., Schmalzbauer, B. S., Hackmann, H., Leemans, R., & Moore, H. (2013). Transdisciplinary global change research: the co-creation of knowledge for sustainability. *Current Opinion in Environmental Sustainability*, *5*(3–4), 420–431. https://doi.org/10.1016/j.cosust.2013.07.001

May, C., Rapley, T., Mair, F.S., Treweek, S., Murray, E., Ballini, L., Macfarlane, A., Girling, M., Finch, T. L. (2015). *Normalization Process Theory On-Line Users' Manual, Toolkit and NoMAD instrument*. Normalization Process Theory.

May, C. R. (2013). Towards a general theory of implementation. *Implementation Science*, *8*(1), 18. https://doi.org/10.1186/1748-5908-8-18

May, C. R. (2015). Applying Normalization Process Theory to the Implementation of Complex Interventions. In D. A. Richards & I. R. Hallberg (Eds.), *Complex interventions in health - an overview of research methods*. Routledge.

May, C. R., Cummings, A., Girling, M., Bracher, M., Mair, F. S., May, C. M., Murray, E., Myall, M., Rapley, T., & Finch, T. (2018). Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. *Implementation Science*, *13*(80), 1–27. https://doi.org/10.1186/s13012-018-0758-1

May, C. R., & Finch, T. (2009). Implementing, embedding, and integrating practices: An outline of normalization process theory. *Sociology*, *43*(3), 535–554. https://doi.org/10.1177/0038038509103208

May, C. R., Finch, T., Ballini, L., MacFarlane, A., Mair, F., Murray, E., Treweek, S., & Rapley, T. (2011). Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. *BMC Health Services Research*, *11*(1), 245. https://doi.org/10.1186/1472-6963-11-245

May, C. R., Johnson, M., & Finch, T. (2016). Implementation, context and complexity. *Implementation Science*, *11*(1), 141. https://doi.org/10.1186/s13012-016-0506-3

Mazzucato, M. (2018). *MISSIONS: Mission-oriented Research & Innovation in the European Union*. https://publications.europa.eu/en/publication-detail/-/publication/5b2811d1-16be-11e8-9253-01aa75ed71a1/language-en

McEvoy, R., Ballini, L., Maltoni, S., O'Donnell, C.A., Mair, F.S., MacFarlane, A. (2014). A qualitative systematic review of studies using the Normalization Process Theory in research implementation processes. *Implementation Science*, *9*(2), 1–22.

McNaughton, R. J., Steven, A., & Shucksmith, J. (2020). Using Normalization Process Theory as a practical tool across the life course of a qualitative research project. *Qualitative Health Research*, *30*(3), 217–227. https://doi.org/10.1177/1049732319863420

Miller, C. A., & Wyborn, C. (2020). Co-Production in Global Sustainability: Histories and Theories. *Environmental Science and Policy*, *113*(February 2018), 88–95. https://doi.org/10.1016/j.envsci.2018.01.016

Natuurpact: ontwikkeling en beheer van natuur in Nederland, 1 (2013) (testimony of Ministerie van Economische Zaken). https://doi.org/10.1007/s13398-014-0173-7.2

Mitchell, C., Cordell, D., & Fam, D. (2015). Beginning at the end: The outcome spaces framework to guide purposive transdisciplinary research. *Futures*, *65*, 86–96. https://doi.org/10.1016/j.futures.2014.10.007

Mommaas, H., & Eweg, R. (2011). Organizing innovations and transitions: Toward a more sustainable fit between innovation strategies and the institutional

environment. In H. C. Van Latesteijn & K. Andeweg (Eds.), *The TransForum Model: Transforming Agro Innovation Toward Sustainable Development* (pp. 1–162). https://doi.org/10.1007/978-90-481-9781-1

Moore, G. F., Evans, R. E., Hawkins, J., Littlecott, H., Melendez-Torres, G. J., Bonell, C., & Murphy, S. (2019). From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation (London, England : 1995)*, *25*(1), 23–45. https://doi.org/10.1177/1356389018803219

Murray, E., Burns, J., May, C. R., Finch, T., O'donnell, C., Wallace, P., & Mair, F. (2011). *Why is it difficult to implement e-health initiatives? A qualitative study*. https://doi.org/10.1186/1748-5908-6-6

Murray, E., Treweek, S., Pope, C., MacFarlane, A., Ballini, L., Dowrick, C., Finch, T., Kennedy, A., Mair, F., Nio Ong, B., Rapley, T., Rogers, A., & May, C. R. (2010). *Normalisation process theory: a framework for developing, evaluating and implementing complex interventions*. https://doi.org/10.1186/1741-7015-8-63

Nederhand, J., Klijn, E. H., Van der Steen, M., & van Twist, M. (2019). The governance of self-organization: Which governance strategy do policy officials and citizens prefer? *Policy Sciences*, *52*(2), 233–253. https://doi.org/10.1007/s11077-018-9342-4

Nederhand, J., Van Der Steen, M., & Van Twist, M. (2019). Boundary-spanning strategies for aligning institutional logics: a typology. *Local Government Studies*, *45*(2), 219–240. https://doi.org/10.1080/03003930.2018.1546172

Ney, S., & Verweij, M. (2015). Messy institutions for wicked problems: How to generate clumsy solutions? *Environment and Planning C: Government and Policy*, *33*(6), 1679–1696.

Nicolini, D. (2012). *Practice Theory, Work, and Organization: An Introduction* (First Edit). Oxford University Press.

Nieminen, M., & Hyytinen, K. (2015). Future-oriented impact assessment: Supporting strategic decision-making in complex socio-technical environments. *Evaluation*, *21*(4), 448–461. https://doi.org/10.1177/1356389015606540

Noble, H., & Heale, R. (2019). Triangulation in research, with examples. *Evidence-Based Nursing*, *22*(3), 67–68. https://doi.org/10.1136/ebnurs-2019-103145

Nordmann, A. (2014). Responsible innovation, the art and craft of anticipation. *Journal of Responsible Innovation*, *1*(1), 87–98. https://doi.org/10.1080/23299460.2014.882064

Norström, A. V, Cvitanovic, C., Löf, M. F., West, S., Wyborn, C., Balvanera, P., Bednarek, A. T., Bennett, E. M., Biggs, R., de Bremond, A., Campbell, B. M., Canadell, J. G., Carpenter, S. R., Folke, C., Fulton, E. A., Gaffney, O., Gelcich, S., Jouffray, J.-B., Leach, M., … Österblom, H. (2020). Principles for knowledge co-

production in sustainability research. *Nature Sustainability*, *3*(3), 182–190. https://doi.org/10.1038/s41893-019-0448-2

Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Polity Press.

O'connor, R. A., Nel, J. L., Roux, D. J., Lim-Camacho, L., Van Kerkhoff, L., & Leach, J. (2019). *Principles for evaluating knowledge co-production in natural resource management: Incorporating decision-maker values*. https://doi.org/10.1016/j.jenvman.2019.109392

O'Sullivan, R. G. (2012). Collaborative Evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and Program Planning*, *35*(4), 518–522. https://doi.org/10.1016/j.evalprogplan.2011.12.005

Oftedal, G. (2014). The role of philosophy of science in Responsible Research and Innovation (RRI): the case of nanomedicine. *Life Sciences, Society and Policy*, *10*(1), 1–12. https://doi.org/10.1186/s40504-014-0005-8

Oliver, K., Kothari, A., & Mays, N. (2019). The dark side of coproduction: do the costs outweigh the benefits for health research? *Health Research Policy and Systems*, *17*(1), 33. https://doi.org/10.1186/s12961-019-0432-3

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: Managing the Responsible Emergence of Science and Innovation in Society* (pp. 27–50). John Wiley.

Owens, S., Rayner, T., & Bina, O. (2004). New agendas for appraisal: Reflections on theory, practice, and research. In *Environment and Planning A* (Vol. 36, Issue 11, pp. 1943–1959). https://doi.org/10.1068/a36281

Pallett, H., & Chilvers, J. (2013). A decade of learning about publics, participation, and climate change: Institutionalising reflexivity? *Environment and Planning A*, *45*(5), 1162–1183. https://doi.org/10.1068/a45252

Pallett, H., & Chilvers, J. (2015). Organizations in the making: Learning and intervening at the science-policy interface. *Environment and Planning A*, *39*(5), 146–166. https://doi.org/10.1068/a45252

Patton, M. Q. (2010). *Developmental Evaluation: Applying Complexity concepts to Enhance Innovation and Use*. Guilford Press.

Paylor, J., & McKevitt, C. (2019). The Possibilities and Limits of "Co-producing" Research. *Frontiers in Sociology*, *4*, 23. https://doi.org/10.3389/fsoc.2019.00023

PBL. (2019). *About PBL*. PBL Netherlands Environmental Assessment Agency.

https://www.pbl.nl/en/aboutpbl

PBL & WUR. (2017). *Lerende evaluatie van het Natuurpact: Naar nieuwe verbindingen tussen natuur, beleid en samenleving*. PBL.

PBL & WUR. (2020). Lerende evaluatie van het Natuurpact 2020 - gezamenlijk de puzzel leggen voor natuur, economie en maatschappij. In *Lerende evaluatie van het Natuurpact 2020: Gezamenlijk de puzzel leggen voor natuur, economie en maatschappij*. PBL.

Pesch, U., Huitema, D., & Hisschemöller, M. (2012). A boundary organization and its changing environment: the Netherlands Environmental Assessment Agency, the MNP. *Environment and Planning C: Government and Policy*, *30*, 487–503. https://doi.org/10.1068/c10150j

Peter, V., Maier, F., Mejlgaard, N., Blocck, C. W., Madsen, E. B., Griessler, E., Wuketich, M., Meijer, I., Woolley, R., Lindner, R., & Bührer, S. (2018). *Monitoring the evolution and benefits of responsible research and innovation in Europe Summarising insights from the MoRRI project Monitoring the evolution and benefits of responsible research and innovation in Europe-Summarising insights from the MoRRI project*. http://europa.eu

Petersen, A. C., Cath, A., Hage, M., Kunseler, E.-M., & Van der Sluijs, J. P. (2011). Post-Normal Science in Practice at the Netherlands Environmental Assessment Agency. *Science, Technology & Human Values*, *36*(3), 362–388. https://doi.org/10.1177/0162243910385797

Pielke, R. A. J. (2007). Four Idealized Roles of Science in Policy and Politics. In *The Honest Broker: Making Sense of Science in Policy and Politics* (pp. 1–7).

Pieterse, E. (2008). *City Futures: Confronting the Crisis of Urban Development*. Zed Books.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, *14*(3), 399–441.

Pohl, C. (2008). From science to policy through transdisciplinary research. *Environmental Science & Policy*, *11*(1), 46–53. https://doi.org/10.1016/j.envsci.2007.06.001

Pohl, C. (2011). What is progress in transdisciplinary research? *Futures*, *43*(6), 618–626. https://doi.org/10.1016/j.futures.2011.03.001

Pohl, C., Rist, S., Zimmermann, A., Fry, P., Gurung, G. S., Schneider, F., Speranza, C. I., Kiteme, B., Boillat, S., Serrano, E., Hadorn, G. H., & Wiesmann, U. (2010). Researchers' roles in knowledge co-production: experience from sustainability research in Kenya, Switzerland, Bolivia and Nepal. *Science and Public Policy*, *37*(4), 267–281. https://doi.org/10.3152/030234210X496628

Polk, M. (2014). *Achieving the promise of transdisciplinarity: a critical exploration of the relationship between transdisciplinary research and societal problem solving*.

https://doi.org/10.1007/s11625-014-0247-7

Polk, M. (2015). Transdisciplinary co-production: Designing and testing a transdisciplinary research framework for societal problem solving. *Futures*, *65*, 110–122. https://doi.org/10.1016/j.futures.2014.11.001

Pollit, C., Van Thiel, S., & Homburg, V. (2007). *New Public Management in Europe: Adaptation and Alternatives*. https://doi.org/10.1057/9780230625365

Popa, F., & Guillermin, M. (2017). Reflexive Methodological Pluralism. *Journal of Mixed Methods Research*, *11*(1), 19–35. https://doi.org/10.1177/1558689815610250

Pyöriä, P. (2005). The Concept of Knowledge Work Revisited. *Article in Journal of Knowledge Management*. https://doi.org/10.1108/13673270510602818

Rau, H., Goggins, G., & Fahy, F. (2018). From invisibility to impact: Recognising the scientific and societal relevance of interdisciplinary sustainability research. *Research Policy*, *47*, 266–276. https://doi.org/10.1016/j.respol.2017.11.005

Raven, R., Van den Bosch, S., & Weterings, R. (2007). Strategic Niche Management and Transition Experiments. From analytical tool to a competence kit for practitioners. *4th Dubrovnik Conference on Sustainable Development of Energy Water and Environment Systems*.

Reckwitz, A. (2002). Toward a Theory of Social Practices A Development in Culturalist Theorizing. *European Journal of Social Theory*, *5*(2), 243–263.

Regeer, B. J., & Bunders, J. F. G. J. G. F. (2003). The epistemology of transdisciplinary research: from knowledge integration to communities of practice. *Interdisciplinary Environmental Review*, *5*(2), 98. https://doi.org/10.1504/ier.2003.053901

Regeer, B. J., & Bunders, J. G. F. F. G. (2009). Knowledge co-creation: Interaction between science and society. In *A Transdisciplinary Approach to Complex Societal Issues. Den Haag: Advisory Council for Research on Spatial Planning, Nature and the Environment/Consultative Committee of Sector Councils in the Netherlands [RMNO/COS]*. Advisory Council for Spatial Planning, Nature and the Environment (RMNO).

Regeer, B. J., de Wildt-Liesveld, R., Van Mierlo, B., & Bunders, J. F. G. (2016). Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation*, *22*(1), 6–28. https://doi.org/10.1177/1356389015623659

Regeer, B. J., Hoes, A.-C., Van Amstel-Van Saane, M., Caron-Flinterman, F. F., & Bunders, J. F. G. (2009). Six Guiding Principles for Evaluating Mode-2 Strategies for Sustainable Development. *American Journal of Evaluation*, *30*(4), 515–537. https://doi.org/10.1177/1098214009344618

Reinecke, S. (2015). Knowledge brokerage designs and practices in four european

climate services: A role model for biodiversity policies? *Environmental Science and Policy*, *54*, 513–521. https://doi.org/10.1016/j.envsci.2015.08.007

Ribeiro, B. E., Smith, R. D. J., & Millar, K. (2016). A mobilising concept? Unpacking academic representations of Responsible Research and Innovation. *Science and Engineering Ethics* , 1–32. https://doi.org/10.1007/s11948-016-9761-6

Ribeiro, C. D. S., Van de Burgwal, L. H. M., & Regeer, B. J. (2019). Overcoming challenges for designing and implementing the One Health approach: A systematic review of the literature. *One Health*, *7*, 1–19. https://doi.org/10.1016/J.ONEHLT.2019.100085

Rijswijk, K., Bewsell, D., Small, B., & Blackett, P. (2015). Reflexive monitoring in New Zealand: Evaluation lessons in supporting transformative change. *Evaluation Journal of Australasia*, *15*(4), 38–43. https://doi.org/10.1177/1035719X1501500405

Rip, A. (2014). The past and future of RRI. *Life Sciences, Society and Policy*, *10*(1), 1–15. https://doi.org/10.1186/s40504-014-0017-4

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in general theory of planning. *Policy Sciences*, *4*, 155–169.

Rog, D. J., Fitzpatrick, J. L., & Conner, R. F. (2012). Context: A Framework for Its Influence on Evaluation Practice. *New Directions for Evaluation*, *135*.

Roger, S., Spaapen, J., Bauer, M., Hogan, E., Revuelta, G., & Stagl, S. (2015). *Indicators for promoting and monitoring responsible research and innovation: report from the expert group on policy indicators for responsible research and innovation* (Issue EUR 26866 EN). https://doi.org/10.2777/9742

Roitman, D.B., & Mayer, J. P. (1982). Fidelity and Reinvention in the Implementation of Innovations. *90th Annual Convention of the American Psychological Association*, 21.

Roitman, David B, & Mayer, J. P. (1982). *Fidelity and Reinvention in the Implementation of Innovations*.

Rolfe, S. (2019). Combining Theories of Change and Realist Evaluation in practice: Lessons from a research on evaluation study. *Evaluation*, *25*(3), 294–316. https://doi.org/10.1177/1356389019835229

Roux, D. J., Nel, J. L., Cundill, G., O'Farrell, P., & Fabricius, C. (2017). Transdisciplinary research for systemic change: who to learn with, what to learn about and how to learn. *Sustainability Science*, *c*, 1–16. https://doi.org/10.1007/s11625-017-0446-0

Roux, D. J., Stirzaker, R. J., Breen, C. M., Lefroy, E. C., & Cresswell, H. P. (2010). Framework for participative reflection on the accomplishment of transdisciplinary research programs. *Environmental Science & Policy*, *13*(8), 733–741. https://doi.org/10.1016/j.envsci.2010.08.002

Sanderson, I. (2002). Evaluation , Policy Learning and Evidence-based Policy. *Public Administration*, *80*(1), 1–22. https://doi.org/10.1111/1467-9299.00292

Sarewitz, D. (2004). How science makes environmental controversies worse. *Environmental Science & Policy*, *7*(5), 385–403. https://doi.org/10.1016/j.envsci.2004.06.001

Sarkki, S., Heikkinen, H. I., & Karjalainen, T. P. (2013). Sensitivity in transdisciplinary projects: A case of reindeer management in Finland. *Land Use Policy*, *34*, 183–192. https://doi.org/10.1016/j.landusepol.2013.03.004

Schäfer, M., Lux, A., & Bergmann, M. (2020). Editorial to the special issue "Transdisciplinary Sustainability Research—Linking research processes and outputs to societal effects." *Environmental Science and Policy*, *107*(February), 206–210. https://doi.org/10.1016/j.envsci.2020.02.018

Schatzki, T. R. (2002). *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*. Pennsylvania State University Press. https://www.jstor.org/stable/10.5325/j.ctt7v38n

Schatzki, T. R. (2016). Practice theory as flat ontology. In G. Spaargaren, D. Weenink, & M. Lamers (Eds.), *Practice theory and research: Exploring the dynamics of social life* (pp. 28–42). Routledge.

Schatzki, T. R., Knorr-Cetina, K., & von Savigny, E. (2001). *The Practice Turn in Contemporary Theory* . Routledge. https://books.google.nl/books/about/The_Practice_Turn_in_Contemporary_Theory.html?id=RfaVpJBB5lgC&redir_esc=y

Scherhaufer, P. (2014). Bridging the Gap Between the Theory and Practices of Stakeholder Participation in Integrated Vulnerability Assessments of Climate Change. *Systemic Practice and Action Research*, *27*(5), 449–463. https://doi.org/10.1007/s11213-013-9294-8

Schmidt, L., Falk, T., Siegmund-Schultze, M., & Spangenberg, J. H. (2020). The Objectives of Stakeholder Involvement in Transdisciplinary Research. A Conceptual Framework for a Reflective and Reflexive Practise. *Ecological Economics*, *176*(June), 106751. https://doi.org/10.1016/j.ecolecon.2020.106751

Schmidt, R. (2016). The Methodological Challenges of Practising Praxeology. In M. Lamers, G. Spaargaren, & D. Weenink (Eds.), *Practice Theory and Research: Exploring the Dynamics of Social Life* (pp. 59–75). Routledge. https://doi.org/10.4324/978131565690-13

Schneider, F., Giger, M., Harari, N., Moser, S., Oberlack, C., Providoli, I., Schmid, L., Tribaldos, T., & Zimmermann, A. (2019). Transdisciplinary co-production of knowledge and sustainability transformations: Three generic mechanisms of impact generation. *Environmental Science and Policy*, *102*, 26–35. https://doi.org/10.1016/j.envsci.2019.08.017

Schneidewind, U., & Augenstein, K. (2012). Analyzing a transition to a sustainability-oriented science system in Germany. *Environmental Innovation and Societal Transitions*, *3*, 16–28. https://doi.org/10.1016/j.eist.2012.04.004

Schoenefeld, J. J., & Jordan, A. J. (2019). Environmental policy evaluation in the EU: between learning, accountability, and political opportunities? *Environmental Politics*, *28*(2), 365–384. https://doi.org/10.1080/09644016.2019.1549782

Scholz, R. W., & Steiner, G. (2015a). The real type and ideal type of transdisciplinary processes: part I—theoretical foundations. *Sustainability Science*, *10*(4). https://doi.org/10.1007/s11625-015-0326-4

Scholz, R. W., & Steiner, G. (2015b). The real type and ideal type of transdisciplinary processes: part II—what constrains and obstacles do we meet in practice? *Sustainability Science*, *10*(4), 527–544. https://doi.org/10.1007/s11625-015-0326-4

Scholz, R. W., & Tietje, O. (2002). *Embedded case study methods: Integrating quantitative and qualitative knowledge*. Sage.

Schön, D. A. (1983). *The Reflective Practitioner*. Basic Books Inc.

Schot, J., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management*, *20*(5), 537–554. https://doi.org/10.1080/09537320802292651

Schuijer, J. W. (2020). *Wading through the mud: Reflections on shaping RRI in practice*. VU University Amsterdam.

Schuitmaker-Warnaar, T. J., Gunn, C. J., Regeer, B. J., & Broerse, J. E. W. (2021). Institutionalizing reflexivity for sustainability: Two cases in health care. *Sustainability*, *13*(11712). https://doi.org/10.3390/su132111712

Schuitmaker, T. J. (2012). Identifying and unravelling persistent problems. *Technological Forecasting and Social Change*, *79*(6), 1021–1031. https://doi.org/10.1016/j.techfore.2011.11.008

Shils, E. A., & Finch, H. A. (1949). *Max Weber On The Methodology of Social Sciences* (First edit). The Free Press.

Shove, E. (2010). Beyond the ABC: climate change policy and theories of social change. *Environment and Planning A*, *42*(6), 1273–1285. https://doi.org/10.1068/a42282

Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2015). Introducing Evidence-Based Principles to Guide Collaborative Approaches to Evaluation: Results of an Empirical Process. *American Journal of Evaluation*, *37*(2), 193–215. https://doi.org/10.1177/1098214015615230

Smith, A., & Stirling, A. (2007). Moving outside or inside? Objectification and

reflexivity in the governance of socio-technical systems. *Journal of Environmental Policy and Planning*, *9*(3–4), 351–373. https://doi.org/10.1080/15239080701622873

Najaarsnota 2018: Motie van het lid Sneller c.s. over beleidsopties voor een volgende neergaande conjunctuur, Pub. L. No. 35095 (2018).

Spaargaren, Gert, Lamers, M., & Weenink, D. (2019). Introduction: using practice theory to research social life. In G. Spaargaren, D. Weenink, & M. Lamers (Eds.), *Practice Theory and Research: Exploring the Dynamics of Social Life* (pp. 19–43). Routledge. https://doi.org/10.4324/978131565690-11

Squires, J. E., Graham, I. D., Hutchinson, A. M., Michie, S., Francis, J. J., Sales, A., Brehaut, J., Curran, J., Ivers, N., Lavis, J., Linklater, S., Fenton, S., Noseworthy, T., Vine, J., & Grimshaw, J. M. (2015). Identifying the domains of context important to implementation science: A study protocol. *Implementation Science*, *10*(1). https://doi.org/10.1186/s13012-015-0325-y

Aanwijzingen voor de Planbureaus, Nr. 3200 (2012) (testimony of Staatscourant). https://wetten.overheid.nl/BWBR0031972/2012-04-01

Stern, E., Saunders, M., & Stame, N. (2015). Standing back and looking forward: Editors' reflections on the 20th Anniversary of Evaluation. *Evaluation*, *21*(4), 380–390. https://doi.org/10.1177/1356389015608757

Sternlieb, F., Bixler, R. P., Huber-Stearns, H., & Huayhuaca, C. (2013). A question of fit: Reflections on boundaries, organizations and social–ecological systems. *Journal of Environmental Management*, *130*, 117–125. https://doi.org/10.1016/j.jenvman.2013.08.053

Stilgoe, J. (2007). *Nanodialogues: Experiments in public engagement with science*. Demos. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Nanodialogues+Experiments+in+public+engagement+with+science#0

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

Stirling, A. (2008). "Opening up" and "closing down": Power, participation, and pluralism in the social appraisal of technology. *Science Technology and Human Values*, *33*(2), 262–294. https://doi.org/10.1177/0162243907311265

Swierstra, T., & Waelbers, K. (2012). Designing a Good Life: A Matrix for the Technological Mediation of Morality. *Science and Engineering Ethics*, *18*(1), 157–172. https://doi.org/10.1007/s11948-010-9251-1

Taibi, E., Gualberti, G., Bazilian, M., & Gielen, D. (2016). A framework for technology cooperation to accelerate the deployment of renewable energy in Pacific Island Countries. *Energy Policy*, *98*, 778–790. https://doi.org/10.1016/j.enpol.2016.03.009

Tangney, P. (2017). What use is CRELE? A response to Dunn and Laing. In *Environmental Science and Policy* (Vol. 77, pp. 147–150). https://doi.org/10.1016/j.envsci.2017.08.012

Tangney, P., & Howes, M. (2016). The politics of evidence-based policy: A comparative analysis of climate adaptation in Australia and the UK. *Environment and Planning C: Government and Policy*, *34*(6), 1115–1134. https://doi.org/10.1177/0263774X15602023

Tengö, M., Brondizio, E. S., Elmqvist, T., Malmer, P., & Spierenburg, M. (2014). Connecting diverse knowledge systems for enhanced ecosystem governance: The multiple evidence base approach. In *Ambio* (Vol. 43, Issue 5, pp. 579–591). https://doi.org/10.1007/s13280-014-0501-3

Turnhout, E., Metze, T., Wyborn, C., Klenk, N., & Louder, E. (2020). The politics of co-production: participation, power, and transformation. *Current Opinion in Environmental Sustainability*, *42*(2018), 15–21. https://doi.org/10.1016/j.cosust.2019.11.009

Turnhout, E., Stuiver, M., Judith, J., Harms, B., Leeuwis, C., Klostermann, J., Harms, B., & Leeuwis, C. (2013). New roles of science in society: Different repertoires of knowledge brokering. *Science and Public Policy*, *40*(3), 354–365. https://doi.org/10.1093/scipol/scs114

Vaidya, A., & Mayer, A. L. (2014). Use of the participatory approach to develop sustainability assessments for natural resource management. *International Journal of Sustainable Development & World Ecology*, *21*(4), 369–379. https://doi.org/10.1080/13504509.2013.868376

Van den Berg, C. F. (2017). Dynamics in the Dutch policy advisory system: externalization, politicization and the legacy of pillarization. *Policy Sciences*, *50*(1), 63–84. https://doi.org/10.1007/s11077-016-9257-x

Van den Hove, S. (2007). A rationale for science-policy interfaces. *Futures*, *39*(7), 807–826. https://doi.org/10.1016/j.futures.2006.12.004

Van der Hel, S. (2020). *New Science Institutions for Global Sustainability*. Universiteit Utrecht.

Van der Hel, Sandra. (2016). New science for global sustainability? The institutionalisation of knowledge co-production in Future Earth. *Environmental Science and Policy*, *61*, 165–175. https://doi.org/10.1016/j.envsci.2016.03.012

Van der Hel, Sandra, & Biermann, F. (2016). The authority of science in sustainability governance: A structured comparison of six science institutions engaged with the Sustainable Development Goals. *Environmental Science and Policy*, *January*, 1–10. https://doi.org/10.1016/j.envsci.2017.03.008

Van der Meer, F.-B., & Edelenbos, J. (2006). Evaluation in Multi-Actor Policy Processes: Accountability, Learning and Co-operation. *Evaluation*, *12*(2), 201–

218. https://doi.org/10.1177/1356389006066972

Van der Meij, M. (2017). *Playful reflection: designing playful reflection on research and innovation*. VU University Amsterdam.

Van der Steen, M., Faber, A., Frankowski, A., & Norbruis, F. (2018). *Opgavegericht evalueren: Beleidsevaluatie voor systeemverandering*. NSOB, Den Haag

Van der Steen, M., Scherpenisse, J., Hajer, M., Van Gerwen, O. J., & Kruitwagen, S. (2015). *Learning by Doing: Government participation in an energetic society*. NSOB, Den Haag

Van Epp, M., & Garside, B. (2019). Towards an evidence base on the value of social learning-oriented approaches in the context of climate change and food security. *Environmental Policy and Governance*, *29*(2), 118–131. https://doi.org/10.1002/eet.1835

Van Hemelrijck, A., & Guijt, I. (2016). Balancing Inclusiveness, Rigour and Feasibility: Insights from Participatory Impact Evaluations in Ghana and Vietnam. *CDI Practice Paper*, *14*. www.ids.ac.uk/cdi

Van Hove, L., & Wickson, F. (2017). Responsible Research Is Not Good Science: Divergences Inhibiting the Enactment of RRI in Nanosafety. *NanoEthics*, *11*(3), 213–228. https://doi.org/10.1007/s11569-017-0306-5

Van Lente, H., Swierstra, T., & Joly, P. B. (2017). Responsible innovation as a critique of technology assessment. *Journal of Responsible Innovation*, *4*(2), 254–261. https://doi.org/10.1080/23299460.2017.1326261

Van Mierlo, B., Regeer, B. J., Van Amstel, M., Arkesteijn, M., Beekman, V., Bunders, J., de Cock Buning, T., Elzen, B., Hoes, A.-C., & Leeuwis, C. (2010). *Reflexive monitoring in action: A guide for monitoring system innovation projects*.

Van Twist, M., Kort, M., & Van der Steen, M. (2015). Assessing and Appraising the Effects of Policy for Wicked Issues: Including Unforeseen Achievements in the Evaluation of the District Policy for Deprived Areas in The Netherlands. *International Journal of Public Administration*, *38*(8), 596–605. https://doi.org/10.1080/01900692.2014.952821

Van Veen, S. C., Verwoerd, L., & Regeer, B. J. (2016). *Characteristics of reflexive evaluation - a literature review conducted in the context of the Natuurpact (2014-2027) evaluation*.

Verwoerd, L., De Wildt-Liesveld, R., & Regeer, B. J. (2017). *The value of reflexive evaluation - a review of the Natuurpact evaluation (2014-2017)*.

Verwoerd, L., Klaassen, P., & Regeer, B. J. (2019). *The value of a reflexive evaluation approach in the eyes of researchers*.

Verwoerd, L., Klaassen, P., & Regeer, B. J. (2021). How to normalize reflexive evaluation? Navigating between legitimacy and integrity. *Evaluation*, *27*(2), 229–

250. https://doi.org/10.1177/1356389020969721

Verwoerd, L., Klaassen, P., Van Veen, S. C., De Wildt-Liesveld, R., & Regeer, B. J. (2020). Combining the roles of evaluator and facilitator: Assessing societal impacts of transdisciplinary research while building capacities to improve its quality. *Environmental Science and Policy*, *103*, 32–40. https://doi.org/10.1016/j.envsci.2019.10.011

Visser, J. Hemerijck, A. (1997). *"A Dutch Miracle". Job growth, welfare reform and corporatism in the Netherlands*. Amsterdam University Press.

Voss, J.-P., Bauknecht, D., & Kemp, R. (2006). *Reflexive Governance for Sustainable Development*. Edward Elgar Publishing.

VWS. (2019). *Startevaluatie Pilot Lerend evalueren VWS Meerjarenprogramma 2018-2022*.

Walter, A. I., Helgenberger, S., Wiek, A., & Scholz, R. W. (2007). Measuring societal effects of transdisciplinary research projects: design and application of an evaluation method. *Evaluation and Program Planning*, *30*(4), 325–338. https://doi.org/10.1016/j.evalprogplan.2007.08.002

Warde, A. (2005). Consumption and Theories of Practice. *Journal of Consumer Culture*, *5*(2), 131–153. https://doi.org/10.1177/1469540505053090

Weiss, C. H. (1981). Measuring the Use of Evaluation. In J. A. Ciarlo (Ed.), *Utilizing Evaluation: Concepts and Measurements Techniques*. Thousand Oaks, CA, US: Sage Publications.

Weiss, C. H. (1995). The haphazard connection: Social science and public policy. In *International Journal of Educational Research* (Vol. 23, Issue 2, pp. 137–150). Pergamon. https://doi.org/10.1016/0883-0355(95)91498-6

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, *7*(2), 225–246. https://doi.org/0803973233

Wenger, Etienne. (1998). Communities of practice: Learning as a social system. In *Systems thinker* (Vol. 9, Issue 5, pp. 2–3).

West, S., Van Kerkhoff, L., & Wagenaar, H. (2019). Beyond "linking knowledge and action": towards a practice-based approach to transdisciplinary sustainability interventions. *Policy Studies*, *40*(5), 534–555. https://doi.org/10.1080/01442872.2019.1618810

Wickson, F., Carew, A. L., & Russell, A. W. (2006). Transdisciplinary research: characteristics, quandaries and quality. *Futures*, *38*(9), 1046–1059. https://doi.org/10.1016/j.futures.2006.02.011

Wickson, Fern, & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: learning from transdisciplinarity. *Journal of Responsible Innovation*, *1*(3), 254–273. https://doi.org/10.1080/23299460.2014.963004

Wiek, A., Talwar, S., O'Shea, M., & Robinson, J. (2014). Toward a methodological scheme for capturing societal effects of participatory sustainability research. *Research Evaluation*, *23*(2), 117–132. https://doi.org/10.1093/reseval/rvt031

Williams, P. (2002). The Competent Boundary Spanner. *Public Administration*, *80*(1), 103–124. https://doi.org/10.1111/1467-9299.00296

Wilsdon, J., Wynne, B., & Stilgoe, J. (2005). The Public Value of Science - Or how to ensure that science really matters. In … *Science. Or how to ensure that science really matters,* …. Demos. https://doi.org/10.13140/RG.2.1.2281.7449

Wise, R. M., Fazey, I., Stafford Smith, M., Park, S. E., Eakin, H. C., Archer Van Garderen, E. R. M., & Campbell, B. (2014). Reconceptualising adaptation to climate change as part of pathways of change and response. *Global Environmental Change*, *28*, 325–336. https://doi.org/10.1016/j.gloenvcha.2013.12.002

Wittmayer, J. M., & Schäpke, N. (2014). Action, research and participation: roles of researchers in sustainability transitions. *Sustainability Science*, *9*, 483–496.

Woltersdorf, L., Lang, P., & Döll, P. (2019). How to set up a transdisciplinary research project in Central Asia: description and evaluation. *Sustainability Science*, *14*, 697–711. https://doi.org/10.1007/s11625-018-0625-7

Wynne, B. (2007). Public Participation in Science and Technology: Performing and Obscuring a Political–Conceptual Category Mistake. *East Asian Science, Technology and Society: An International Journal*, *1*(1), 99–110. https://doi.org/10.1007/s12280-007-9004-7

Wynne, Brian. (2006). Public Engagement as a Means of Restoring Public Trust in Science – Hitting the Notes, but Missing the Music? *Public Health Genomics*, *9*(3), 211–220. https://doi.org/10.1159/000092659

Yanow, D. (2000). Underlying Assumptions of an Interpretive Approach: The Importance of Local Knowledge. In *Conducting Interpretive Policy Analysis* (pp. 1–26). SAGE Publications Inc. https://doi.org/10.4135/9781412983747.n1

Yarime, M., Trencher, G., Mino, T., Scholz, R. W., Olsson, L., Ness, B., Frantzeskaki, N., & Rotmans, J. (2012). Establishing sustainability science in higher education institutions: Towards an integration of academic development, institutionalization, and stakeholder collaborations. *Sustainability Science*, *7*(SUPPL. 1), 101–113. https://doi.org/10.1007/s11625-012-0157-5

Yin, R. K. (2003). Designing case studies. *Qualitative Research Methods*, 359–386.

Zhang, W. (2011). Power shifts in a changing world order. The role of the European Union and the position of the Netherlands. In *Power shifts in a changing world order. The role of the European Union and the position of the Netherlands*.

Zscheischler, J., Rogga, S., & Lange, A. (2018). The success of transdisciplinary research for sustainable land use: individual perceptions and assessments.

Sustainability Science, *13*, 1061–1074. https://doi.org/10.1007/s11625-018-0556-3

Zuiderent-Jerak, T. (2007). Preventing implementation: Exploring interventions with standardization in healthcare. *Science as Culture*, *16*(3), 311–329. https://doi.org/10.1080/09505430701568719

Zwart, H., Landeweerd, L., & Van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the European research funding arena from 'ELSA' to 'RRI.' *Life Sciences, Society and Policy*, *10*(1). https://doi.org/10.1186/s40504-014-0011-x

Zweekhorst, M. B. M., Broerse, J. E. W., & Bunders, J. F. G. (2002). Institutionalizing a transdisciplinary approach to technology development in a Bangladeshi NGO. *Interdisciplinary Environmental Review*, *4*(1), 43. https://doi.org/10.1504/ier.2002.053875

Summary

Introduction and background

The complex and intractable character of today's environmental and sustainability issues present unprecedented challenges to science and policy alike. In response, there is called for more reflexive modes of knowledge production. With such modes, academic and non-academic actors collaborate in processes of knowledge co-production geared towards societal transformation and sustainable development. Despite knowledge co-production's increasing popularity, it has been argued that, in practice, it appears to deviate little from conventional technocratic ideas on the interactions between science and policy. Policy researchers in the Global North who aspire knowledge co-production are shown to run into societal, political, cultural and institutional barriers as the respective science-policy systems appear to privilege more classical modes of knowledge production. Consequently, co-production's transformative power often falls short of its potential to contribute to sustainability. [

Addressing these barriers, scholars have called for the institutionalisation of reflexive research. Yet, the process by which institutionalisation might be achieved and what its outcomes might look like so far have received relatively little empirical and theoretical attention. This thesis aims to make up for this, and does so by exploring the process by which knowledge co-production in policy evaluation – as an enactment of reflexive research – becomes *normalised* at a Dutch knowledge institute positioned at the intersection of science and policy, viz. the PBL Netherlands Environmental Policy Assessment Agency (Dutch: *Planbureau voor de Leefomgeving*, or PBL).

The main research question addressed in this thesis is as follows:

> *How does the process of normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency take shape?*

With normalisation, I refer to the process by which knowledge co-production becomes implemented, embedded and integrated into an organisation's standing research practice. Normalisation does not imply that co-production has replaced or subsumed all other approaches to policy research; rather, I refer to a situation where co-production in policy evaluation has become *just as normal* as other standing approaches. I adopted this understanding of normalisation from Normalisation Process Theory (NPT)), a theoretical framework to study and advance the

normalisation of newly introduced practices, on which I expand in **Chapter 2**. The framework conceptualises the complex interactions between actors who aspire knowledge co-production and its normalisation, and the (unconducive) contextual structures – the predominant rules and norms that guide their practice – they may encounter. Notably, NPT has particular attention for the work policy researchers do to (un)successfully navigate these structures, which arguably affects the transformative potential of the knowledge co-production process they engage with.

Research design

I adopted a practice-based approach to empirically study PBL's (attempted) practice of co-production (**Chapter 3**). The PBL has an authoritative status on the Dutch science-policy interface and a long-standing positivist-technocratic orientation towards policy research. In recent years, the PBL has made its ambition for innovative and reflexive research modes to better address complex socio-environmental problems increasingly explicit. At the same time, however, its policy researchers cannot elude the more traditional expectations on the organisation's role and function that reside within the organisation itself and its wider policy context. In my studies, the organisation serves as paradigmatic case to study the work policy researchers do to implement and normalise co-production. The insights presented in this thesis holds relevance for policy researchers (and their organisations) who find themselves in similar positions and with similar aspirations.

Over the course of my research (2015-2021), I participated as researcher-practitioner in various evaluation projects in which a knowledge co-production approach was adopted. This allowed me a 'view from the trenches' on the emerging practice of co-production. I conducted a total of four studies for which I made use of various methods for data collection – such as participant observation, semi-structured interviews, focus group discussions and document analysis – to address different aspects of the normalisation process. These include the challenges policy researchers face, the activities they undertake to address these, the influence of contextual developments on normalisation and the potential of reflexive monitoring to support normalisation.

The main empirical focus of this thesis is the Natuurpact reflexive evaluation (Dutch: *Lerende evaluatie van het Natuurpact*; NP) a longitudinal large-scale policy research programme conducted by the PBL in cooperation with Wageningen University & Research (WUR). The NP research project runs from 2014-2027 and involves multiple three-year cycles. In the project an explicit knowledge co-production approach is adopted. While hardly the organisation's first undertaking of knowledge co-production, the NP project was its first endeavour of knowledge co-production in

policy evaluation of considerable scale and duration. I was involved as reflexive monitor, for which I combined the roles of facilitator of reflexive learning and impact assessor, to both help develop capacity for co-production and to review the project's policy impact. Additionally, two other policy evaluations at the PBL with a knowledge co-production approach served as cases in my research: the Inter-Administrative Programme Vibrant Rural Areas reflexive evaluation (Dutch: *Lerende evaluatie van het Interbestuurlijk Programma Vitaal Platteland*; VRA) and the Regional Deals for advancing regional wellbeing reflexive research programme (Dutch: *Lerend kennisprogramma voor Regio Deals voor regionale brede welvaart*; RD), which provided the opportunity for more comparative analyses of policy researchers' doings and sayings with normalising co-production.

Main conclusions and discussion

As regards how the normalisation process took shape, I found that the NP research team's decisions on how to operationalise key features of knowledge co-production were informed by disparate institutional logics – the modernist and reflexive – on what policy evaluation is and 'ought to do' (**Chapter 4**). The collaborative character of co-production that characterised the NP project evoked criticism from the PBL community for its perceived risk for objectivity and independence, and the organisation's credibility in general. These golden standards for rigorous and valid policy research trace back to the predominant discourse on policy science, rooted in modernist logic. In a similar vein, the collaborative character made (some) policy actors apprehensive to take on the active role and ownership required for successful collaboration, as it rubbed up against entrenched expectations on the roles of policy researchers and actors. The modernist logic prevailing, knowledge co-production lacked legitimacy as valid approach to policy evaluation in the eyes of PBL colleagues and (some of the) involved policy actors. This was also found for the VRA and RD projects (**Chapter 5**). The lack of legitimacy for co-production in both the organisational and policy context presents the first part of policy researchers' main challenge for its implementation and normalisation.

The research project teams were observed to invest significant time and effort to legitimise co-production to ensure the organisational support and involvement of policy actors required for their projects' success (**Chapters 4, 5** and parts of **8**). The actions the teams undertook were directed at establishing alignment between prevailing structures for policy research (such as rules and norms; rooted in the modernist logic), and co-production's key features (embodying reflexive logic). In **Chapters 4** and **5** I identify two types of alignment activities. First, the negotiation of prevailing structures to stretch and restructure these to also encompass key features

of co-production. For example, it became increasingly regular practice to extend peer-review communities to policy and societal actors, by which the organisation's golden standard of objectivity was implicitly expanded to intersubjectivity. Second, when rules appeared too rigid to negotiate, policy researchers were observed to modify co-production's theoretical-ideal key features to ensure alignment. While a certain degree of modification is likely necessary for an innovation to have viability when introduced into a new context, scholars also warn against its over-modification as this may cause the loss of the innovation's purpose integrity. It is here that the second part of the main challenge that policy researchers faced for normalisation comes to the fore: developing legitimacy for knowledge co-production, without compromising the integrity of its original purpose by over-modifying its theoretical-ideal features.

In my research, I posit that such over-modification may at least partially explain why some co-production practices are found to be tokenistic and deviate little from their technocratic counterparts. This appeared especially the case when policy researchers were less familiar with co-production's key features. However, my research also demonstrated that, as experience grew and the consequences of over-modification became materially manifest, integrity loss was remedied by the re-modification of certain features. **Chapter 5** shows how, for instance, the project teams over the course of their projects strived to increase the diversity of involved stakeholders to enhance the wider relevance and usability of their projects' findings. Moreover, I found that the temporal less-rigorous application of (aspects of) key features may be permissible to develop the legitimacy and commitment required for co-production's initiation. This arguably requires advanced understanding of matters of integrity and aspects of co-production that are non-negotiable to prevent over-modification and unrepairable loss of integrity. Also, it points towards the importance of a long-term take on normalisation processes.

Chapter 8 additionally sheds lights on the development of specific reflexive skills by the NP project team, including sensitivity to disparate problem framings and institutional realities. These observations stress the importance of reflexive learning for the normalisation of knowledge co-production, as also highlighted by other scholars. Nonetheless, I found that the learning that took place tended to be of a somewhat superficial character. Often, learning occurred ad hoc, and only insofar as was required to resolve immediate (legitimacy) issues and continue with the projects (without, for instance, truly resolving deeper matters pertaining to interdisciplinary or epistemological confusion). It appeared that policy researchers were more occupied with the 'hardware' of co-production (e.g., organising group interviews and multi-stakeholder workshops, and other more procedural aspects) as opposed to its 'software' (i.e., reflecting upon underlying values and ideals, and how these may

conflict with prevailing rules). Scholars argue that attention for software is elemental for normalisation, because the necessary cultural change may otherwise lag behind, resulting in integrity loss. Here lies a potentially opportune role for reflexive monitoring to foster learning-by-doing and promote co-production integrity, upon which I reflect in **Chapters 7** and **8**.

In these chapters I make the case for reflexive monitoring in action to simultaneously promote and assess knowledge co-production in evaluation, thereby contributing to its normalisation. Reflexive monitoring is a methodology aimed at enhancing the reflexivity of the project team to allow them to act in greater accordance with their ambitions and ideals in light of the institutional, societal and political context. I reflect upon the dual roles that I and co-workers from the VU Athena Institute adopted during the NP project as facilitators of reflexive learning and as assessors of the project's quality and impact. While combining these roles were found to work complementary, we also experienced difficulties with compelling the team to pause and reflect on the values, presumptions and normative orientations guiding their practice. They were more occupied with our impact assessment, which, while formally purposed to inform learning processes for the next project cycle, to them was a powerful tool to legitimise co-production by attesting to its scientific rigour and impact. Their preoccupation with legitimising co-production led the NP project team to (unwittingly) outsource the responsibility for maintaining the integrity of their approach to us: their focus on its hardware and 'doing' co-production abstained attention for its software and 'thinking' about it. For reflexive monitoring to successfully contribute to normalisation, the respective project team should experience ownership over their reflexive learning process, and structural reflection should be adequately embedded within the knowledge co-production process.

In addition to providing detail to the process of normalisation, **Chapters 5** and **6** of this thesis discuss how contextual developments affect how rigidly unconducive contextual structures are maintained, thereby delimiting the space that policy researchers have to implement and normalise co-production without over-modification. I posit that such contextual developments affect what in literature is conceptualised as contextual readiness for normalisation. In **Chapter 5** I discuss several organisational developments, including the instalment of an organisation-wide Community of Practice as platform to discuss the projects' approach, which are revealed as important contributing factors for normalisation. My findings corroborate previous studies that suggest that organisational learning is crucial for successful normalisation. This type of learning pertains to the 'orgware' of co-production, meaning organising the supporting structures for the aforementioned hardware and software (e.g., sufficient capacity, training, changes in organisational procedures).

As regards the readiness of the policy context, **Chapter 6** reports on the influence of political developments as these affect how policy actors and researchers involved with the NP project perceive its impact and value its co-production approach. Respondents were observed to mobilise four ideal-typical impact rationales to make sense of the project's impact: the *accountability, instrumental, network* and *transformative* rationale. The four rationales are shown to differ in their degree of appreciation for knowledge co-production, amongst other things. Which rationale is recruited, so my research suggests, is affected by political developments, such as policy deadlines drawing near. This presented a risk for the normalisation of co-production, as policy actors were observed to increasingly mobilise the accountability rationale and favour more classical functions of policy evaluation (e.g., fulfilling purposes of accountability and compliance). Similar observations have been made in other studies, pointing out that, as political tensions grow, policy actors are inclined to default into conventional science-policy interactions as these are widely assumed to provide indisputable, independent and objective knowledge. This allows them to shift blame and avoid accountability for public decisions. My research draws attention to important questions, e.g., is it practically possible to implement and normalise knowledge co-production (with integrity) in highly political settings? At the same time, however, I observed policy actors progressively mobilise the transformative rationale, as they increasingly viewed nature policy as running into systemic barriers, especially in interaction with other policy domains. Consequently, their perceived urgency for system learning and transformative change increased. Political developments hence may also enhance contextual readiness for knowledge co-production, as actors may experience increased urgency for system change and reflexive ways of working.

Contributions to research and practice

This thesis aims to promote reflexive research's advancement in science-policy systems by deepening understanding of how a normalisation process of a practice of knowledge co-production in policy evaluation takes shape. It is a complex, dynamic and emerging process, that involves the on-going orchestration of alignment by negotiating contextual structures and modifying co-production's key features. During this process, rather than a focus on a 'perfect' practice, focusing on what steps can be taken to establish a foothold for co-production may be more fruitful. A (temporal) less-rigorous application of co-production's key features may be permissible in light of this foothold and to accommodate to unconducive structures. Nevertheless, for normalisation to be successful, the structures that make up the contexts in which co-production is introduced need to change in sync with the new practice itself. To fully appreciate and allow for these change processes, a long-term perspective is necessary.

Practically, this thesis provides useful vantage points to support this process. Normalisation involves policy researchers and research teams to navigate legitimising co-production on the one hand and staying true to its theoretical ideals on the other, in a type of unchoreographed *dance*. To know which foot is next, policy researchers require reflexive capacity. In light of this, I argue that more purposeful investment in reflexive learning and consideration for co-production's software is necessary. This asks structural reflection, and its adequate embedding in the research process from its onset. In further support for normalisation, I found that organisations may encourage contextual readiness by supplying the required orgware, thereby providing a responsive institutional environment to support innovation. Additionally, continued attention for co-production's hardware is necessary to ensure the quality and impact of the more methodological and procedural aspects the approach. As regards the readiness of the policy context, policy researchers – and their organisations for that matter – require sensitivity to political-administrative tension to deduce whether and how co-production might be a suitable way forward. While outside of the scope of this thesis, for successful science-policy knowledge co-production, normalisation should also occur within the policy context. Likely, in this context notions similar to software, hardware and orgware may be fruitful to assess and promote co-production's normalisation.

To conclude, the necessary changes towards a more reflexive research practice do not occur in dramatic revolutions, but have an evolutionary and incremental character. But together, small steps can take you a long way and may lead to significant changes. It is by taking small steps that organisations and their policy researchers may gradually develop the space and capacity to walk the walk of co-production, and to move towards a practice of reflexivity.

Acknowledgements

While the purpose of writing and defending a PhD thesis is without a doubt to demonstrate one's capability to conduct independent academic research, the process I went through the past six (seven?) years can in no way be considered a purely solitary effort. In this final section I'd like to take a moment to express my gratitude to all who have joined me on this journey.

First and foremost I am extremely indebted to **Barbara**, my promotor. Bar, your advice, leadership and friendship have been invaluable. I remember how nervous I was during our first encounters during my internship on the Natuurpact reflexive evaluation project back in 2015. I was convinced that you would soon see how out of place I was (imposter syndrome, anyone?). Your openness, enthusiasm and honest curiosity for my opinions and thoughts encouraged me to increasingly find my own voice and grow more confident, both professionally and personally. You have my trust and friendship forever (but are we 3's or are we 5's?).

Additionally, I would like to express my sincere appreciation for the other members of my PhD team. **Pim**, you have been, and continue to be, such a wonder to work with. It appears to me that you thrive in the face of complexity, a calming feat that never ceases to amaze and which has been a true foothold whenever the whole of concepts, frameworks, theories, paradigms, principles, mechanisms, criteria, rationales, modalities, empirics – you name it – threatened to get the best of me. I will never forget what you said when I told you I dreaded writing this thesis' theoretical background because I found it so dauntingly complex (*"But Lisa, you've already written this a thousand times in different ways."*). It (almost) wrote itself when I found this to indeed be true. **Eva**, I feel so lucky to have met you! Not only as part of my PhD team, but also as colleague and dear friend at the PBL you have been invaluable guiding me through the organisation and its sometimes obscure norms and rules. You have been a wonderful sparring partner, and your encouragement and attention to detail have been of great support. I am so pleased to see how our mutual interest and expertise for reflexive research and transformative learning in complex environmental policy issues continuous to grow and finds traction in our latest endeavours with the LiON programme. I have greatly enjoyed our walks and talks in Amelisweerd, and I hope there will be many more to come. **Anne**, while you are a more recent addition to the team, you have my sincere thanks for your support, lightness and positivity, and critical perspective. I am forever impressed by your brain and the amount of references you are able to effortlessly reproduce out of the top of your head – an

honest walking encyclopaedia. To all, thanks for the laughs and the trust that I would conclude this PhD journey to a successful end. I look forward to continuing our collaborations in the future, in which I have no doubt.

First, as parttime posted worker in 2017-2018, and from 2019 onwards as employee, which I combined with my work at the university (0.8 fulltime equivalent (FTE)/0.2 FTE, resp.). At the PBL, I was involved with various research projects in which a co-production approach was adopted. In addition, I was tasked with supporting the professionalising and embedding of reflexive research within the organisation. This provided me with a 'view from the trenches', and allowed me to experience the practical concerns of researchers with enacting co-production first hand.

I am also deeply indebted to the PBL organisation and its researchers for allowing me a 'view from the trenches' to witness the ongoing development of their reflexive practice. Initially as posted worker and later also as employee, I was granted ample opportunity to develop myself professionally and further my interest into reflexive research and knowledge co-production. My PBL colleagues worked to make me feel right at home. In particular, my gratitude goes out to members of the Natuurpact reflexive evaluation project team – **Rob, Irene, Wiebren, Alwin** and **Dirk-Jan**. Over the years you shared with me your insecurities, doubts, convictions, eye-openers and overall lessons learned that to a large extent provided the empirical grounds for the research in this thesis. In a similar vein, my thanks goes out to members of the project teams of the Inter-Administrative Programme Vibrant Rural Areas reflexive evaluation and the Regional Deals for advancing regional wellbeing participatory research programme – **Hiddo, Pia, Lia, Martha, Mark, Emil, Femke D., Saskia, Marieke** and, of course, **Anet** – and many other colleagues driving PBL's reflexive practice, including our Chief Scientist, **Bas**. In particular I want to mention **Femke V.** for the inspiring talks and her infectious enthusiasm. Thank you everyone for the interviews, your contributions during focus group sessions, workshops and seminars and many other more informal conversations that shaped my thinking and understanding of your research practice. You made yourself vulnerable in sharing your experiences and reflections, and I am humbled by your trust.

During my journey there have been quite a few other colleagues from the VU University Athena Institute who joined me parts of the way. To start, **Hilde**, you are such an asset to our Athena-team. Your critical view, dare to voice alternative explanations and questions, and (at times hilarious) sobering remarks have been a joy, as have been our more personal talks about work and life in general. **Evelien**, your sharp mind, eye for detail and seemingly inexhaustible drive for conceptualising,

theorising and analysing ('getting into the nitty gritty' – the more complex the better?) are a true inspiration.

Of course these acknowledgements would not be complete without noting my dear co-occupants of room U520. **Nicole, Anouk** and **Wieke**, I have enjoyed our time together immensely. Thank you for the many many many coffees, Tony Chocolonely's and chats. Additionally, I'd like to thank **Jantien** and **Aafke** for the laughs and dances and **Durwin** for the much appreciated motivational talks (and the occasional sneaky cigarette when we were still young and foolish). To each of you: as the saying goes, indeed sharing our PhD plights reduced their weight by half. Our time at Athena has come to a rather abrupt end due to the COVID-19 pandemic, some of us finishing their thesis and finding new challenges elsewhere during a (partial) lock-down. I hope there will be an opportunity in the near future to celebrate our victories in a more Athena-like manner.

I'd like to thank my close friends **Judith** and **Cecile** for their everlasting belief and listening ears for whenever I felt overwhelmed and needed an outlet (be it in the shape of bottles of wine or oat milk cappuccinos). **Tom** and **Isabelle**, it has been so great to share our PhD experiences, despite my branching off into a rather different scientific discipline (marine biology seems a lifetime ago) and your travels to Oceania. Also, **Fleur, Ruth, Jen, Nadja** and **Imke**, thank you for the encouragement and cheering me on.

I have the privilege of a warm, supporting family. Natuurlijk de **van 't Landjes**, maar vooral **Mama, papa,** bonusmama **Len** en zus **Elke**: ik weet dat dit hele promoveer-gebeuren en wat ik nou allemaal doe in Den Haag niet altijd goed te volgen is – sorry dat ik zo slecht ben in het uitleggen ervan in normale-mensen taal – maar ik had het zonder jullie steun en liefde niet gekund. You encouraged me to chase my dreams and ambitions and believed in me always.

Above all I would like to thank **Mike**, my rock, my best friend, my love. The past year has been such a rollercoaster for us. We bought a house and moved to Amersfoort and became the proud parents of Jona, a wonderful feat that also presented a sudden and very strict impeding deadline for this PhD, all the while every other aspect of our lives was affected by the ongoing pandemic. There is no one else in this world I would so happily survive a lock-down with. You have worked so hard to unburden me wherever and whenever possible and this PhD would not have been possible without you. Thanks for keeping me sane (even though you drive me nuts sometimes) and keeping me going. And finally, of course, **Jona**. You are my sun. I can't wait to continue our new lives as a family of our own.

Action created the space for learning-by-doing – an iterative cycle based on the audacity to fail (at the beginning).

– Lang et al., 2012

The complex and intractable character of today's environmental and sustainability issues present unprecedented challenges to science and policy alike. In response, there are calls for more reflexive modes of knowledge production. With such modes, academic and non-academic actors may collaborate in processes of knowledge co-production geared towards societal transformation and sustainable development. However, despite its increasing popularity, knowledge co-production in practice is observed to deviate little from conventional technocratic ideas on science-policy interactions, compromising its transformative potential. This thesis sheds light on what happens when theoretical ideals for knowledge co-production meet real-life settings. It provides detail on how the normalisation process of knowledge co-production in policy evaluation takes shape by centralising the work policy researchers do for its successful implementation. By pairing theoretical with practical insights, this thesis seeks to contribute to the inclusion of practices of reflexivity in science-policy systems.