DOCTORAL THESIS

# Governing in Turbulence: Leveraging Experimentation and Data Capabilities for Building Adaptive Governance

Kerli Onno

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2026

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
28/2026

# Governing in Turbulence: Leveraging Experimentation and Data Capabilities for Building Adaptive Governance

KERLI ONNO

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Business and Governance
Ragnar Nurkse Department of Innovation and Governance
This dissertation was accepted for the defence of the degree 09/04/2026

**Supervisor**: Professor Ringa Raudla
School of Business and Governance
Tallinn University of Technology
Tallinn, Estonia

**Co-supervisor**: Professor Veiko Lember
School of Business and Governance
Tallinn University of Technology
Tallinn, Estonia

**Opponents**: Professor Sreeja Nair
Lee Kuan Yew School of Public Policy
National University of Singapore
Singapore

Professor Valerie Pattyn
Institute of Public Administration
Leiden University
Leiden, The Netherlands

**Defence of the thesis**: 01/06/2026, Tallinn

**Declaration:**

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Kerli Onno

---
signature

European Union
European Regional
Development Fund

Investing
in your future


ISSN 2585-6898 (publication)
ISBN 978-9916-80-488-9 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-489-6 (PDF)
DOI https://doi.org/10.23658/taltech.28/2026

Onno, K. (2026). *Governing in Turbulence: Leveraging Experimentation and Data Capabilities for Building Adaptive Governance* [TalTech Press]. https://doi.org/10.23658/taltech.28/2026

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
28/2026

# Valitsemine muutlikes oludes: eksperimenteerimise ja andmevõimekuste kasutamine kohaneva valitsemise toetamiseks

KERLI ONNO

TAL TECH
KIRJASTUS

# Contents

## List of Publications

This PhD thesis relies on five original studies (published, accepted for publication or under review), each referred to throughout the thesis with a respective Roman numeral:

I Raudla, R., Sarapuu, K., Juuse, E., Harbuzova, N., Onno, K., Vallistu, J., & Cepilovs, A. (2023). To experiment or not to experiment in tax policy? Halduskultuur–The Estonian Journal of Administrative Culture and Digital Governance, 23(1), 27–48.

II Onno, K. (2025). The Role of Data Capabilities in Developing Public Sector Dynamic Capabilities: Lessons from the COVID-19 Pandemic. International Journal of Public Administration, 48(11), 736–747.

III Raudla, R., Sarapuu, K., Vallistu, J., Onno, K., & Harbuzova, N. (2025). The politics of experimental policymaking: the influence of blame avoidance and credit claiming. Policy Sciences, 58(2), 245–266.

### Other Publications

IV Onno, K., Raudla, R., Sarapuu, K., Juuse, E., & Vallistu, J. Public sector experimentation in times of crises: exploring the paradoxes (available on request)

V Raudla, R., Sarapuu, K., Vallistu, J., & Onno, K. Learning from policy experiments (available on request)

## Author's Contribution to the Publications

Contribution to the papers in this thesis are:

I A co-authored article. The author of the doctoral thesis was not the lead author but was responsible for specific parts of the paper. For example, the author contributed to the literature review, data collection, and writing the results.

II A single-authored article. The author of the doctoral thesis was responsible for designing and conducting all the stages of the study, including developing the framework of the study, collecting and analysing data, and writing the results.

III A co-authored article. The author of the doctoral thesis was not the lead author but was involved in data collection and provided input to specific parts of the article.

IV A co-authored article. The author of the doctoral thesis was the lead author of the article and was responsible for designing the study and writing most of the content of the article.

V A co-authored article. The author of the doctoral thesis was not the lead author but was involved in collecting the data and writing the paper.

## Introduction

Governments around the world operate in turbulent conditions, making unpredictability and uncertainty, but also time shortages and conflicting values (Deverell, 2010, p. 34; Ansell & Trondal, 2018) a part of their everyday working environment. This requires public sector organisations to operate in a world characterised by consistently high degrees of uncertainty and interconnectedness (Hatfield-Dodds et al., 2007; Janssen & Van der Voort, 2020; Koontz et al., 2015; Scarlett, 2013; Hurlbert, 2018), imperfect and incomplete knowledge (Hasselman, 2017; Pattyn et al., 2021), and where "unpredictable change has been the norm rather than the exception ( icl u et al., 2020, p. 168). During periods of constant and continuous turbulence that may be regarded as a "normal state of affairs" (Ansell & Trondal, 2018, p. 6), public sector organisations may also face sudden and short-term crises. This thesis views crisis as "a serious threat to the basic structures or the fundamental values and norms of a social system, which—under time pressure and uncertain circumstances—necessitates making critical decisions" (Rosenthal et al., 1989, p. 10). Crises may differ in their scope, consequences, or affected stakeholders, but it is often agreed that they lead to a "sudden (negatively perceived) change of conditions" and include "components of threat, urgency and uncertainty" (Hooren, et al. 2014, p. 2), which makes them challenging to tackle.

Under such conditions, traditional governance models often prove too rigid. As a result, adaptive governance has emerged as a salient approach (Dietz et al., 2003; Chaffin et al., 2014; Nolte & Lindenmeier, 2024; Sharma-Wallace et al., 2018; Brunner, 2010). Adaptive governance refers to governance systems that self-organise as a result of learning and interaction and are supported by institutional arrangements that enable continuous learning, collaboration, and flexibility (Rijke et al., 2012). As Folke et al. (2005, p. 444) explain, in a system with high adaptability, "the actors have the capacity to reorganize the system within desired states in response to changing conditions and disturbance events," which indicates that adaptive governance emerges from the adaptive capacity of public sector organisations. However, despite adaptive governance being widely promoted as a means of managing complexity and uncertainty, it is often argued that it remains a normative and descriptive concept that lacks sufficient operationalisation (Zhao et al., 2024; Janssen & Van der Voort, 2020; Rijke et al., 2012).

Drawing on institutional theories, organisation theories, public administration and public policy literature, this thesis aims to fill this gap by operationalising the concept through the exploration of capabilities that underlie adaptive governance. The literature highlights a broad set of capabilities, e.g., institutional flexibility, multi-level coordination, and stakeholder collaboration but especially continuous learning (Rijke et al., 2012; Armitage et al., 2008) as a crucial enabler of adaptive governance. This thesis focuses on experimentation and data capabilities that help to operationalise learning under uncertainty but also provide justification and direction for policy change—without experimentation and data capabilities, responding to changes may remain reactive or *ad hoc*. The thesis argues that experimentation and data capabilities serve as means of adjusting organisational routines—defined as "repetitive, recognizable patterns of interdependent actions" (Feldman & Pentland, 2003, p. 95)—to respond to changes in the surrounding environment and improve the adaptive capacity of public sector organisations. However, to support adaptive governance, these capabilities should become routinised and institutionalised ways of making decisions and developing policies.

Furthermore, experimentation and data capabilities may encounter significant barriers that hinder the development of adaptive governance (Eshuis & Gerrits, 2021).

To understand the conditions and antecedents needed for adaptive governance to emerge, this study explores the factors that influence experimentation and the development of data capabilities. The thesis follows an inductive analytical strategy and, drawing on empirical studies, outlines the key factors influencing the use of experimentation and data capabilities in three dimensions of governance—polity, policy, and politics (Lange et al., 2013). It is argued that while data capabilities and experimentation serve as sources of dynamism, as they support the adaptive capacity of public sector organisations, different factors influence their potential to lead to continuous and long-term capability building and routine adaptations that are required for adaptive governance.

Additionally, the implications of crises for experimentation and data capabilities are explored to complement this discussion. While crises may provide urgency and "windows of opportunity" (Ansell & Trondal, 2018) to induce changes in organisational routines, less is known about how crisis conditions influence experimentation and the development of data capabilities. The thesis outlines four paradoxes—"contradictory yet interrelated elements that exist simultaneously" (Smith & Lewis 2011, p. 386)—that stem from crisis conditions and the logics of experiments. This indicates that crises call for flexible experiments that operate on a short timescale. Also, the results indicate that while high-level data capabilities support public organisations in responding to a crisis, the temporality of crisis conditions may hinder a wider use of the results of experiments or long-term routine changes.

# 1 Research Objectives

To understand the role of data capabilities and experimentation in enhancing adaptive governance, this thesis proposes three research questions that are explored in five research articles:

*1. How do experimentation and data capabilities help public sector organisations to adapt?*

*2. Which factors influence experimentation and data capabilities?*

*3. How do crisis conditions influence experimentation and data capabilities?*

First, the doctoral thesis aims to explore how experimentation and data capabilities help public sector organisations to adapt to changing conditions through the modification of routines (**RQ1**). To explore the ways in which data capabilities and experimentation provide public sector organisations with the means to adapt, Chapter 3 adopts the dynamic capability approach (Teece et al., 1997), which has gained significant attention from researchers (e.g., Ambrosini et al., 2009; Eisenhardt & Martin, 2017; Winter, 2003). Both data capabilities and experimentation serve as sources of dynamism by enabling public sector organisations to sense the surrounding environment (more specifically, to identify and assess opportunities) and to seize those opportunities by mobilising resources. However, it is argued that their potential to transform organisational routines that ensure continuous renewal (**Article II**; Conboy et al., 2020; Teece, 2007) may be hindered by several factors outlined in this thesis (see RQ2).

The single-authored **Article II** is a case study that uses the framework of dynamic capabilities to understand how the data capabilities of the Estonian Tax and Customs Board (ETCB) enabled the change of organisational routines in the face of the COVID-19 pandemic. The article explores three organisational routines that were highly dependent on data capabilities: (1) allocating financial support to vulnerable groups in society; (2) evaluating the economic impact of the pandemic; and (3) providing management information to the ETCB board. The study focuses on crisis-induced changes in these three routines: it traces actions and patterns of action (Feldman, 2016, p. 29), explores the role of data capabilities in routine adaptation, and therefore enhances our understanding of the ways in which data capabilities support developing adaptive capacity in public sector organisations.

While **Article II** focused on the implications of data capabilities, **Articles I**, **III**, **IV**, and **V** focus on experimentation as a mechanism to gain new knowledge and learn. Experiments are considered to be a valuable approach in policy learning: they help public sector organisations test new ideas and gain information about their effectiveness and feasibility, identify the best alternatives of action, and scale up ideas that show promising results in the testing phase (**Article I**). In this thesis, the way experimentation may support the adaptive capacity of public sector organisations through routine changes or through the creation of new routines is explored. By examining the promises and pitfalls of experimentation in public policy, **Article I** seeks to understand the promises and challenges of experimentation, more specifically in tax policy.

While RQ1 considered experimentation and data capabilities as potential sources of dynamism, the second research question aims to understand the factors that influence experimentation and data capabilities (**RQ2**). As Hasselman (2017) explains, adaptive governance requires flexible and responsive legal and administrative arrangements; institutions that are able to learn and solve problems but also relate to various

stakeholders; and politics that value transparency, legitimacy, and accountability. To provide insights into the second research question, this thesis brings together the inductively derived factors that influence experimentation and the execution of data capabilities in public sector organisations. To provide analytical clarity, these factors are grouped into three dimensions: polity (organisational settings), policy, and politics (Lange et al., 2013).

First, while focusing on the organisational and managerial factors, this thesis suggests that path dependency and agency (ownership) significantly influence the organisational capacity to modify its existing routines to adapt to the surroundings. These arguments are illustrated by the case study of the ETCB (**Article II**), and the comparative study of how Estonian and Finnish civil servants perceive experimentation (**Article V**). **Article II** explores the ways public sector organisations use their data capabilities to support the transformation of routines. It is argued that while path dependency is one of the most significant factors that can support or hinder the execution of these capabilities in public sector organisations, the role of agency is important in overcoming the status quo and ensuring continuous change in routines. Furthermore, **Article V** argues that cumulative experience with experimentation influences civil servants' perceptions of whether to launch large-scale policy experiments. Moreover, agency and ownership are outlined as some of the most important aspects that underlie learning from experiments.

Second, experimentation helps identify the best policy alternatives and is therefore related to the concept of adaptive policymaking, which refers to policies that adapt over time as conditions change and learning takes place (Swanson & Bhadwal, 2009). However, different types of experiments allow for different ways of learning (**Articles I**, **IV**, **V**). Chapter 4.2 outlines the promises and challenges of different types of experimentation in policy learning. The results based on the comparative study of Finland and Estonia indicate that while experimentation can be seen as a fruitful mechanism for policy learning, the considerations of the feasibility and usability of different types of experiments may vary considerably based on the contextual factors in the two countries. While Finnish civil servants regarded RCTs as the most reliable and convincing learning mechanism, Estonian interviewees preferred design experiments owing to the diversity of information they provide.

Third, politics often defines which decisions are made within public organisations, and which capabilities are built and exercised. This, in turn, shapes how policies are designed and implemented. In Chapter 4.3, the interaction of the political setting and the logic of experimentation is discussed by focusing on the political factors that influence learning from experiments (**Article V**) and the motivations behind political support for experimentation (**Article III**). More specifically, blame avoidance and credit claiming are discussed as factors that influence political support for experimentation, and empirical results illustrate the different perceptions of Finnish and Estonian civil servants.

The third research question seeks to understand the implications of crises for experimentation and data capabilities (**RQ3**). In Chapter 5, this thesis seeks to understand how public sector organisations executed their data capabilities during the COVID-19 pandemic and the role of these capabilities in the adaptation of organisational routines. Additionally, the paradoxes that emerge in the interplay of experimentation logics and crisis conditions are explored, outlining the promises and challenges of different types of experiments in crisis conditions (**Article IV**). It is argued that the temporality of crisis conditions hinders long-term routine adaptations, as the urgency and necessity for such adaptations remain temporary as well. Also, crisis conditions and experimentation logics

create paradoxes that challenge the execution and learning potential of experiments. However, these challenges may be tackled by choosing generative experiments as an intervention method, since they provide flexibility in rapidly changing conditions.

To illustrate the focus of the current thesis, Figure 1 presents a framework that gives an overview of the study's analytical approach. It also shows how the proposed research questions provide a comprehensive basis for the thesis and thereby help to achieve the research objectives.

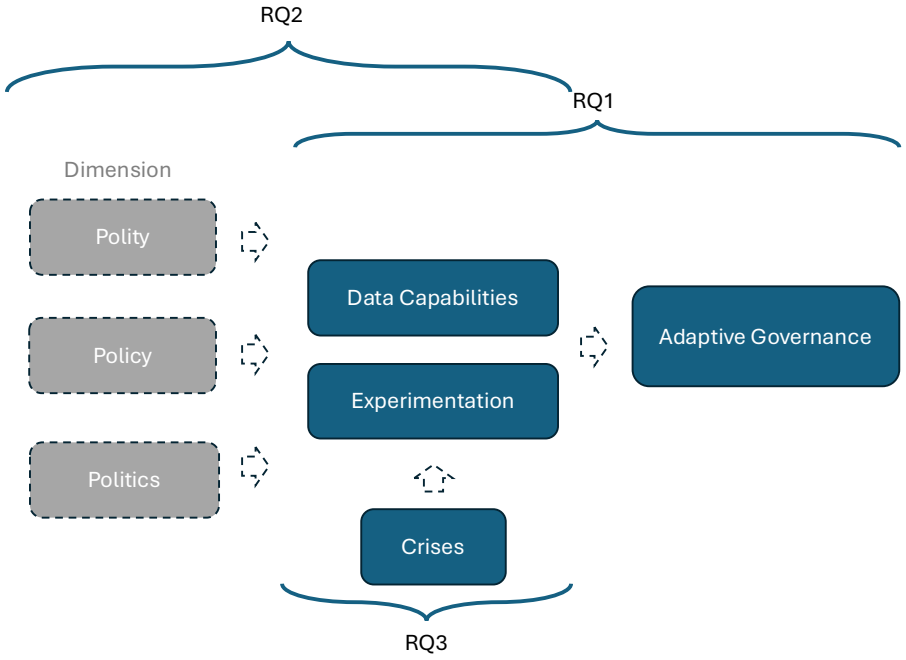


*Figure 1. The framework proposed by the author of the thesis.*

The elements and relationships that this framework presents are combined with empirical evidence from the interviews conducted with civil servants and policymakers in Estonia and Finland.

Table 1 summarises the contribution of all five articles to the proposed research questions.

*Table 1. Publications and their contribution to the research questions.*

| **Study** | *How do experimentation and data capabilities help public sector organisations to adapt?* | *Which factors influence experimentation and data capabilities?* | *How do crisis conditions influence experimentation and data capabilities?* |
|---|---|---|---|
| **Article I** | The article focuses on experimentation as a learning mechanism that enables dealing with complexity and uncertainty. | The article outlines the general promises and pitfalls of experimentation in tax policy and discusses the promises and challenges of different types of experiments. | - |
| **Article II** | The article focuses on the case of the ETCB's data capabilities to understand how they support improvements in existing (inter-organisational) routines and contribute to dynamic capability building. | The article outlines the role of path dependency and agency in developing dynamic capabilities. | Results indicate that while data capabilities provided organisations with the means to modify existing routines, the changes remained temporary as the crisis conditions abated. However, the existing level of data capabilities defined the fluency of crisis response. |
| **Article III** | The article explores the role of politicians and political decision-making in experimentation. In other words, it explores the way political motivations may influence experimentation. | The article aims to understand how credit claiming and blame avoidance influence the decision to launch experiments. | - |

| | | | |
|---|---|---|---|
| **Article IV** | The article seeks to understand the implications of crises for experimentation. | - | The article proposes four paradoxes that stem from the interplay between crisis conditions and the logics of experiments (conflicting timeframes, lack of resources, intolerance for failure, and the limitations of scalability). Crisis conditions require fast and flexible decision-making, which indicates that certain types of experiments (generative experiments) may be more suitable in crisis conditions than others (RCTs). |
| **Article V** | The article aims to understand how experiments enable policy learning. | The article focuses on three lenses: research, political, and managerial—to understand the factors that influence learning from experimentation. | - |

The thesis is structured around these three research questions. Chapter 2 gives a concise overview of the methodology of the thesis. Chapter 3 discusses the underlying logic of data capabilities and experimentation by drawing on the dynamic capabilities' framework, which helps explain how data capabilities and experimentation serve as sources of dynamism that support adaptive governance. Chapter 4 explores the factors that influence data capabilities and experimentation based on the interviews with Estonian and Finnish civil servants. The factors are discussed across three higher-order dimensions, beginning with those related to polity (Subsection 4.1), followed by policy (Subsection 4.2), and the political dimension (Subsection 4.3). Chapter 5 discusses the implications of crises for experimentation and data capabilities. Concluding discussion and limitations of the current research follow.

# 2 Methodology

To reach the aims of the study, this thesis draws on five articles. Of these five articles, three are comparative (**Articles III**, **IV**, **V**), one is a theoretical analysis of existing literature on experimental policymaking (**Article I**), and one is a single case study (**Article II**). The case study method was chosen because it supports an in-depth exploration of the chosen phenomenon and enables examining it in a broader context (Baxter & Jack, 2008; Yin, 2009). All articles were written as part of the research project "Experimental approaches and institutional innovations in the domains of fiscal and financial policy," funded by the Estonian Research Council (Grant PRG1125). **Articles V** and **III** were case studies that explored how Finnish and Estonian public officials perceived the opportunities and challenges of using experimental approaches. The case selection was based on the fact that although these countries have very similar political settings (unitary parliamentary systems with proportional electoral rules and coalition governments), their experiences with experimenting follow different paths. While Finland has extensive experience with policy experiments, emphasised by the programme "Experimental Finland" in 2015–2019 (an initiative of Prime Minister Juha Sipilä's Government), Estonia has not implemented comparable large-scale strategic programmes to support experimental approaches. In Estonia, policy experimentation is a relatively recent development, largely driven by bottom-up initiatives and supported by European Union funds to conduct "innovation sprints" and "experimentation masterclasses". (**Article V**, p. 9) Therefore, these two cases provide a fertile ground for understanding the key factors that affect learning from experiments.

**Article II** is an explanatory case study conducted to understand whether, and under what conditions, organisational routines change when faced with crisis conditions. Eriksson (2013) notes that examining dynamic capabilities requires careful sampling, as dynamic capabilities cannot be identified ex ante and it may be difficult to find relevant organisations to study. The Estonian Tax and Customs Board (ETCB) can be considered a good starting point for exploring dynamic capabilities. As Estonia is considered a successful e-state (Margetts & Naumann, 2017; Tammpuu & Masso, 2018), such high-level technological capacity supports the execution of data capabilities and the ETCB can be considered one of the most advanced public agencies in Estonia in deploying data capabilities (Lember et al., 2018). In the case study, three routines were examined, two of which were inter-organisational: (1) allocating financial support to vulnerable groups in society, (2) evaluating the potential impact of COVID-19 restrictions; (3) providing management information to the board of the ETCB.

Given the ontological and epistemological viewpoints chosen for this thesis, all five articles employ qualitative research methods. As Forman and Damschroder (2007) argue, the aim of qualitative research is to develop a deep understanding of a studied phenomenon rather than to make generalisations based on statistical data. They emphasise that qualitative approaches make it possible to understand underlying processes and present different points of view, motivations, and experiences (Forman & Damschroder, 2007). This thesis builds on various data collection methods used across the articles, including literature reviews (**Articles I–V**), interviews (**Articles II**, **III**, **IV**, **V**), and document analysis (**Articles II**, **III**, **IV**, **V**). Therefore, qualitative interview data was included in all of the articles.

**Articles III**, **IV**, and **V** drew on the interviews carried out with civil servants in Finland and Estonia between July 2022 and July 2023. The interviews were conducted in the Government Office / Prime Minister's Office, financial and fiscal policy institutions, and

other public organisations to understand the civil servants' perceptions and experiences with experimentation. Using the same interview protocol in both countries, 66 interviews were conducted in total—32 in Estonia and 34 in Finland. In **Article II**, nine interviews with civil servants from four different public organisations (the ETCB, Enterprise Estonia, the Estonian Unemployment Insurance Fund, and the Ministry of Communications and Economic Affairs) were conducted between July 2021 and March 2022. These interviews provided insights into the three organisational routines examined in the article.

*Table 2. Methodological approaches of the articles.*

| **Article** | **Aims and research questions of the study** | **Methodology** | **Data collection method** | **Focus of the analysis** |
|---|---|---|---|---|
| **Article I** | To explore the promises and challenges of experimentation and different types of experiments in tax policy | Comparative theoretical framework | Scientific literature on experimental policymaking and types of experiments | Comparative analysis of the promises and challenges of experimental approaches in the context of tax policy |
| **Article II** | To understand how organisational routines were adapted to rapidly changing conditions and to explore whether, and how, the knowledge and experience gained during the pandemic contributed to the long-term improvement of existing routines. RQs: (1) How did data capabilities enhance the adaptation of routines during the COVID-19 pandemic? (2) Do data capabilities support the development of dynamic capabilities in the public sector, and if so, under what conditions? | Explanatory single case study | Document analysis and semi-structured interviews with six civil servants from the ETCB, one from the Estonian Unemployment Insurance Fund, one from Enterprise Estonia, and one from the Ministry of Communications and Economic Affairs | Focus on organisational routines as research objects and changes in these routines that were induced by the COVID-19 pandemic |

| | | | | |
|---|---|---|---|---|
| **Article III** | The aim of the article is to advance the understanding of the politics of experimentation by analysing how policy actors' perceptions of blame avoidance and credit claiming influence experimental policymaking. RQs: (1) How have the motivations of blame avoidance and credit claiming influenced experimental policymaking in the two countries? (2) Which contextual factors have shaped blame avoidance and credit claiming motivations in public policy experiments? | Comparative case study (Finnish and Estonian public officials) | 66 semi-structured interviews conducted with public officials in Finland and Estonia (2022–2023), document analysis | Focus on the political setting of policy experimentation |
| **Article IV** | In the article, a paradox lens is adopted to understand the interplay between crisis conditions and the logics of experimentation. RQ: What paradoxes emerge in experimental policymaking during crises? | Qualitative single case study | 66 semi-structured interviews conducted with public officials in Finland and Estonia (2022–2023), document analysis | Focus on the interplay of crisis conditions and the logics of experimentation |

| | | | | |
|---|---|---|---|---|
| **Article V** | Three key lenses for examining instrumental policy learning from experiments are proposed: the research lens, the managerial lens, and the political lens. Based on these three lenses, the factors that facilitate and hinder learning from experiments are analysed | Comparative case study (Finnish and Estonian public officials) | 66 semi-structured interviews conducted with public officials in Finland and Estonia (2022–2023), document analysis | Focus on factors that influence policy learning |

The interviews explored the experiences of civil servants who were involved in providing COVID-19 pandemic response (**Article II**) and the perceptions of Estonian and Finnish public officials regarding the benefits and pitfalls of policy experiments, as well as their own experience with experimenting (**Articles III**, **IV**, **V**). The main data analysis method used in the articles was qualitative content analysis. In **Articles III**, **IV**, and **V**, the main coding software was MAXQDA, while Nvivo was used in **Article II**. The interviews were recorded and transcribed. Multiple data collection methods were used (document analysis and interviews) to reduce bias. Document analysis was also used to prepare for the interviews and understand the context and background of the topics explored.

This thesis has several methodological limitations. The interviews for **Articles III**, **IV**, and **V** were conducted in Estonia and Finland, comparing the experiences of civil servants in these two countries. However, these states may be regarded as similar in terms of the key features of their political institutions (parliamentary systems, proportional electoral systems, and coalition governments). Examining states with different governance settings could therefore further reveal factors that influence dynamic capability building. Additionally, **Article II** is a case study of the routines in one public sector organisation. Although it provides deep insight into the processes and changes brought about by the pandemic, and into the logic of the effects of dynamic capabilities, the sample size may have limited generalisability. Moreover, the insights may differ for other types of crises or institutional settings, and research could make use of other similar case studies to better understand the dynamics of routine adaptations. Furthermore, detecting the changes induced by dynamic capabilities and understanding their development would benefit from longitudinal data that could capture the dynamics of routine adaptations over a longer period.

# 3 Data Capabilities and Experimentation: Enhancing Adaptive Capacity Through Routine Modifications

Adaptive governance requires public sector organisations that are able to learn and adjust under conditions of uncertainty. High adaptive capacity ensures operational continuity in turbulent conditions but also the ability to develop informed policies despite complexity and lack of knowledge. To understand how data capabilities and experimentation support the adaptive capacity of public sector organisations, the perspective of dynamic capabilities is applied. Dynamic capabilities are routinised practices that help policymakers and public sector organisations to adapt to changing circumstances, to transform, or innovate (Teece et al., 1997; Ambrosini et al., 2009; Eisenhardt & Martin, 2017; Winter, 2003; Piening, 2013). Although both data capabilities and experimentation provide public sector organisations with means to sense their surroundings and learn, their underlying logic differs in several important aspects that are illustrated in Figure 2.

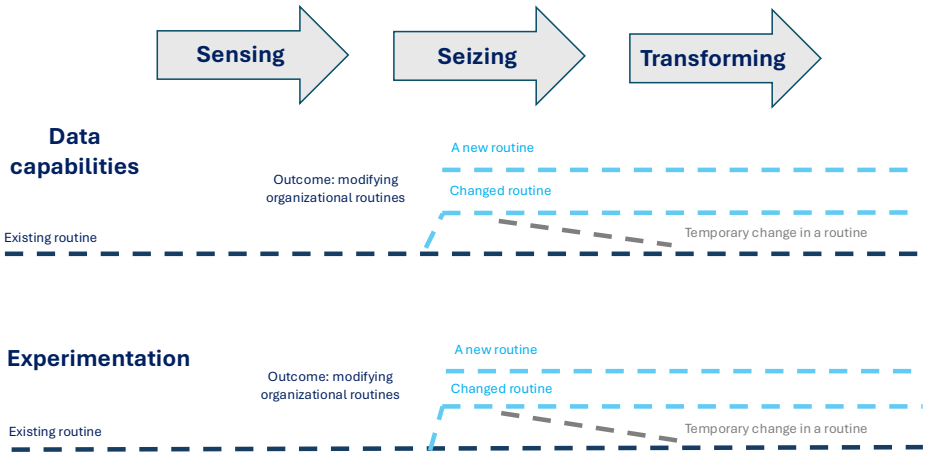


*Figure 2. The logics of experimentation and data capabilities based on the framework of dynamic capabilities. Figure proposed by the author of this thesis, based on the work of Teece et al. (1997).*

It is argued that dynamic capabilities support an organisation's ability to sense the surrounding environment (more specifically, to identify and assess opportunities), to seize those opportunities by mobilising resources, and to transform by performing activities that ensure continuous renewal (**Article II**; Conboy et al., 2020; Teece, 2007). This is achieved by modifying existing organisational routines, which are viewed as the "building blocks" of capabilities (Edmondson & Zuzul, 2016, p. 182). Table 3 outlines how the concepts of data capabilities and experimentation relate to adaptive governance, explains their learning logic, and describes their relationship to existing organisational routines.

*Table 3. From data capabilities and experimentation to adaptive governance.*

| Dynamic capability | Definition | How it relates to adaptive capacity? | Learning logic | Relationship to existing routines |
|---|---|---|---|---|
| **Data capabilities** | Ability to "collect, process, and analyse different types of information" (Mayne et al., 2020, p. 40). | Enable continuous learning and feedback, which are central to adaptive capacity. Data capabilities support acting upon the changes in the environment and transforming existing routines for adaptation. | Data capabilities are routinised practices that provide information about, and enable the monitoring of, changes in surrounding conditions. | Support the modification of existing organisational routines by providing the information and knowledge generated from data. |
| **Experimentation** | Policy-relevant tests "undertaken by government organization(s) to learn about the impacts of a new policy solution, which can be used as evidence for further policy decisions" (**Article I**) | Generates iterative learning that enables the modification of potential policy solutions based on feedback, thereby increasing adaptive capacity. | Experimentation may be used as an *ad hoc* learning strategy. Iterative learning helps to understand the implications of the intervention. | Existing routines may be influenced and altered, or new routines created after the scaling of the intervention. |

Data capabilities refer to the combination of "roles and skills, technologies, and data management practices and processes" (Farmer et al., 2022, p. 64) that enable organisations to extract value from data and use it for decision-making. In **Article II**, the logic of data capabilities, defined as the ability to "collect, process, and analyse different types of information" (Mayne et al., 2020, p. 40), is explained through the dynamic capabilities' framework:

- It is argued that organisations need to anticipate changes and pressures in their environment by sensing novel information (van Assche et al., 2022). *Sensing* capabilities support an institution's ability to detect changes in its environment and thereby help it to adapt proactively (Duchek, 2014). Such data processes relate to understanding the conditions of target groups, revealing the changing needs of service users, and interpreting other changes and uncertainties (**Article II**; Conboy et al., 2020).

- *Seizing* capabilities support identifying and implementing a solution (Duchek, 2014) by reconfiguring resources to respond to unexpected events. The underlying data processes relate to building competencies, supporting decision-making, or mobilising resources to address opportunities (**Article II**; Conboy et al., 2020), but also providing *ad hoc* solutions to emerging challenges.
- *Transforming* capabilities refer to continuous changes in operational routines. Transformation can be achieved by disrupting existing routines, e.g., through the introduction of a new technology or a work practice (**Article II**; Piening, 2013), and the underlying processes are related to reconfiguring capabilities or re-engineering processes (Conboy et al., 2020).

Besides executing routine data capabilities, experimentation can be regarded as a complementary route for enhancing adaptive capacity and as a mechanism to respond to pressing policy problems. Experiments may be undertaken to initiate *ad hoc* organisational learning and are defined as policy-relevant tests "undertaken by government organization(s) to learn about the impacts of a new policy solution, which can be used as evidence for further policy decisions" (**Article I**). However, experimentation relies on certain capabilities to succeed (e.g., collaborative or data-analytic capabilities). Therefore, in this thesis, experimentation capability refers to the ability to design, implement, and evaluate interventions (e.g., RCTs, pilots, design experiments) that could be used for learning. Similarly to data capabilities, and to understand the potential of experimentation to generate routine changes, the logics of experimentation are explained through the concepts of sensing, seizing, and transforming:

- First, experiments provide policymakers and civil servants with information on "what works", thereby enhancing their *sensing* ability by helping to validate or disprove the "hunches" underlying proposed ideas. However, experiments are usually launched when there is already sufficient information about potential solution(s) that need to be tested to validate the initial idea or to understand the impacts of the intervention (Bravo-Biosca, 2020). This distinguishes experimentation from data capabilities, which support organisations in sensing changes in their environment that require a response.
- Second, the act of launching an experiment can be regarded as *seizing*, as it aims to evaluate alternative courses of action to identify the one with the highest value (Haynes et al., 2012). However, changes introduced through experiments are always reversible and temporary, meaning that it is possible to return to the previous state (McFadgen & Huitema, 2017), if necessary. Also, experiments have "a clearly structured learning strategy, defined ex ante rather than as an afterthought" (Bravo-Biosca, 2020, p. 195), while routine adaptations initiated by data capabilities may not require such a fixed approach (**Article II**).
- Third, when successful results emerge, experiments are scaled and they can lead to *transforming* existing practices, or even the creation of new routines to adapt to changing circumstances. However, as the main goal of experimentation is learning, experiments could be regarded as successful even if they do not lead to scaling or a transformation phase, while they help avoid policy failures (Ansell & Bartenberger, 2016).

Next, this thesis explores the factors that hinder or support the development of experimentation and data capabilities in the public sector. These factors enable us to understand the conditions that support the development of adaptive governance.

# 4 From Adaptive Capacity to Adaptive Governance: Factors Influencing Experimentation and Data Capabilities

Although differing in their underlying logic, data capabilities and experimentation provide public sector organisations with the means to learn and adjust to changing conditions, serving as sources of dynamism that support adaptive capacity. Wilhelm et al. (2022) argue that the three factors (sensing, seizing, transforming) that are included in the concept of dynamic capabilities may differ in their routineness in organisations: the frequency of their execution and structuring (existence of rules, formal procedures, or standards). Therefore, while organisations may frequently apply routine dynamic capabilities, they may also perform "experiential activities such as prototyping, real-time information, multiple options, and experimenting that generate immediate knowledge quickly" (Eisenhardt & Martin, 2017, p. 1117), which indicates that the availability of experimentation and data capabilities, even when executed in an *ad hoc* manner, provides organisations with the means to advance their adaptive capacity.

However, high adaptive capacity within public sector organisations may not automatically lead to adaptive governance. For example, necessary data may be available, but the capabilities for using it in decision-making may be lacking, or evidence from experimentation may be generated, but it may be ignored for political or other reasons. Therefore, developing adaptive governance requires such capabilities to become institutionalised and a routine part of decision-making. It has been argued that institutionalisation refers to the "development of more or less regular and fixed patterns" (Eshuis & Gerrits, 2021, p. 278) that can become embedded in regular policy processes and shape the way governance operates.

Next, the thesis explores the factors that influence experimentation and data capabilities to understand the conditions that support or hinder their development. These factors, which were inductively derived from interviews with Finnish and Estonian civil servants, are grouped into three higher-order dimensions: polity, policy, and politics (Lange et al., 2013). The *polity* perspective explores the factors that influence experimentation and data capabilities at the institutional level. Here, two of the most significant factors that emerged from the research were path dependency and the role of agency. The *policy* perspective focuses on experimentation as a learning mechanism for adaptive policymaking. The implications of different types of experiments are outlined and explained based on the cases of Estonian and Finnish governance institutions. The *politics* perspective examines the role and motivation of politicians in supporting and engaging in experiments, as policy experimentation is highly influenced by political settings.

## 4.1 Building Adaptive Public Sector Organisations: Interplay of Path Dependency and Agency

Adaptive governance is facilitated by flexible institutions that are able to learn in the face of uncertainty and complexity. The ability of public sector organisations to access new information and the capacity to act on that information, but also their willingness to take risks (McLain & Lee, 1996) influence the way they adapt to changing circumstances. This indicates that the two factors—existing capabilities that are formed through path dependency and agency that helps to overcome the status quo—have a significant role in adjusting to rapidly changing circumstances. Next, we explore these two factors in more detail.

### 4.1.1 Path Dependency: Shaping the Ability to Cope with Turbulence

All institutions possess a certain level of capabilities, capacities, and behavioural patterns that shape the range of options available for improvement. A wide range of studies has examined the antecedents and existing capabilities necessary for data-driven decision-making (Mikalef et al., 2017; Sharma et al., 2014). Similarly, the ability of public sector organisations to cope with rapidly changing conditions depends on their previous cumulative experience, practices, and knowledge.

This argument was supported by the case study of the ETCB, which revealed specific data capabilities that enhance the ability of public organisations to cope with suddenly changed conditions and the expectations of stakeholders (**Article II**). These include the ability to automate processes and data flows that support decision-making, provide open data in response to information requests from stakeholders, redistribute work to manage sudden spikes in the workload of certain officials, and deliver e-services when application volumes rise significantly and remote consultations are required. However, it was argued that the ability to provide these solutions in highly turbulent times and crisis conditions was only possible because the necessary competence and technological prerequisites were already in place in the studied organisations. In other words, once these capabilities are established, they reinforce themselves through accumulated experience and expertise, making it more feasible for organisations to deliver such solutions in rapidly changing conditions that require adaptation.

The same logic can be applied to experimentation, as the capabilities of public sector organisations shape their ability to experiment, learn, and collaborate. Besides data capabilities, various collaborative, communicative, and managerial capabilities are required for experimentation, as it often involves a large number of stakeholders: organisations, individuals, or other public officials at different levels of government (**Article I**; Stoker, 2010). For example, design experiments require close collaboration with stakeholders who may hold conflicting values and interests, which demands strong leadership, inclusion, and collaboration capabilities to reach consensus and ensure acceptability (**Article I**; **Article IV**; Bauknecht et al., 2020). Similarly, various data-analytical capabilities are necessary to effectively learn from experiments, especially from randomised controlled trials (RCTs) that require specialised competences to analyse data with complex statistical methods or to test hypotheses (**Article IV**), which in turn determine the relevance of the results of an experiment.

Furthermore, our results indicate that public sector organisations that have previously engaged in experimentation develop procedural know-how, tolerance for uncertainty, and mechanisms for learning from failure, thereby lowering the perceived risk of further policy experimentation. In Finland, the Sipilä Government's national policy agenda "Experimenting Finland" (Leino & Akerman, 2021) supported the integration of experimentation into the policymaking process and legitimised it as a learning strategy in the eyes of the public. By contrast, in Estonia, the lack of such experience may explain the prevailing hesitancy to launch large-scale experiments. (**Article III**) Similarly, past experience and accumulated knowledge may significantly influence the attitudes of the public and the media towards experimentation. Experiments that are conducted in a transparent manner and lead to visible learning outcomes may increase public and media tolerance towards new experiments, as these are then perceived as necessary. (**Article III**)

Therefore, path dependency, including cumulative experiences and knowledge, shape the conditions that may support or hinder experimentation and the level of data

capabilities. Path dependency may constrain adaptive capacity by reinforcing already established practices, processes, and routines. However, it does not emerge automatically from past trajectories and cumulative experiences but is actively constructed by actors who seize opportunities for change and purposefully work towards overcoming the constraints to develop further adaptive capacity. This indicates the importance of agency, especially the willingness, motivation, and incentives to experiment and develop data capabilities.

### 4.1.2 Managerial Factors and Agency: Seizing Opportunities for Adaptation

Overcoming prevalent patterns, improving existing capabilities, and ensuring long-term development in the pursuit of adaptive governance require strong leadership that questions established routines (path dependency) and is willing to mobilise resources for building adaptive capacity. Next, we explain the role of agency in the development of data capabilities and experimentation.

First, while organisational routines can be adapted temporarily by using the knowledge generated by data capabilities (**Article II**), the adaptive capacity of public sector organisations is supported by their continuous enactment. The role of agency was proposed as one of the most important factors influencing such long-term change in organisational routines (**Article II**). While organisations may face situations (e.g., crises) that open "windows of opportunity" to change existing routines with the support of data capabilities, agency ultimately shapes whether the underlying data capabilities that enable such change are developed or sustained in the long term. For example, although shifting tasks from people to machines (e.g., through automated data flows or modern technologies to perform certain tasks) provided considerable benefits in crisis conditions, the long-term development of these capabilities for modifying organisational routines depends on the willingness and motivation of the civil servants responsible for their execution **(Article II**).

Furthermore, the data capabilities that support change in organisational routines in the public sector are influenced by the legislative environment. While the COVID-19 pandemic induced routine adaptations, one reason these changes remained temporary was the reliance on short-term "legislative fixes" that enabled additional inter-organisational data use that did not continue after the pandemic (**Article II**). In other words, using data in other organisational routines and practices presupposes the development of further legal grounds, which is highly dependent upon agency and the willingness to pursue the change once the urgency has passed, but also political context and interests (see also Chapter 4.3). For example, building the surveillance capacity of public sector organisations could be considered politically sensitive, even when the effectiveness of certain routines that rely on access to more relevant data may appear obvious (**Article II**).

Second, **Article V** argues that managerial factors, which include the characteristics, actions, and practices of managers, strongly influence learning from experiments. For example, agency influences whether adequate planning takes place (Greenberg & Poole, 2007; Mostert et al., 2007) and whether sufficient time and attention are devoted to learning (Bos et al., 2013; Farrelly & Brown, 2011; Mattocks, 2021). Challenges in the planning or evaluation phase were frequently outlined by our interviewees, as they may undermine learning from experiments. For example, the planning phase may be too rushed, or the evaluation phase insufficient or absent altogether. Furthermore, agency defines whether experiments have clear ownership and well-defined goals and roles

(Armitage et al., 2008; Heikkila et al., 2024). Our results indicate that Estonian and Finnish civil servants consider commitment and ownership as important aspects that facilitate learning from experiments. Several Estonian civil servants also pointed out that the lack of clear goals and performance metrics has undermined learning from experiments, as without them, evaluation may remain subjective, allowing attention to focus only on positive aspects and limiting opportunities to learn from failure (**Article V**). Finnish civil servants suggested that RCTs require the establishment of clear goals and performance metrics that fosters learning. However, vaguely defined goals may challenge the evaluation of the effects of an intervention, as the target groups involved in the experiment may have different interests (**Article V**).

While public sector organisations leverage existing capabilities to design, implement, and evaluate policies, experimentation could also be considered a policy-level learning mechanism that enables to gather knowledge about the best courses of action (**Article I**). Next, we discuss the potential of experiments to enhance adaptive policymaking and the possibilities that various types of experiments provide for policy learning.

## 4.2 Experimental Governance: Iterative Learning for Adaptive Policymaking

It has been argued that adaptive policymaking relies on learning over time, as it operates on the best available information and continually experiments with alternatives to identify the best course of action (Nair, 2020). For this reason, policy experimentation has been proposed as a key strategy for managing increasingly complex policy challenges (**Articles I**, **III**, **V**), Furthermore, experimentation supports the inclusion of relevant stakeholders and helps avoid rigid or merely formal policy processes (e.g., consultations without inclusion) that would hinder their adaptive capacity (Cosens et al., 2017).

Next, this thesis examines the role of experimentation in policy learning and the characteristics of various experimental approaches.

### 4.2.1 Learning Underpins Adaptive Policymaking

Central to understanding the role of experimentation in adaptive policymaking is the question of learning, which helps to challenge the status quo—established "ideas, knowledge and normative assumptions" (Dockx et al., 2023, p. 12). Policy learning is a mechanism that enables to respond to complex issues (Zaki et al., 2023) and occurs when the understandings and beliefs related to public policy are revised and updated (Dockx et al., 2023; Dunlop & Radaelli, 2013; Goyal & Howlett, 2024). It is essential to the ability of the system to respond to new circumstances, which makes experimentation an important part of adaptive governance (Eshuis & Gerrits, 2021). Policy experiments serve as tools that help to intentionally gather and act upon new information, evidence, or data (**Article V**) and enable iterative learning. Therefore, experimentation helps to develop effective and adaptive policies in several ways (**Article I**):

- Experiments provide an opportunity to test a planned policy on a small scale, offering policymakers information about the impacts of the intervention in a real-world setting (Ettelt et al., 2015; Millo & Lezaun, 2006) and in interaction with actual target groups.
- Experiments provide an opportunity to develop evidence-based policies (e.g., Stoker & John, 2009) and allow policymakers to evaluate the effectiveness of different policy alternatives (Bravo-Biosca, 2020; Haynes et al., 2012; Nair & Howlett, 2016),

test possible scenarios, and assess their consequences and outcomes (**Article I**; Bailey et al., 2017) to identify the most suitable course of action. The value of experimentation becomes especially clear in complex situations where uncertainty is high and decision-makers lack complete information, especially regarding the impacts of the intervention (**Article I**; Ansell & Bartenberger, 2016; Bauknecht et al., 2020; Bravo-Biosca, 2020; Nair & Howlett, 2016).

- Experiments help to avoid large-scale failures and reforms with unanticipated results (**Article III**; Heilmann, 2008; Nair & Howlett, 2016). This includes the overall costs of a policy, as experimentation provides insight into competing policy options and enables policymakers to outline the alternatives with the highest value (**Article I**; Adkins & Ylöstalo, 2018; Checkland et al., 2021; Haynes et al., 2012) or to assess whether desired results emerge and whether unintended effects occur (**Article I**; Haynes et al., 2012; Nair & Howlett, 2016). While large-scale interventions take time and may be too rigid for constantly changing environments, the smaller-scale interventions that experimentation facilitates support adaptive policymaking.
- In addition, in situations where various actors interact and are involved, experiments help to reduce informational asymmetries and build acceptability among stakeholders (Nair & Howlett, 2016; Philipsen et al., 2021). They also reduce uncertainty related to the behaviour of the target groups of the policies (Burtless, 1995; Nair & Howlett, 2016), which may be dynamic and change over time.

Policymakers may choose between different types of experiments to learn about potential interventions and policy solutions. Next, we discuss the possibilities and challenges associated with different types of experiments and their potential in supporting adaptive policymaking.

### 4.2.2 Learning from Different Types of Experiments

Public sector organisations may choose from a variety of experimental designs, each with its own strengths and weaknesses. In this thesis, we discuss the differences between three types of experiments that differ in their learning logic: randomised controlled trials (RCTs), pilot projects, and design experiments.

First, *randomised controlled trials* (RCTs) are regarded as the gold standard of experimentation, as they provide a rigorous approach to testing the effects of a potential policy solution (e.g., Baldassarri & Abascal, 2017; Dur et al., 2025; Pearce & Raman, 2014) and generate results that are considered to be "unbiased", "internally valid", and "free from selection bias" (Greenberg & Morris, 2005, p. 229). RCTs prioritise causal explanations through the random assignment of subjects, dividing them into multiple groups (including a control group) that receive different treatments (Bravo-Biosca, 2020; Pearce & Raman, 2014). However, it has been argued that RCTs may place limits on policy learning (e.g., Ko & Shin, 2017; Straßheim, 2024), as they tend to test interventions with a relatively narrow scope, are challenged when impacts are heterogeneous, and may face problems of external validity (e.g., Bedecarrats et al., 2017; Ko & Shin, 2017).

Besides RCTs, generative approaches to experimentation suggest that using pilots and design experiments has advantages over RCTs in promoting learning (Ansell & Bartenberger, 2016). While they do not offer such a rigorous approach to experimenting, they are more adaptive in testing novel solutions, as feedback on their effects is gathered iteratively until desired results emerge (Bailey et al., 2017; Nair, 2020; Ko & Shin, 2017). Collecting information continuously and across several rounds contributes to updating policy beliefs and knowledge about the effects of a potential solution, which enables

policymakers to adapt it to changing circumstances. For example, *non-randomised pilots* allow the testing of a new policy on a small set of target groups or jurisdictions before scaling it up (**Articles I**, **III**; Farrelly, 2008; Nair & Howlett, 2016). While pilots may operate on a fairly short and flexible timescale, it has been argued that the time needed for piloting policy solutions may vary significantly depending on the scale, complexity, and other demands of the intervention. (**Article IV**; Nair & Howlett, 2016) In addition, *design experiments* focus strongly on stakeholder experience by providing an iterative method for gathering feedback. In design experiments, the policy solution is refined until the desired results emerge (**Article I**; Ansell & Bartenberger, 2016; Stoker & John, 2009). Just like pilots, design experiments can operate on various time scales, which indicates that they could also be used to develop short-term interventions. However, the full impact of design cycles may only become apparent after years of experimenting (**Article IV**; Stoker & John, 2009).

The contrast between how different types of experiments support learning was well-illustrated by Finnish and Estonian policymakers (**Article V**). In Finland, policymakers generally agreed that RCTs are the best interventions for providing information to support policy change, as they allow for in-depth evaluation of the impacts of the intervention, produce reliable and valid results, and therefore make the resulting knowledge easier to use in the decision-making process. For example, while the importance of RCTs in ensuring policy learning was emphasised, the interviewees pointed out that the findings of non-randomised pilots can be contested, as alternative explanations have not been ruled out. However, Finnish policymakers also acknowledged that learning from RCTs can be constrained when an intervention is exposed to shifts in context or to conditions of high uncertainty, as conducting RCTs requires the possible alternatives of action to be defined in advance. By contrast, Estonian policymakers argued that, alongside RCTs, other types of designs also provide opportunities for learning. This was illustrated by the fact that none of the interviewees in Estonia suggested RCTs as a precondition for learning from an experiment, while they discussed design experiments as "potentially the most fruitful" designs for gathering diverse information. While RCTs may be perceived as a convincing experimental method, they were not perceived as being superior to other types of experiments. (**Article V**)
Therefore, Finnish and Estonian civil servants have considerably different views when it comes to learning from experiments. Similarly, the political settings they operate in reflect different attitudes towards experimentation. To extend the discussion to the political dimension of adaptive governance, we discuss how political motivations influence whether public sector organisations engage in large-scale experiments.

## 4.3 Implications of Political Settings on Experimentation

Public sector organisations are always influenced by political settings, and by focusing on the political dimension, we can understand how the political context shapes and influences learning from experiments. Policy experiments cannot be considered politically neutral (Pearce & Raman, 2014, p. 398; **Article V**)—political context may define where attention is directed, which initiatives are supported with resources, which groups of society and policy problems are addressed with experimental approaches, and how ambitious these initiatives are (Meadowcroft, 2009). Next, we explore the motivations of politicians to engage in experimentation.

### 4.3.1 Policy Learning is Shaped by Political Context

While experiments provide policymakers with evidence (Stoker & John, 2009), there are several ways in which the political setting influences learning from experiments. As **Article III** demonstrated, the differences in the perceptions of Finnish and Estonian civil servants were significant—while half of the Finnish interviewees pointed to political considerations as crucial to shaping the process of learning from experiments, Estonian interviewees reported no experiences of such instances. More specifically, this thesis outlines three ways in which experimental approaches may clash with the prevailing logic of "political decision-making, policy and electoral cycles, electoral considerations, and value conflicts" (**Article I**, p. 32; Bauknecht et al., 2020; Bravo-Biosca, 2020; Burtless, 1995).

First, politicians may follow a certain logic in decision-making that outweighs the value of evidence in implementing policies—they may prioritise ideological considerations, voter preferences, and party agendas instead of the evidence that experiments provide (**Article I**). During the interviews, examples were provided where, due to political interests, insufficient time was dedicated to planning an experiment, or learning from experiments was skipped due to political considerations. For example, a large-scale policy pilot in Finland concerned with delegating employment services from the central government to municipalities was scaled up two years before the experimental evidence was due to be published. While the preliminary results did not show any significant differences between the experimental group and the control group—and therefore did not confirm the positive impacts of the intervention—the decision to scale the solution was politically motivated and did not follow an experimental logic (**Article V**). Furthermore, it has been proposed that policy agendas may change when there is a time lapse between the launch of an experiment and the receipt of the results, which means that policy actors may become less motivated to learn from the experiment over time (Oakley, 1998). To illustrate this argument, our interviewees pointed out that when a change in government takes place, the experiments of the previous government may not attract as much interest, which hinders learning from them (**Article V**).

Second, it is noted that politicians may not seek to revise and update their policy beliefs with the help of experiments but rather use them to confirm pre-existing ideas (Brodkin & Kaufman, 2000; Sanderson, 2002) or to advocate for policy decisions already made (Huitema et al., 2018; Sanderson, 2002). The results of our study (**Article III**) revealed that if experimental evidence contradicts dominant political ideology or governmental agenda, the results of the experiment do not change the existing views—instead, the gathered evidence is ignored and learning terminated. It was also pointed out that while researchers view experiments as tools for generating knowledge, politicians want experiments to deliver results and expect "to find at least something that did work" (**Article V**). While it has been argued that politicians may even pressure the evaluators of experiments to present positive evidence for solutions they prefer (Nair & Howlett, 2016), our interviewees did not outline specific examples in which direct political pressure was used to shape the results of the experiment.

Third, politicians may be concerned about the signals that experimenting in policymaking sends to the electorate (**Article III**), as experiments entail a risk of failure, which could be used to blame them for such failures. While experimentation could be regarded as a valuable strategy for providing *ad hoc* solutions to pressing policy problems, the public may see policy experimentation as a way to postpone important decisions, waste time, or perceive it as "doing too little" to solve a salient problem (**Article I**). Also, testing a policy before launching a large-scale reform may help gather

valuable information and avoid potential large-scale failures, but it also provides an opportunity to demonstrate interest in socially relevant policy problems without initiating a comprehensive reform (**Article III**; Heilmann, 2008; Nair & Howlett, 2016). This may be politically beneficial if addressing a particular problem is associated with unpopular, or even painful, decisions (**Article III**; Howlett, 2014; Nair & Howlett, 2016). Our results confirmed that political considerations may support launching experiments, but that does not mean that policy actors are interested in learning from them. Instead, experiments can be launched for other reasons (e.g., for stalling decisions, as bargaining compromises, as signals for taking action) (**Article V**).

Considering that experimentation involves admitting to a lack of existing knowledge about the possible impacts of a proposed solution (Bauknecht et al., 2020; Farrelly, 2008), politicians may hesitate to acknowledge such uncertainty. Therefore, elected officials often weigh decisions through a blame-credit calculus: they prefer actions that maximise visible credit and minimise exposure to blame (**Article III**).

### 4.3.2 Avoiding Blame or Claiming Credit: Political Motivations for Launching Experiments

While overcoming the contradictions between political settings and experimentation could be key to promoting adaptive governance, the motivations of politicians to engage in experimentation require deeper insight. It is argued that political actors expect that claiming credit leads to more votes, while receiving blame damages their electoral prospects (Hansson, 2018; Howlett, 2014; Leong et al., 2023). For that reason, "they seek to avoid failures for which they can plausibly be held responsible and will claim credit for policy successes" (Leong & Howlett, 2017, p. 603). The mechanisms of blame avoidance and credit claiming are argued to influence politicians' decisions (e.g., Hinterleitner, 2017; Howlett, 2014; Leong & Howlett, 2017; Leong et al., 2023), including the decision to launch large-scale policy experiments. There are several specific factors that influence these blame avoidance and credit claiming mechanisms, which often underpin political decision-making.

First, the urgency of a policy problem or the saliency of the policy field affects policymakers' blame-credit calculations (Howlett, 2014; Leong & Howlett, 2017) when deciding whether to initiate an experiment. If the problem is not visible, policymakers may not consider it important enough to pursue wide reforms, and taking small steps through experimentation could be seen as dealing with the problem, which gives policymakers an opportunity to glean credit or avoid blame (Howlett, 2014; Leong & Howlett, 2017). By contrast, when the policy problem is salient and urgent (such as a crisis response), policymakers may prefer large-scale actions over small steps (Howlett, 2014), as these offer the possibility of greater credit. As the results of **Article III** demonstrate, Estonian officials indicate that delays in solving urgent issues that experiments may introduce could be seen as problematic, as they may tempt politicians to implement solutions immediately instead of experimenting. Furthermore, while in some cases, Finnish officials have been able to overcome pressures to implement immediate solutions to urgent policy problems and launched large-scale experiments, it is still acknowledged that crises create pressures that undermine experimentation (**Articles III**, **IV**).

Second, the scope of the experiment could also affect the blame-credit calculus, as large-scale policy experiments attract public attention (Oakley, 1998), which may amplify the size of projected credit or blame. Our results indicate that in Finland, politicians

do not hesitate to launch large-scale experiments with high visibility. In Estonia, by contrast, politicians have not launched large-scale experiments and have garnered some political credit from small-scale design experiments.

Third, existing literature and the results of **Article III** demonstrate that the media is perceived as one of the relevant factors in shaping perceptions of blame avoidance and credit claiming (see, for example, Hinterleitner & Sager, 2015). As the media tends to reinforce negativity bias and is willing to attribute blame to individual politicians, it likely affects their blame-credit calculus in experimental policymaking (**Article III**). Our results indicate that in Finland, the fear of negative media attention has not prevented or discouraged policy experiments. However, Estonian officials considered the anticipated media attention to experimentation and failure as an important mediator of blame, including expectations of amplified negative media attention. It was also pointed out that politicians are focused on avoiding scandals and unfavourable attention in the media. (**Article III**) Furthermore, the results of our study indicate that if experiments become an important part of the policymaking process and they are legitimate in the eyes of the public, initiating and conducting experiments may feel less risky for policymakers. It was well illustrated by our empirical results, where Finnish officials pointed to their past experience with policy experiments, which were strongly supported by the Sipilä Government's "Experimenting Finland" programme that legitimised the use of experiments in the public sector and led to greater tolerance of failure in the pursuit of learning. By contrast, Estonia lacks experience with large-scale experiments, and, as a result, politicians are perceived as more hesitant to launch experiments (**Article III**).

Fourth, it is argued that public trust influences governments' and politicians' blame avoidance behaviour towards policy experiments (**Article III**; Boin et al., 2009; Hinterleitner & Sager, 2015). Our results reveal that trust is an important factor in explaining the comparative outcomes of Estonia and Finland. The governments in societies with high trust may be more likely to experiment, as they are more successful in advancing their own narrative, and this high level of trust also fosters public confidence that the government is experimenting for a useful purpose. Similarly, policymakers may feel that high public trust provides them with room for failures without being blamed. Several Finnish interviewees argued that without public trust, experimentation would be greatly hindered. Estonian public officials pointed to the lack of trust in government and the perception that launching experiments would generate negative reactions from the public. In Estonia, it was perceived that experiments that end in failure are considered a waste of time and public resources, rather than a learning opportunity, which stands in contrast to the views of Finnish public officials. (**Article III**)

As the results of **Article III** demonstrate, there is a stark contrast between Estonian and Finnish civil servants regarding their perceptions of the credit and blame associated with public policy experiments. While in Finland public policy experiments are largely regarded as a way to avoid blame, Estonian public officials perceive policy experiments as predominantly "risky" and as entailing considerable potential for blame. The results of our study illustrate how the perceptions of the same learning mechanism may differ due to various factors, including institutional and political settings.

Public institutions operate in increasingly turbulent environments, which demand quick adaptation to unexpectedly changing circumstances. Next, we further discuss the implications of crises on experimentation and data capabilities to understand whether, and under what conditions, crises may support adaptive governance.

# 5 Implications of Crises: Leveraging Data Capabilities and Experimentation

Public sector organisations around the world operate in an increasingly turbulent environment, which has led to a continuous need to prepare for and keep up with crises that challenge established administrative routines and management techniques (Willi et al., 2020; Christensen et al., 2016; Lenz, 2024) and often shift the priorities of policymaking and public administration. Crises give rise to a state of urgency that stems from the pressure of external stakeholders (e.g., citizens, politicians, or other governmental agencies) to find more effective solutions (Hogan & Feeney, 2012) or from the recognition that current ways of working are ineffective or unsuitable for crisis response. This creates fertile ground for seizing "windows of opportunity" (Hooren et al., 2014; Caprotti & Cowley, 2017) to implement novel solutions. Our results confirm that extraordinary times call for thinking differently, because there is no option to continue doing what has always been done (**Article IV**). Such periods push public sector organisations towards new solutions and increase the willingness of organisations and policymakers to overcome inertia (**Articles II, IV**).

Data capabilities and experiments can help mitigate the implications of crises, as they reduce uncertainties and help gather information, thereby supporting decision-making and enabling change motivated by sudden urgency. However, crisis conditions have several implications for experimentation and the potential of data capabilities to induce long-term routine adjustments.

## 5.1 Crisis Conditions and the Logics of Experiments Create Paradoxes

Given that public sector experiments support organisational learning and provide evidence in uncertain situations (Willi et al., 2020), they could be viewed as a crucial approach in crisis governance. However, conducting policy experiments during a crisis can present various challenges that stem from the interplay between the experimentation approach and crisis conditions. More specifically, we identified four paradoxes—"contradictory yet interrelated elements that exist simultaneously" (Smith & Lewis 2011, p. 386)—that help us to understand the possibilities for conducting experiments in crisis conditions.

First, experimentation and crisis conditions create conflicting timeframes. In other words, crises create a sense of urgency, whereas experiments take time. While experimentation may be "a key activity to practice the effectual logics of aiming to control rather than predict uncertain events" (Björklund et al., 2020, p. 7), our interviewees also noted that the pressure from a diverse set of stakeholders creates impatience when developing crisis responses, and that quick, visible results are preferred (**Article IV**). During crises, governments and public organisations are pressured to make rapid decisions and develop new policies to avoid further escalation of the negative impacts of crises (Bessant et al., 2015; Sharma et al., 2022). While learning is considered to be one of the most important goals of experiments (van der Heijden, 2018), during crises, it is not enough to learn "what does not work". Our interviewees pointed out that during crises, the pressure for fast decision-making and successful solutions is exceptionally high, and solutions are developed within "hours and days rather than months or years" (**Article IV**).

Second, crises bring about an intolerance for failure, but failure and learning from mistakes are inherent to the experimental process. Although experiments may provide valuable insights for deciding about the best policy solution, experiments also require a

safe space where failure is tolerated. Nair and Howlett (2016) indicate that under uncertainty, policymakers face a considerable challenge of making decisions with limited, incomplete, or no information about the future policy environment. However, to restore a sense of control, they are expected to act with certainty and avoid failure. This creates a paradoxical situation where certainty is expected despite uncertain conditions. Moreover, politicians may prefer to appear decisive and competent in addressing crises, reducing their incentives to launch risky policy experiments (Bils & Izzo, 2022, p. 1). Similarly, **Article III** indicated that elected representatives may want to be perceived as bold and taking charge, which may lead them to steer clear of experimentation to avoid blame for delays.

Third, crises bring about a lack of certain resources that are crucial for conducting experiments, especially cognitive capacity. While our interviewees generally agreed that crises induce innovative thinking and the development of novel solutions, they agreed that experiments require significant cognitive resources (**Article I**, **IV**) that may be scarce during crises. Individuals working in crisis conditions face complex and unfamiliar situations and their decisions have high stakes (Dionne et al., 2018). As crises frequently hit specific areas and require deep subject knowledge and competence to provide solutions, some civil servants may be even more prone to burnout (**Article II**), or such expertise may be difficult to find.

Fourth, experiments help to understand the effects of a policy solution, but the temporality of crises may limit the wider use of their results, creating limitations of scalability. It has been argued that even when experiments show promising results, scaling them up may prove to be problematic since conditions change rapidly. Experiments conducted during a crisis have the potential to become "seeds for transitions" (Meinherz & Fritz, 2023, p. 509); however, they are often context-specific and limited in their replicability (Nair & Howlett, 2016), helping to achieve small wins instead of institutional changes (van den Ende et al., 2024; Savaget et al., 2024; van der Heijden, 2018). Therefore, while crises provide "windows of opportunity" (Hooren et al., 2014; Caprotti & Cowley, 2017) to undertake policy changes, our interviewees argued that crises create extraordinary working conditions, and the solutions developed as crisis responses may not be applicable once post-crisis normalcy returns. For example, it was pointed out that experiments that aim to produce long-term value may presuppose stability, whereas crises are temporary. As the saying goes, "extraordinary times call for extraordinary measures", and solutions from experiments conducted during crises are unlikely to last beyond the crisis. (**Article IV**)

These four paradoxes significantly influence how policymakers perceive the feasibility of conducting experiments during crises. However, different types of experiments provide different kinds of learning and may therefore equip them with the means to overcome these tensions. Based on the analysis of these four paradoxes, we argue that in crisis conditions, policymakers may feel more inclined to use generative experiments (pilot projects, design experiments) that provide greater flexibility in rapidly changing conditions than randomised controlled trials (RCTs). While their fixed and reliable methodology makes RCTs the "gold standard" of experimentation, they are very time-consuming and rigid at the same time (Bell & Peck, 2016; **Article IV**) and may not provide public sector organisations with the means to adapt to rapidly changing conditions.

While generative experiments may help policymakers identify the best policy solutions, data capabilities may also support the adjustment of existing organisational routines

when new information emerges. Next, this thesis explains how the data capabilities of the Estonian Tax and Customs Board enabled temporary routine adaptations and provided insights for future capability building to enhance adaptive capacity.

## 5.2 Temporality of Crises Leads to Temporary Routine Modifications

While crises create a sense of urgency and a rapid need to change and improve organisational routines, this thesis aims to understand whether such temporary conditions may generate long-term routine changes. **Article II** revealed that the COVID-19 pandemic induced three types of changes:

- *Temporary routine changes* — the sense of urgency faded after the crisis. For example, the ability to automate data flows and decision-making helped public sector organisations provide information to various stakeholders. We found evidence of organisations that had been examining individual financial aid applicants manually and were able to switch to mass decisions. However, it was noted that making mass decisions in pandemic-related areas and providing services to novel target groups were temporary solutions, and there was no need for these actions to continue afterward. Similarly, during the COVID-19 pandemic, public sector organisations faced the need to acquire in-depth knowledge (e.g., technical specifics, data structures, combined with expert knowledge of tax legislation) in areas they had not previously encountered and would not need after the crisis. For example, the results of **Article II** outlined that the legislative environment restricted long-term changes in routines, as certain legislative solutions that were required for the pandemic response were developed for crisis conditions only.
- *On-demand solutions were developed* — capacity to provide certain solutions increased but did not become routinised after the crisis. For example, the ability to provide simple dashboards of detailed data, combine it with macro-level indicators, and provide more real-time data became crucial in the face of crisis. Although most of the automatic management reports that were outlined in the results of **Article II** were not used after the crisis, they could be activated on demand—creating the possibility to monitor the economic situation should the need for rapid decision-making arise again.
- *Long-term routine adjustment*. This included, for example, the increased capacity to provide open data that helped to cope with increased workload, enquiries, and information requests from stakeholders. More specifically, the data of fuel sales was made available on the ETCB homepage to reduce the number of individual enquiries. This data continued to be provided after the pandemic, which indicates a continuous routine change. (**Article II**) Although the examples are small-scale, they illustrate that during crises, long-term routine adaptations may occur.

As **Article II** argues, long-term changes in the studied routines were rather rare due to the temporality of crisis conditions. Also, crises bring about an increased need to prioritise crisis response, which means that projects aiming for long-term value or tackling complex societal problems (e.g., demographic challenges), may be discontinued and replaced by developing *ad hoc* solutions that help mitigate the crisis at hand (**Article IV**). Therefore, while experimentation and data capabilities provide public sector organisations with valuable means to sense their surroundings and seize the opportunities they provide, their potential to support long-term and continuous routine adaptation is limited by the temporality of crisis conditions

# 6 Concluding Discussion

Adaptive governance has emerged as a concept in response to uncertain and complex environments that entail interconnectedness, non-linear interactions, and persistent change (Scarlett, 2013). Such conditions have led to a need to make sense of how public sector organisations adapt to their surroundings, how they gather information, learn, and eventually change. The aim of this thesis was to explore experimentation and data capabilities as sources of dynamism that support the adaptive capacity of public sector organisations and explore the factors that influence their wider adoption and institutionalisation in the pursuit of adaptive governance. This thesis argues that while public sector organisations may use data capabilities and experimentation to enhance their adaptive capacity by enabling modifications to routines, the institutionalisation of adaptive capacity into governance practice, which produces adaptive governance, may be hindered by several factors arising from the dimensions of polity, policy, and politics. Next, the main results are discussed based on the proposed research questions.

## 6.1 RQ1: How Do Data Capabilities and Experimentation Help Public Sector Organisations to Adapt?

This thesis combined the concepts of experimentation and data capabilities and used the dynamic capabilities framework to explore the underlying mechanisms through which public sector organisations adapt to their surroundings. As illustrated in Figure 2 (see p. 19), experimentation and data capabilities enable public sector organisations to sense their surroundings, seize opportunities, and transform routines in order to adapt to changing conditions. However, the two capabilities differ in their underlying logic.

First, while data capabilities support an organisation's ability to detect changes in its environment and thereby increase its sensing ability, experimentation helps to assess the implications of potential solutions and proposed ideas. Experiments are usually launched when there is already enough information on potential solution(s) that need to be tested, setting them apart from data capabilities, which support organisations in sensing changes in their environment that require a response. Second, data capabilities provide input for decision-making related to building competences or mobilising resources to address detected problems and opportunities, enabling public sector organisations to seize opportunities and implement the best solutions. Similarly, conducting experiments contributes to seizing opportunities as it helps to understand and decide upon the different paths of action. However, changes made during experiments are reversible and temporary, while routine adaptations may not be implemented with such precondition. Third, data capabilities enable the modification of existing organisational routines to adapt to changing conditions and support transformation. However, modifying or creating new routines through experimentation is depends on whether the intervention is scaled. Therefore, while experimentation and data capabilities provide public sector organisations with the means to sense and seize opportunities, the transformation phase and continuous adaptation to changing conditions do not automatically follow and remain dependent on other factors, often stemming from different dimensions of governance (polity, policy, politics).

Therefore, experimentation and data capabilities could be seen as mutually reinforcing capabilities. Data capabilities may help public sector organisations detect problems that require experimentation to find the best solution. High-level data capabilities enable the analysis and evaluation of the effects of experiments in depth and

ensure the reliability of the results. Vice versa, experiments may support public sector organisations in focusing data capabilities on the most pressing and strategic problems. Experiments may also provide insights into where data capabilities should be developed to increase organisational learning. Therefore, experimentation and data capabilities support public sector organisations' adaptive capacity, which, when becoming routinised and institutionalised for decision-making and policy formulation, supports the development of adaptive governance.

## 6.2 RQ2: Which Factors Influence Experimentation and Data Capabilities?

RQ1 explored data capabilities and experimentation as mechanisms that enable public sector organisations to sense the surrounding environment and to respond to these changes. However, it is argued that their potential to bring about long-term transformation of organisational routines in ways that ensure continuous renewal (**Article II**; Conboy et al., 2020; Teece, 2007) may be hindered by several factors across the polity, policy, and politics dimensions of governance. RQ2 examined these factors to provide insights into the conditions that support adaptive governance, as well as the challenges that hinder it. For greater analytical clarity, these inductively derived factors were grouped into three broader dimensions: polity (organisational settings), policy, and politics.

First, the interaction between path dependency and agency was outlined as a crucial factor influencing experimentation and the development of data capabilities at the organisational (polity) level.

Factors related to existing capacities and experience influence the deployment of experimentation and data capabilities. For example, it was outlined in **Article II** that existing routines made it easier to respond to the expectations induced by the COVID-19 pandemic. However, public sector organisations that lacked the capabilities needed for such adaptation found it difficult to build them when the crisis hit (**Article II**). Furthermore, as the empirical results of **Article I** demonstrate, cumulative experience with experimental policymaking has an important role in launching policy experiments. In other words, if a government has prior experience with experimentation and it has become an integral part of the policymaking process, it is easier to launch large-scale policy experiments in the future. By contrast, limited experience with experimentation or previous ambiguous outcomes may foster scepticism and resistance to experimentation.

Besides path dependency, the role of agency was outlined by our interviewees as crucial to overcoming the status quo. While experimentation and data capabilities provide information and opportunities for learning, public officials' "willingness, motivation, and incentives to change organizational routines" (**Article II**, p. 9) often define whether the development of certain capabilities becomes strategically coordinated and purposeful. For example, **Article II** demonstrated that it is easier to respond to crises when an organisation wields high-level data capabilities (path dependency) and the capabilities are further developed when it is able to seize the "windows of opportunity" revealed by crisis conditions. Similarly, when launching public policy experiments, factors related to agency, especially ownership, are expected to boost learning, while it can be undermined by unclear roles and insufficient commitment (**Article V**).

In the policy dimension, experimentation provides a learning mechanism to challenge the status quo. Experiments help to gather iterative feedback on changing conditions so as to avoid large-scale failures and choose the policy solution with the highest value. However, various contextual factors influence how civil servants perceive the possibilities of using experimentation as a policymaking tool. The results of this thesis indicate that even in similar geographical and institutional contexts (for example, Finnish and Estonian civil servants), perceptions of the feasibility of experiments may vary considerably (**Article V**).

In Finland, policymakers generally agreed that RCTs are the best interventions to provide information for policy changes, as they allow for an in-depth evaluation of the impacts of the intervention, their results are reliable and valid, and therefore the resulting knowledge is easier to use in the decision-making process. However, it was acknowledged that learning from RCTs can be challenging in rapidly changing contexts or under high uncertainty. By contrast, Estonian policymakers agreed that besides RCTs, other types of designs provide opportunities for learning. None of the interviewees in Estonia suggested RCTs as a precondition for learning and RCTs were not perceived as being superior to other types of experiments. (**Article V**) Therefore, while launching and conducting experiments, policymakers have an opportunity to decide whether to focus on the reliability and validity of the results that RCTs provide, or to choose generative experiments that enable flexibility and opportunities to learn under rapidly changing conditions.

Third, political motivations influence experimentation and can significantly influence decisions to launch large-scale policy experiments. As illustrated in **Article III**, an important driver of political motivations for launching experiments is the credit-blame calculus: the calculation between the potential for credit claiming and the risk of being held responsible for failures. In Finland, policy actors generally regard experiments as sources of credit, because testing the solution before implementing large-scale reforms is seen as a rational and cautious approach to complex problems. By contrast, in Estonia, experiments are perceived as politically risky instruments because they generate few opportunities for credit claiming, while potential failure is perceived to create considerable blame. As a result, Estonian policymakers feel that society expects them to innovate without experimenting, while in Finland, experimentation is preferred as it provides a safe space to learn from failures.

Furthermore, public trust and the reactions of the media need to be emphasised, as they play a significant role in politicians' support for experimentation. If experimentation is legitimised in the eyes of the public and takes place in a society with a high level of trust, experimenting and learning from failures may feel less risky. This is especially relevant for political actors, as expected media reactions may define their risk tolerance when engaging in experimentation.

*Table 4. Factors influencing experimentation and the development of data capabilities.*

| Influencing factor | How does it influence adaptive governance? |
|---|---|
| **Path dependency.** Previous experience with experimentation and the existing level of data capabilities. | Defines the *status quo* and the level of existing capabilities, which influences the possibilities or sets the limits for the development of adaptive capacity. Adaptive capacity is supported by capabilities that provide public sector organisations with the means to sense their surroundings, provide input for seizing, and may lead to the transformation of routines. |
| **Agency**. Ownership, willingness, and motivation to change. | The development of adaptive capacity depends on whether capabilities are continuously renewed and executed, as agency helps to overcome path dependency. While crisis conditions provide "windows of opportunity" and create urgency for change, managerial motivations help to overcome the *status quo* and often determine whether the capabilities are developed in the long term. |
| **Learning** enables organisations to challenge the *status quo*. | Routinised data capabilities provide organisations with constant feedback from their surroundings, which may lead to routine adaptations where necessary. Iterative learning from experiments provides public sector organisations with the information on the most suitable courses of action under conditions of limited or no information. Learning leads to change and adaptation when the scaling phase of experiments occurs. |
| **Type of an experiment**. RCTs, pilot projects, design experiments. | The type of an experiment defines the learning logic. Choosing the wrong method may hinder policy making as it may not provide reliable input for decision-making. While RCTs provide valid and reliable information in tackling long-term policy problems, in turbulent conditions, generative experiments are seen as a more fruitful means to learn about the surroundings. Adaptive capacity increases and routines are changed only when scaling phase occurs. |
| **Political settings.** Motivations of politicians and the blame-credit calculus. | Political motivations and the blame-credit calculus may either promote or hinder the use of experiments, depending on whether experiments are seen as opportunities for credit claiming or sources of blame. Political settings influence all phases of experimentation: whether experiments are launched, how they are conducted, how learning takes place, and whether they are scaled. Therefore, adaptive governance is shaped by whether experimentation is seen as a valuable policymaking tool that needs to be institutionalised or as something to be avoided. |
| **Crisis conditions** | Crises create a state of urgency to respond to pressing policy problems. However, conducting experiments in crisis conditions creates a set of paradoxes that limit the possibilities available for experimentation. The routine adaptations made in crisis conditions tend to remain rather temporary. |

Building experimentation and data capabilities in the long term is related to the costs and benefits of *ad hoc* problem solving (Winter, 2003), which means that building and maintaining dynamic capabilities could be relatively costly, and at times, providing temporary solutions to temporary challenges may be more efficient. However, building adaptive governance requires routinised and institutionalised capabilities. It has been argued that while adaptive governance may be easily institutionalised at a cognitive level, the real transformational and enduring change will not follow when the "underlying ideas regarding governance remain unchanged" or when adaptive governance is pursued solely as a pilot (Eshuis & Gerrits, 2021, p. 292). As illustrated by the comparative studies of Finnish and Estonian governance institutions, the strategic and politically supported experimentation framework ("Experimental Finland" under Sipilä's Government) has significantly helped to legitimise the approach and to lower the perceived risks associated with experimentation.

Besides these factors, crisis conditions have further implications for experimentation and the development of data capabilities. While crises may reveal opportunities for improvement, the temporality of urgency may limit long-term routine adaptations.

## 6.3 RQ3: How Do Crisis Conditions Influence Building Data Capabilities and Experimentation?

This thesis argues that the level of existing capabilities shapes the ability of public sector organisations to modify their routines in response to a crisis (**Article II**). Given that established routines go through an adaptation process until they stabilise (Razak, 2017, p. 12), this thesis argues that crises may not provide the necessary timeframe for new routines to become stable and repetitive. As **Article II** illustrated, when the "state of urgency" created by the crisis subsided, the changes made to routines were reversed, thereby hindering long-term changes from taking effect. Certain data capabilities outlined in **Article II**, however, may increase organisational dynamic capabilities in turbulent conditions in the future (e.g., the ability to provide open data, automate data flows, provide simple presentations of data). This indicates that while high-level data capabilities support crisis response, such extraordinary conditions may reveal a previously unnoticed need for certain capabilities, as well as gaps in competences or processes that, when strategically developed, may lead to further capability building towards adaptive governance.

Similarly, this includes the capabilities to conduct generative experiments, which **Article IV** identifies as being more flexible and effective than RCTs for addressing rapidly changing crisis conditions. However, **Article IV** also highlights the limitations of scalability as one of the paradoxes that emerge from the interplay between crisis conditions and experimentation logics. This indicates that during crises, when surrounding conditions change rapidly and uncertainty is high, experiments may not provide useful insights for non-crisis contexts.

The results of this thesis indicate that experimentation and data capabilities help to modify the routines of public sector organisations so as to respond to rapid changes in their environments. These modifications take place within existing (and path dependent) organisational routines, which means that the level of existing capabilities influences how challenging it is to achieve routine adaptations and novel solutions in the face of crises. However, this thesis suggests that while experimentation and data capabilities

provide public sector organisations with valuable means to sense their surroundings and seize the opportunities they provide, they may be limited in their potential to provide long-term and continuous routine adaptation due to the temporality of crisis conditions.

## Limitations and suggestions for further research

The research presented in this thesis has several limitations.

First, the focus of the current thesis is primarily on the fiscal and financial policy field. Broadening the research to other areas would help to expand knowledge on the factors that influence the development of adaptive governance. Second, the research consists of only one case study aimed at understanding the development of dynamic capabilities. This indicates that the insights into the ways data capabilities induce changes in organisational routines may be limited. Additional case studies would be needed to understand whether the results of this study can be generalised. Third, further research should be conducted to understand what constitutes an experimentation capability. While this thesis has examined data capabilities and outlined the specific processes and actions that contribute to data capabilities, experimentation capability would benefit from similar attempts to enhance our understanding of the concept.

Further research should be conducted to understand the potential of experimentation and data capabilities to enhance adaptive governance. The research would benefit from the inclusion of a broader selection of policy fields to understand different logics and the variety of factors that influence experimental governance. In addition, exploring the experiences of other states or regions with different experiences, cultural backgrounds, and institutional settings may provide valuable insights into this stream of research. Finally, a more comprehensive approach should be developed that incorporates different types of crisis and non-crisis conditions to identify additional factors and their interactions that support or hinder the development of adaptive governance

## List of Figures

# List of Tables

## References


Adkins, L., & Ylöstalo, H. (2018). Experimental policy, price and the provocative state. *Distinktion: Journal of Social Theory*, *19*(2), 152–169.

Ambrosini, V., Bowman, C., & Collier, N. (2009). Dynamic capabilities: An exploration of how firms renew their resource base. *British Journal of Management*, *20*, S9–S24.

Ansell, C. K., & Bartenberger, M. (2016). Varieties of experimentalism. *Ecological Economics*, *130*, 64–73.

Ansell, C., & Trondal, J. (2018). Governing turbulence: An organizational-institutional agenda. *Perspectives on Public Management and Governance*, *1*(1), 43–57.

Armitage, D., Marschke, M., & Plummer, R. (2008). Adaptive co-management and the paradox of learning. *Global Environmental Change*, *18*(1), 86–98.

Bailey, S., Checkland, K., Hodgson, D., McBride, A., Elvey, R., Parkin, S., Rothwell, K., & Pierides, D. (2017). The policy work of piloting: Mobilising and managing conflict and ambiguity in the English NHS. *Social Science & Medicine*, *179*, 210–217.

Baldassarri, D., & Abascal, M. (2017). Field experiments across the social sciences. *Annual Review of Sociology*, *43*(1), 41–73.

Bauknecht, D., Bischoff, T. S., Bizer, K., Führ, M., Gailhofer, P., Heyen, D. A., Proeger, T., & Von der Leyen, K. (2020). Exploring the pathways: Regulatory experiments for sustainable development–An interdisciplinary approach. *Journal of Governance & Regulation*, *9*(3), 49–71.

Baxter, P., Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, *13*(4), 544–559. Available at: https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1573&context=tqr/

Bédécarrats, F., Guérin, I., & Roubaud, F. (2019). All that glitters is not gold. The political economy of randomized evaluations in development. *Development and Change*, *50*(3), 735–762.

Bell, S. H., & Peck, L. R. (2016). On the feasibility of extending social experiments to wider applications. *Journal of MultiDisciplinary Evaluation*, *12*(27), 93–111.

Bessant, J., Rush, H., & Trifilova, A. (2015). Crisis-driven innovation: The case of humanitarian innovation. International journal of innovation management, *19*(6), 1540014.

Bils, P., & Izzo, F. (2022). *Policymaking in times of crisis* [Working Paper].

Björklund, T. A., Mikkonen, M., Mattila, P., & van der Marel, F. (2020). Expanding entrepreneurial solution spaces in times of crisis: Business model experimentation amongst packaged food and beverage ventures. *Journal of Business Venturing Insights*, *14*, e00197.

Boin, A., 't Hart, P., & McConnell, A. (2009). Crisis exploitation: political and policy impacts of framing contests. *Journal of European Public Policy*, *16*(1), 81–106.

Bos, J. J., Brown, R. R., & Farrelly, M. A. (2013). A design framework for creating social learning situations. *Global Environmental Change*, *23*(2), 398–412.

Bravo-Biosca, A. (2020). Experimental innovation policy. *Innovation Policy and the Economy*, *20*(1), 191–232.

Brem, A., Nylund, P. A., & Roshani, S. (2024). Unpacking the complexities of crisis innovation: a comprehensive review of ecosystem-level responses to exogenous shocks. *Review of Managerial Science*, *18*(8), 2441–2464.

Brodkin, E. Z., & Kaufman, A. (2000). Policy experiments and poverty politics. *Social Service Review*, 74(4), 507–532.

Brunner, R. D. (2010). Adaptive governance as a reform strategy. *Policy Sciences*, *43*(4), 301–341.

Burtless, G. (1995). The case for randomized field trials in economic and policy research. *Journal of Economic Perspectives*, *9*(2), 63–84.

Caprotti, F., & Cowley, R. (2017). Interrogating urban experiments. *Urban Geography*, *38*(9), 1441–1450.

Chaffin, B. C., Gosnell, H., & Cosens, B. A. (2014). A decade of adaptive governance scholarship: synthesis and future directions. *Ecology and Society*, *19*(3).

Christensen, T., Lægreid, P., & Rykkja, L. H. (2016). Organizing for crisis management: Building governance capacity and legitimacy. *Public Administration Review*, *76*(6), 887–897.

Checkland, K., Hammond, J., Coleman, A., Macinnes, J., Mikelyte, R., Croke, S., Billings, J., Bailey, S., & Allen, P. (2023). "Success" in policy piloting: Process, programs, and politics. *Public Administration*, *101*(2), 463–480.

Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2020). Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *European Journal of Operational Research*, *281*(3), 656–672.

Cosens, B. A., Craig, R. K., Hirsch, S. L., Arnold, C. A. T., Benson, M. H., DeCaro, D. A., Garmestani, A. S., Gosnell, H., Ruhl, J. B., & Schlager, E. (2017). The role of law in adaptive governance. *Ecology and Society*, *22*(1), 1.

Davis, D. D., & Korenok, O. (2023). Policy experiments and financial regulation: Using laboratory methods to evaluate responses to the 2007–2009 financial crisis. *Journal of Economic Surveys*, *37*(3), 657–687.

Deverell, E. (2010). *Crisis-induced learning in public sector organizations* (Doctoral dissertation, Försvarshögskolan (FHS)).

Dietz, T., Ostrom, E., & Stern, P. C. (2003). The struggle to govern the commons. *Science*, *302*(5652), 1907–1912.

Dionne, S.D., Gooty, J., Yammarino, F. J., & Sayama, H. (2018). Decision making in crisis: A multilevel model of the interplay between cognitions and emotions. *Organizational Psychology Review*, 8(2–3), 95–124.

Dockx, E., Verhoest, K., Langbroek, T., & Wynen, J. (2023). Bringing together unlikely innovators: do connective and learning capacities impact collaboration for innovation and diversity of actors? *Public Management Review*, *25*(6), 1104–1127.

Duchek, S. (2014). Growth in the face of crisis: The role of organizational resilience capabilities. *Academy of Management Proceedings*, No. 1, 13487.

Dunlop, C. A., & Radaelli, C. M. (2013). Systematising policy learning: From monolith to dimensions. *Political Studies*, *61*(3), 599–619.

Dur, R., Non, A., Prottung, P., & Ricci, B. (2025). Who's Afraid of Policy Experiments? *The Economic Journal, 135*(666), 538–555.

Edmondson, A., & Zuzul, T. (2016). Teaming routines in complex innovation projects. In R. Howard-Grenville, C. Rerup, A. Langley, & H. Tsoukas (Eds.), *Organizational routines: How they are created, maintained, and changed*. Oxford University Press.

Eisenhardt, K. M., & Martin, J. A. (2017). Dynamic capabilities: what are they? *The SMS Blackwell Handbook of Organizational Capabilities*, 341–363.

Eriksson, T. (2013). Methodological issues in dynamic capabilities research: a critical review. *Baltic Journal of Management*, *8*(3), 306–327.

Eshuis, J., & Gerrits, L. (2021). The limited transformational power of adaptive governance: a study of institutionalization and materialization of adaptive governance. *Public Management Review*, *23*(2), 276–296.

Ettelt, S., Mays, N., & Allen, P. (2015). Policy experiments: investigating effectiveness or confirming direction? *Evaluation*, *21*(3), 292–307.

Farmer, J., McCosker, A., Albury, K., & Aryani, A. (2022). Data capability through collaborative data action. In *Data for Social Good: Non-Profit Sector Data Projects* (pp. 63–88). Springer Nature Singapore.

Farrelly, R. (2008). Policy on trial: randomised trials are the best tool we have for finding out if policies really work. *Policy: A Journal of Public Policy and Ideas*, *24*(3), 7–12.

Farrelly, M., & Brown, R. (2011). Rethinking urban water management: experimentation as a way forward? *Global Environmental Change*, *21*(2), 721–732.

Feldman, M. S. (2016). Past, Present, and Future. *Organizational routines: How they are created, maintained, and changed*, 23–46.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*(1), 94–118.

Forman, J., & Damschroder, L. (2007). Qualitative content analysis. Rmt: *Empirical Methods for Bioethics: A Primer* (pp. 39–62). Emerald Group Publishing Limited.

Goyal, N., & Howlett, M. (2024). Types of learning and varieties of innovation: how does policy learning enable policy innovation? *Policy & Politics*, *52*(4), 564–585.

Greenberg, D. H., & Morris, S. (2005). Large-scale social experimentation in Britain: What can and cannot be learnt from the employment retention and advancement demonstration? *Evaluation*, *11*(2), 223–242.

Greenberg, D., & Poole, L. (2007). Designing a social experiment for the UK: how it was done and some lessons learned. *Evidence & Policy*, 3(1), 5–29.

Hansson, S. (2018). The discursive micro-politics of blame avoidance: Unpacking the language of government blame games. *Policy Sciences*, *51*(4), 545–564.

Hasselman, L. (2017). Adaptive management; adaptive co-management; adaptive governance: what's the difference? *Australasian Journal of Environmental Management*, *24*(1), 31–46.

Hatfield-Dodds, S., Nelson, R., & Cook, D. C. (2007). Adaptive governance: An introduction and implications for public policy.

Haynes, L., Goldacre, B., & Torgerson, D. (2012). Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. Cabinet Office – Behavioural Insights Team. Available at: http://dx.doi.org/10.2139/ssrn.2131581.

Heikkila, T., Gerlak, A. K., & Smith, B. (2024). Diagnosing individual barriers to collective learning: how governance contexts shape cognitive biases. *Journal of European Public Policy*, 31(7), 2026–2049.

Heilmann, S. (2008). Policy experimentation in China's economic rise. *Studies in Comparative International Development*, *43*(1), 1–26.

Hinterleitner, M. (2017). Reconciling perspectives on blame avoidance behaviour. *Political Studies Review, 15*(2), 243–254.

Hinterleitner, M., & Sager, F. (2015). Avoiding blame—a comprehensive framework and the Australian Home Insulation Program Fiasco. *Policy Studies Journal*, *43*(1), 139–161.

Hood, C. (2010). The Blame Game: Spin, Bureaucracy, and Self-Preservation In Government.

Hooren, F. V., Kaasch, A., & Starke, P. (2014). The shock routine: Economic crisis and the nature of social policy responses. *Journal of European Public Policy*, *21*(4), 605–623.

Hogan, J., & Feeney, S. (2012). Crisis and policy change: The role of the political entrepreneur. *Risk, Hazards & Crisis in Public Policy*, *3*(2), 1–24.

Howlett, M. (2014). Why are policy innovations rare and so often negative? Blame avoidance and problem denial in climate change policy-making. *Global Environmental Change*, *29*, 395–403.

Huitema, D., Jordan, A., Munaretto, S., & Hildén, M. (2018). Policy experimentation: core concepts, political dynamics, governance and impacts. *Policy Sciences*, 51, 143–159.

Hurlbert, M. A. (2018). *Adaptive Governance of Disaster.* Springer International Publishing.

Janssen, M., & Van der Voort, H. (2020). Agile and adaptive governance in crisis response: Lessons from the COVID-19 pandemic. *International Journal of Information Management*, *55*, 102180.

Ko, K., & Shin, K. (2017). How Asian countries understand policy experiment as policy pilots? *Asian Journal of Political Science*, *25*(3), 253–265.

Koontz, T. M., Gupta, D., Mudliar, P., & Ranjan, P. (2015). Adaptive institutions in social-ecological systems governance: A synthesis framework. *Environmental Science & Policy*, *53*, 139–151.

Lange, P., Driessen, P. P., Sauer, A., Bornemann, B., & Burger, P. (2013). Governing towards sustainability—conceptualizing modes of governance. *Journal of Environmental Policy & Planning*, *15*(3), 403–425.

Leino, H., & Åkerman, M. (2022). The politics of making Finland an experimenting nation. *Critical Policy Studies*, *16*(4), 441–459.

Lember, V., Kattel, R., & Tõnurist, P. (2018). Technological capacity in the public sector: The case of Estonia. *International Review of Administrative Sciences*, *84*(2), 214–230.

Lenz, A. (2024). Managing crises as if no one is watching? Governance dilemmas from a public perspective. *Public Administration*, *102*(4), 1397–1417.

Leong, C., & Howlett, M. (2017). On credit and blame: Disentangling the motivations of public policy decision-making behaviour. *Policy Sciences*, *50*(4), 599–618.

Leong, C., Howlett, M., & Safaei, M. (2023). Blame avoidance and credit-claiming dynamics in government policy communications: Evidence from Leadership Tweets in Four OECD Countries during the 2020–2022 COVID-19 Pandemic. *Policy and Society*, *42*(4), 564–585.

Margetts, H., & Naumann, A. (2017). *Government as a platform: What can Estonia show the world.* University of Oxford, *1*(1), 1–41.

Mattocks, K. (2021). Policy experimentation and policy learning in Canadian cultural policy. *Policy Sciences*, 54(4), 891–909.

Mayne, Q., De Jong, J., & Fernandez-Monge, F. (2020). State capabilities for problem-oriented governance. *Perspectives on Public Management and Governance*, *3*(1), 33–44.

McFadgen, B., & Huitema, D. (2017). Are all experiments created equal? A framework for analysis of the learning potential of policy experiments in environmental governance. *Journal of Environmental Planning and Management*, *60*(10), 1765–1784.

McLain, R. J., & Lee, R. G. (1996). Adaptive management: promises and pitfalls. *Environmental Management*, *20*(4), 437–448.

Meadowcroft, J. (2009). What about the politics? Sustainable development, transition management, and long-term energy transitions. *Policy Sciences*, *42*(4), 323–340.

Meinherz, F. X., & Fritz, L. (2024). The crisis justified the urgency, but now we have to go back to the rule of law: Urban mobility governance during Covid-19. *Environmental Politics*, *33*(3), 508–529.

Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J., & Olsen, D. (2017). Big data analytics capability: antecedents and business value.

Millo, Y., & Lezaun, J. (2006). Regulatory experiments: Genetically modified crops and financial derivatives on trial. *Science and Public Policy*, *33*(3), 179–190.

Mostert, E., Pahl-Wostl, C., Rees, Y., Searle, B., Tàbara, D., & Tippett, J. (2007). Social learning in European river-basin management: barriers and fostering mechanisms from 10 river basins. *Ecology and Society*, *12*(1).

Nair, S. (2020). Designing policy pilots under climate uncertainty: A conceptual framework for comparative analysis. *Journal of Comparative Policy Analysis: Research and Practice*, *22*(4), 344–359.

Nair, S., & Howlett, M. (2016). Meaning and power in the design and development of policy experiments. *Futures*, *76*, 67–74.

Nolte, I. M., & Lindenmeier, J. (2024). Creeping crises and public administration: a time for adaptive governance strategies and cross-sectoral collaboration? *Public Management Review*, *26*(11), 3104–3125.

Oakley, A. (1998). Public policy experimentation: lessons from America. *Policy Studies*, *19*(2), 93–114.

Pattyn, V., Matthys, J., & Hecke, S. V. (2021). High-stakes crisis management in the Low Countries: Comparing government responses to COVID-19. *International Review of Administrative Sciences*, *87*(3), 593–611.

Pearce, W., & Raman, S. (2014). The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy Sciences*, *47*(4), 387–402.

Philipsen, S., Stamhuis, E. F., & De Jong, M. (2021). Legal enclaves as a test environment for innovative products: Toward legally resilient experimentation policies. *Regulation & Governance*, *15*(4), 1128–1143.

Piening, E. P. (2013). Dynamic capabilities in public organizations: A literature review and research agenda. *Public Management Review*, *15*(2), 209–245.

Razak, A. R. A. (2017). *An investigation into the micro-dynamics of routine flexibility*. Lancaster University.

Rijke, J., Brown, R., Zevenbergen, C., Ashley, R., Farrelly, M., Morison, P., & van Herk, S. (2012). Fit-for-purpose governance: A framework to make adaptive governance operational. *Environmental Science & Policy*, *22*, 73–84.

Rivlin, A. M. (1973). Social experiments: the promise and the problems. *The Brookings Bulletin*, *10*(3), 6–9.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, *80*(1), 1–22.

Savaget, P., Roulet, T., & Ventresca, M. (2024). Experimentation in wicked situations: How activists construct pragmatic action frames. *Organization Studies*, *45*(10), 1493–1521.

Scarlett, L. (2013). Collaborative adaptive management: challenges and opportunities. *Ecology and Society*, *18*(3).

Sharma-Wallace, L., Velarde, S. J., & Wreford, A. (2018). Adaptive governance good practice: Show me the evidence! *Journal of Environmental Management*, *222*, 174–184.

Sharma, G. D., Kraus, S., Srivastava, M., Chopra, R., & Kallmuenzer, A. (2022). The changing role of innovation for crisis management in times of COVID-19: An integrative literature review. *Journal of Innovation & Knowledge*, 7(4), 100281.

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), 433–441.

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, *36*(2), 381–403.

Stein, H., Cunningham, S., & Carmody, P. (2021). The rise of "behavioral man": Randomized controlled trials and the "new" development agenda. *Human Geography*, *14*(1), 62–75.

Stoker, G. (2010). Exploring the promise of experimentation in political science: Micro-foundational insights and policy relevance. *Political Studies*, *58*(2), 300–319.

Stoker, G., & John, P. (2009). Design experiments: Engaging policy makers in the search for evidence about what works. *Political Studies*, *57*(2), 356–373.

Straßheim, H. (2024). A world of evidence: the global spread and silent politics of evidence cultures. *Policy and Society*, *43*(4), 414–431.

Swanson, D., & Bhadwal, S. (Eds.). (2009). *Creating adaptive policies: A guide for policymaking in an uncertain world*. IDRC.

Tammpuu, P., & Masso, A. (2018). 'Welcome to the virtual state': Estonian e-residency and the digitalised state as a commodity. *European Journal of Cultural Studies*, *21*(5), 543–560.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, *28*(13), 1319–1350.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.

icl u, T., Hin ea, C., Andrianu, B. (2020). Adaptive and turbulent governance. a s of governing that foster resilience. The case of the covid-19 pandemic. *Transylvanian Review of Administrative Sciences*, *16*(SI), 167–182.

Van Assche, K., Beunen, R., Verweij, S., Evans, J., & Gruezmacher, M. (2022). Policy Learning and Adaptation in governance; a Co-evolutionary Perspective. *Administration & Society*, *54*(7), 1226–1254.

Van Den Ende, M. A., Hegger, D. L., Mees, H. L., & Driessen, P. P. (2024). The transformative potential of experimentation as an environmental governance approach: The case of the Dutch peatlands. *Environmental Policy and Governance*, *34*(5), 524–537.

Van der Heijden, J. (2018). From Leaders to Majority: A Frontrunner Paradox in Built-Environment Climate Governance Experimentation. *Journal of Environmental Planning and Management*, *61*(8), 1383–1401.

Wilhelm, H., Maurer, I., & Ebers, M. (2022). (When) are dynamic capabilities routine? A mixed-methods configurational analysis. *Journal of Management Studies*, *59*(6), 1531–1562.

Willi, Y., Nischik, G., Braunschweiger, D., & Pütz, M. (2020). Responding to the COVID-19 crisis: transformative governance in Switzerland. *Tijdschrift voor economische en sociale geografie*, *111*(3), 302–317.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, *24*(10), 991–995.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). Sage.

Zaki, B.L., Pattyn, V. & Wayenberg, E., (2023). Policy learning type shifts during creeping crises: A storyboard of COVID-19 driven learning in Belgium. *European Policy Analysis*, *9*(2), pp. 142–166.

Zhao, G., Hui, X., Lu, Y., & Zhang, Y. (2024). Progress in adaptive governance research and hotspot analysis: a global scientometric visualization analysis. *Discover Sustainability*, *5*(1), 234.

## Acknowledgements

It was 2020 when I started my PhD journey, full of passion and an urge to make sense of the academic world of public administration. However, I was not prepared for the depth of knowledge that I was exposed to, especially through my supervisors, Ringa Raudla and Veiko Lember. Throughout the entire journey, they served as mediators between me and the knowledge "hidden" within the walls of our university that I was about to explore.

A special thank you to Ringa—you are a true professional, incredibly swift, and have an admirable ability to delve into different topics. You inspired me greatly along the way. It was an honour and a privilege to be part of our research project on public policy experimentation. Although it was not part of my initial plan, I can confidently say that this opportunity and a chance to work together with such bright colleagues (Ringa Raudla, Külli Sarapuu, Egert Juuse, Nastassia Harbuzova, Johanna Vallistu, and Aleksandrs Cepilovs) led me here today with much less solitude and stress than I had prepared myself for. I enjoyed our mutual efforts, discussions, and discoveries!

In the last few years, another important event happened—I became a mother. While it has been a blessing, the new life brought challenges of its own and made me realise even more how kind, considerate, and patient our team was. The support of my supervisors and our research team was the reason this journey did not stop, even in the hardest times, and small steps have led to this moment now. And of course—my sweet little Keron—what teamwork this journey has been! As time passes, you probably won't remember much, but when you noticed that mommy was aloof at times, this thesis might have been the reason. I will put this in writing so you can read this THANK YOU in capital letters when you grow up and learn to speak English.

Also, my family has a wonderful ability to push me just the right amount, while providing me with a safe space to feel low and tired, which has been crucial in helping me pick myself up and keep going. Thank you, my partner Martin, and my father Raiki. But most importantly, thank you, Tiiu, for being an incredible mother, woman and a role model. I would have loved for you to have had had the chance to share this joy and sense of achievement.

The research presented in this doctoral thesis was funded by Estonian Research Agency (PRG1125).

# Abstract

## Governing in Turbulence: Leveraging Experimentation and Data Capabilities for Building Adaptive Governance

Governments around the world operate in turbulent conditions, making unpredictability and uncertainty, but also time shortages and conflicting values (Deverell, 2010, p. 34; Ansell & Trondal, 2018) a part of their everyday working environment. Under such conditions, traditional governance models often prove too rigid. As a result, adaptive governance has emerged as a salient approach (Dietz et al., 2003; Chaffin et al., 2014; Nolte & Lindenmeier, 2024; Sharma-Wallace et al., 2018; Brunner, 2010). Adaptive governance refers to governance systems that self-organise as a result of learning and interaction and are supported by institutional arrangements that enable continuous learning, collaboration, and flexibility (Rijke et al., 2012). As Folke et al. (2005, p. 444) explain, in a system with high adaptability, "the actors have the capacity to reorganize the system within desired states in response to changing conditions and disturbance events," which indicates that adaptive governance emerges from the adaptive capacity of public sector organisations. However, despite adaptive governance being widely promoted as a means of managing complexity and uncertainty, it is often argued that it remains a normative and descriptive concept that lacks sufficient operationalisation (Zhao et al., 2024; Janssen & Van der Voort, 2020; Rijke et al., 2012).

Drawing on institutional theories, organisation theories, public administration and public policy literature, this thesis aims to fill this gap by operationalising the concept through the exploration of capabilities that underlie adaptive governance. The literature highlights a broad set of capabilities, e.g., institutional flexibility, multi-level coordination, and stakeholder collaboration but especially continuous learning (Rijke et al., 2012; Armitage et al., 2008) as a crucial enabler of adaptive governance. This thesis focuses on experimentation and data capabilities that help to operationalise learning under uncertainty but also provide justification and direction for policy change—without experimentation and data capabilities, responding to changes may remain reactive or *ad hoc*. The thesis argues that experimentation and data capabilities serve as means of adjusting organisational routines—defined as "repetitive, recognizable patterns of interdependent actions" (Feldman & Pentland, 2003, p. 95)—to respond to changes in the surrounding environment and improve the adaptive capacity of public sector organisations. However, to support adaptive governance, these capabilities should become routinised and institutionalised ways of making decisions and developing policies.

Furthermore, experimentation and data capabilities may encounter significant barriers that hinder the development of adaptive governance (Eshuis & Gerrits, 2021). To understand the conditions and antecedents needed for adaptive governance to emerge, this study explores the factors that influence experimentation and the development of data capabilities. The thesis follows an inductive analytical strategy and, drawing on empirical studies, outlines the key factors influencing the use of experimentation and data capabilities in three dimensions of governance—polity, policy, and politics (Lange et al., 2013). To reach the aims of the study, this thesis draws on five articles. Of these five articles, three are comparative (**Articles III**, **IV**, **V**), one is a theoretical analysis of existing literature on experimental policymaking (**Article I**), and one is a single case study (**Article II**). All five articles employ qualitative research methods.

This thesis builds on various data collection methods used across the articles, including literature reviews (**Articles I–V**), interviews (**Articles II**, **III**, **IV**, **V**), and document analysis (**Articles II**, **III**, **IV**, **V**). Therefore, qualitative interview data was included in all of the articles.

It is argued that while data capabilities and experimentation serve as sources of dynamism, as they support the adaptive capacity of public sector organisations, different factors influence their potential to lead to continuous and long-term capability building and routine adaptations that are required for adaptive governance. Also, it has been argued that while adaptive governance may be easily institutionalised at a cognitive level, the real transformational and enduring change will not follow when the "underlying ideas regarding governance remain unchanged" or when adaptive governance is pursued solely as a pilot (Eshuis & Gerrits, 2021, p. 292). As illustrated by the comparative studies of Finnish and Estonian governance institutions, the strategic and politically supported experimentation framework ("Experimental Finland" under Sipilä's Government) has significantly helped to legitimise the approach and to lower the perceived risks associated with experimentation.

Additionally, the implications of crises for experimentation and data capabilities are explored to complement this discussion. While crises may provide urgency and "windows of opportunity" (Ansell & Trondal, 2018) to induce changes in organisational routines, less is known about how crisis conditions influence experimentation and the development of data capabilities. The thesis outlines four paradoxes—"contradictory yet interrelated elements that exist simultaneously" (Smith & Lewis 2011, p. 386)—that stem from crisis conditions and the logics of experiments. This indicates that crises call for flexible experiments that operate on a short timescale. Also, the results indicate that while high-level data capabilities support public organisations in responding to a crisis, the temporality of crisis conditions may hinder a wider use of the results of experiments or long-term routine changes.

# Lühikokkuvõte

## Valitsemine muutlikes oludes: eksperimenteerimise ja andmevõimekuste kasutamine kohaneva valitsemise toetamiseks

Avaliku sektori organisatsioonid tegutsevad üha turbulentsemates tingimustes, mille tulemusena on ebakindlus, ettearvamatus ja määramatus osa nende igapäevasest toimimisest. Sellised tingimused, mida võib pidada uueks normaalsuseks (Ansell & Trondal, 2018), nõuavad võimekusi, mis aitavad teha kiireid ja paindlikke otsuseid olukorras, kus informatsiooni ei ole kunagi piisavalt (Hasselman, 2017). Lisaks tuleb valmistuda ja silmitsi seista kriisidega, mis kujutavad endast ootamatut ja lühiajalist, kuid tõsist ohtu süsteemi põhistruktuuridele – väärtustele ja normidele, mis toob kaasa ajalise surve ja ebakindluse ning eeldab kiirete otsuste tegemist (Rosenthal jt 1989, lk 10).

Käesolev doktoritöö uuris kohanevat valitsemist (Dietz jt, 2003; Chaffin jt, 2014; Nolte & Lindenmeier, 2024; Sharma-Wallace jt, 2018; Brunner, 2010) kui lähenemisviisi, mis toetab avaliku sektori võimet kohaneda muutustega pideva õppimise, koostöö ja paindlike institutsioonide abil (Rijke jt, 2012; Armitage jt, 2008). Kuigi tegemist on laialdaselt kasutusel oleva kontseptsiooniga, peetakse seda sageli liialt normatiivseks ja kirjeldavaks (Zhao jt, 2024; Janssen & Van der Voort, 2020; Rijke jt, 2012). Käesolev doktoritöö adresseeris seda lünka, uurides kohaneva valitsemise aluseks olevaid võimekusi ja keskendudes eelkõige eksperimenteerimisele ja andmevõimekustele. Kasutades dünaamiliste võimekuste raamistikku, selgitati doktoritöös eksperimenteerimise ja andmevõimekuste loogikat ja viise, kuidas nad võimaldavad muutusi organisatsiooni olemasolevates rutiinides – "korduvates, äratuntavates ja omavahel seotud tegevuste mustrites" (Feldman & Pentland, 2003, lk 95).

Kuigi andmevõimekused ja eksperimenteerimine võimaldavad avaliku sektori organisatsioonidel olla dünaamilisemad ja toetavad nende kohanemisvõimet, mõjutavad nende rakendamist ja arengut mitmed faktorid, mis tuletati induktiivselt intervjuudest Soome ja Eesti avalike teenistujatega. Selgus, et kui Soome varasem kogemus eksperimenteerimisega on võimaldanud seda lähenemisviisi poliitikakujundamisel legitimeerida, siis Eesti avalikud teenistujad tajuvad eksperimenteerimist pigem riskantse lähenemisena. Lisaks uuriti käesolevas doktoritöös kriiside mõju nendele võimekustele – kuigi kriisid võivad pakkuda "võimaluste aknaid" uute lahenduste elluviimiseks (Ansell & Trondal, 2018), siis kriiside ajutine iseloom seab piirangud pikaajaliste muutuste läbiviimiseks organisatsiooni rutiinides.

# Appendix

**Article I**

Raudla, R., Sarapuu, K., Juuse, E., Harbuzova, N., **Onno, K.**, Vallistu, J., & Cepilovs, A. (2023). To experiment or not to experiment in tax policy? Halduskultuur–The Estonian Journal of Administrative Culture and Digital Governance, 23(1), 27–48.

Ringa Raudla, Külli Sarapuu, Egert Juuse, Nastassia Harbuzova, Kerli Onno, Johanna Vallistu, Aleksandrs Cepilovs. 2023 "To Experiment or not to Experiment in Tax Policy?"
*Halduskultuur. The Estonian Journal of Administrative Culture and Digital Governance* 23(1), 27-48

# To Experiment or not to Experiment in Tax Policy?

RINGA RAUDLA, KÜLLI SARAPUU, EGERT JUUSE, NASTASSIA HARBUZOVA, KERLI ONNO, JOHANNA VALLISTU, AND ALEKSANDRS CEPILOVS
RAGNAR NURKSE DEPARTMENT OF INNOVATION AND GOVERNANCE, TALLINN UNIVERSITY OF TECHNOLOGY, ESTONIA

## Abstract

Despite the increasing use of experiments in policy-making the suitability of field experiments in the public sector context is still under debate. In this article, we focus on experimenting in the field of tax policy and ask: what are the promises and pitfalls of using experimental approaches in tax policy? While the existing discussions on tax policy experimentation focus on randomized controlled trials from a legal perspective, we adopt a broader view and provide a more comprehensive discussion by synthesizing insights from the fields of political science, public policy, public administration, and governance. Our analysis encompasses randomized controlled trials, non-randomized policy pilots and design experiments. We summarize the existing knowledge on using field experiments in policy-making and discuss the implications of the knowledge for experimenting in tax policy. We seek to offer a more holistic and critical take on whether we should promote the use of experimental approaches in this domain.

**Keywords:** policy experiments, tax policy, randomized controlled trials, policy pilots, design experiments

## 1. Introduction

Converging trajectories in various disciplines point to the increasing importance of using experimental approaches in policy-making (e.g. Ansell and Bartenberger 2016; John 2014; Ettelt et al. 2015a; Lee and Ma 2020; McFadgen and Huitema 2017; McGann et al. 2018). This has been aided by parallel developments in academia and practice. First, the term "experimentalist governance" has been used to describe developments in the European Union's approach to policy-making in various policy sectors (Rangoni and Zeitlin 2021). Second, the rise of the so-called "randomistas", who advocate the use of randomized controlled trials (RCTs) in development (as evidenced by Banerjee, Duflo and Kremer getting the Nobel prize in economics in 2019), has strongly influenced debates on development economics, with spillovers to other areas as well (e.g. Kvangraven 2020; Leao and Eyal 2019; Stein et al. 2021). Third, experimentalism has been embraced by behavioral economics and behavioral public policy (e.g. Jones and Whitehead 2018; Strassheim 2020). Fourth, the calls to integrate design thinking – which is also experimentalist in its logic – into policy-making have recently been made by a number of public policy scholars (e.g. Stoker and John 2009; McGann et al. 2018). In response to the pressures to be "smarter" and more "innovative", many governments have created nudge-type government units and public sector innovation labs that employ various

kinds of experimental methods (e.g. John 2014; McGann et al. 2018; Tõnurist et al. 2017).

At the same time, the world is becoming increasingly uncertain and complex, warranting experimentation as a way to cope (Rangoni and Zeitlin 2021; Voß and Simons 2018). For example, the need to deal with climate change has precipitated an avalanche of sustainability experiments (e.g. Ansell and Bartenberger 2016; McFadgen and Huitema 2017, 2018 McFadgen 2019). Fast technological developments (especially in ICT) and their pervasive impacts on economic and social spheres have triggered extensive discussions on experimental legislation and regulation (e.g. Philipsen et al. 2021; Ranchordas 2013).

The suitability of experimenting in the public sector context is still under debate, especially with regard to specific policy fields. In this article, we focus on experimenting in the field of tax policy. More specifically, our research question is: What are the promises and pitfalls of using experimental approaches in tax policy? The focus has been inspired by various considerations. Tax policy entails the use of taxes for a number of purposes: generating revenue for the government, shaping the behavior of individuals, redistributing income, and stabilizing the economic cycle (e.g. Kay 1990). Given that *consistency* and *horizontal equity*[1] are considered to be the core values in this policy area, using experimental approaches – which by nature entail disruption and potentially differential treatment of citizens – is likely to be more controversial in tax policy than in many other policy fields. At the same time, it is a policy sector where the complex and rapidly evolving context (e.g. changing technology, globalization, challenges presented by climate change) gives rise to considerable new uncertainties. Experiments could be viewed as generating useful knowledge about the novel challenges and the best ways to address them (Rangoni and Zeitlin 2021). Given the importance of taxes in generating revenues for the state, getting tax reforms "right" has very high stakes (Werner and Riedl 2019).

Recently, rather bold proposals for making more extensive use of field experiments in tax policy have been put forward by Abramowicz (2019). While Abramowicz (2019) approached the issue from the legal perspective and focused specifically on the promises of *randomized* experiments, we adopt a broader view. We provide a comprehensive discussion about the suitability of policy experiments in the field of tax policy by synthesizing insights from the fields of political science, public policy, public administration, and governance. With regard to the different *types* of experiments, our analysis encompasses both randomized controlled trials and non-randomized policy pilots as well as design experiments. We summarize the existing knowledge on using experimental approaches and discuss the implications of the knowledge for experimenting in the field of tax policy. With that, we hope to offer a more holistic (and also a more critical) take on whether we should promote the use of experimental approaches in this domain.

It is worth emphasizing that in this article we focus on tax *policy* rather than tax *administration*. The use of field experiments to improve tax compliance (e.g. by using deterrence messages, referring to the tax behavior of others, increasing the moral costs of non-compliance in tax offices' communication with taxpayers) has been extensively discussed in the existing studies (for useful overviews, see Hallsworth 2014; Mascagni 2018; Pomeranz and Vila-Belda 2019). Furthermore, our focus is not on whether *research* in public administration, public policy and political science should make (more) use of experimental methods (as discussed e.g. in Druckman et al. 2006; Stoker 2010). Instead, we are interested in whether *policy-makers*, in

1 Consistency refers to taxing equivalent transactions in a similar way. Horizontal equity means that taxpayers who are in a similar position should face a similar tax burden (e.g. Elkins 2006; Musgrave 1990).

devising and improving policy measures in the tax policy domain, should make more extensive use of experimental approaches, as the current *Zeitgeist* seems to be prescribing. Also, our focus is on field experiments and not on laboratory experiments.[2]

The article is structured as follows. Section 2 explains what we mean by the term "policy experiment". Sections 3 and 4 give an overview of the potential benefits and pitfalls of making more extensive use of experimental approaches in tax policy, respectively. While sections 3 and 4 focus on issues that are relevant for all experimental approaches (irrespective of the specific design choices), section 5 analyzes the specific promises and challenges presented by different types of designs. Section 6 concludes.

## 2. Definition of policy experiment

Policy experiment is a protean concept (Karvonen and van Heur 2014), and in the existing literature, various definitions have been offered (for overviews of these debates, see, e.g., Ansell and Bartenberger 2016; Bauknecht et al. 2020). In some disciplines (e.g. economics), the term "experiment" typically has a rather narrow meaning and refers to a randomized controlled trial (RCT) (Ansell and Bartenberger 2016; Burtless 1995; Werner and Riedl 2019). On the other extreme, any policy reform – like the reduction in the top marginal rates of the income tax in the US in the 1980s – has been labelled as an "experiment" (Burtless 1995; Druckman et al. 2006; Heldeweg 2015; Nair and Howlett 2016).

In this article, we opt for the middle ground. The definition of policy experiment we employ is as follows: it is a policy relevant test undertaken by government organization(s) to learn about the impacts of a new policy solution, which can be used as evidence for further policy decisions (Bravo-Biosca 2020, 195; Heldeweg 2015, 183; McFadgen and Huitema 2017, 1768; Nair and Howlett 2016, 69; Millo and Lezaun 2006, 179). In an experiment, a new policy solution is "tried out or tested in a restricted environment in terms of time, space, scope and/or actors", but it is "intended to provide a proof of principle that subsequently could have the potential of wider societal relevance" (Heldeweg 2015, 183). Experiments are "temporary and reversible interventions without permanent policy consequences" (McFadgen and Huitema 2017, 1767).

We regard a policy experiment as "a process that generates learning through an explicit intention to test new ideas" (McFadgen and Huitema 2017, 1765). As Bravo-Biosca (2020, 195) explains, an experiment is "intentionally set up to learn", has "a clearly structured learning strategy", and generates "new information, evidence, or data". Therefore, if a government just tries out something new, it does not amount to a policy experiment unless the systems and processes required to learn from it are also established. This includes a timeframe for checking results and deciding whether to continue the experiment, adjust it, discontinue or scale up (Bravo-Biosca 2020). In other words, we follow the understanding whereby policy experimentation does not entail "freewheeling trial and error or spontaneous policy diffusion" but is "purposeful and coordinated activity geared to producing novel policy options that are injected into official policymaking", with the goal to scale them up, if successful (Heilmann 2008, 3).

2 Useful overviews of laboratory experiments that could be insightful for tax policy have been provided by Werner and Riedl (2019) and Alm and Malézieux (2021).

## 3. The promises and benefits of using experimental approaches

The advocates of the experimental turn posit that experimental approaches can improve the substantive quality of policies as well as the policy-making process. First, from the perspective of *policy content*, the key aspects that speak for experimentation are the complexity of the system policy-makers seek to influence and uncertainty about the impacts of an intervention (Ansell and Bartenberger 2016; Bauknecht et al. 2020; Bravo-Biosca 2020; Hughes et al. 2020; Lee and Ma 2020; Lee et al. 2009; Millo and Lezaun 2006; Nair and Howlett 2016; Voß and Simons 2018). Experimental approaches are considered especially useful for understanding complex systems where actors, institutions and policy continuously evolve and interact in various ways (Bravo-Biosca 2020) or are influenced by novel and disruptive technological developments (Bauknecht et al. 2020; Philipsen et al. 2021; Ranchordas 2013, 2015a, 2015b; Van Gestel and Van Dijck 2011).

Therefore, the key promise of policy experiments, regardless of the specific design, is to provide policy-makers with information that would otherwise not be available (Abramowicz 2008; Ettelt et al. 2015b; Millo and Lezaun 2006; Philipsen et al. 2021). Experiments allow policy-makers to test the effectiveness of a policy on a smaller scale, but in a real world setting, before it is rolled out on a larger scale (Bailey et al. 2017; Checkland et al. 2021; Farrelly 2008; Hughes et al. 2020; Lee et al. 2009; Philipsen et al. 2021; Ranchordas 2013). In addition to allowing the exploration of whether intended impacts materialize, experiments can offer information about unintended and adverse effects (Haynes et al. 2012; Hughes et al. 2020; Lee and Ma 2020; Nair and Howlett 2016; Philipsen et al. 2021; Ranchordas 2013), reduce informational asymmetries with regard to policy's acceptability for stakeholders (Nair and Howlett 2016 Philipsen et al. 2021), and alleviate uncertainty about how the target group behaves in response to different measures (Burtless 1995; Nair and Howlett 2016; Ranchordas 2013). Similarly, experiments can offer information regarding the impacts of different policy design options and help policy-makers choose between them (Bravo-Biosca 2020; Haynes et al. 2012; Hughes et al. 2020; Lee and Ma 2020; Nair and Howlett 2016). In the context of disruptive technological developments, experiments help to diminish the pacing and information gaps between technological innovations and legislation (Ranchordas 2013, 2015c).

These promises of experimenting reflect the information needs also in the field of tax policy. For example, if policy-makers were considering the use of tax credits for promoting the adoption of more environmentally sustainable investments and business practices by companies, an experiment could offer information about the degree of take-up of such incentives. Since taxpayers' perceptions of tax incentives can be distorted (Werner and Riedl 2019), it can be difficult to make linear predictions about their effects. Experiments can offer useful insights into how such incentives are actually perceived. With regard to choosing the best tax policy design, an experiment may allow policy-makers to compare the effects of tax credits, tax deductions and tax exemptions, and opt for the solution that is likely to have the largest impact on the adoption of environmentally sustainable practices. In assessing the effects of tax credits on the behavior of businesses, policy experimenters can examine whether the size of the tax credit makes a difference. Experiments could also be used for assessing which aspects of the policy are having the largest effect (Haynes et al. 2012). For example, if the government employs, in parallel, tax credits but also subsidies for promoting environmentally

sustainable investments, it may be interested in which instrument is responsible for driving the policy outcomes. In an era of disruptive technological developments, experiments can be useful for testing how to best tax digital nomads, digital transactions, digital data, carbon, or ecological footprints (Akdogan 2021; Shome 2021). In such instances, experiments also allow governments to test their own ability to impose taxes on novel phenomena (Millo and Lezaun 2006).

Second, from the perspective of *policy process*, experimentation can improve the quality of deliberations over policy. In uncertain situations, policy experimentation can help to build consensus and create space for political bargains by focusing policy-makers' attention on policy consequences and outcomes (Abramowicz 2008; McFadgen 2019; Millo and Lezaun 2006; Nair and Howlett 2016; Lee and Ma 2020; Ranchordas 2013, 2015a, 2015c). This can be particularly beneficial in tax policy, where the partisan debates tend to be polarized and more influenced by policy hunches (or even dogmas) and less by information about the actual impacts of different solutions (Abramowicz 2019). By presenting a possible future, experimenters can mitigate conflict in policy reforms that may otherwise be politically unpalatable (Bailey et al. 2017; Nair and Howlett 2016; Ranchordas 2013, 2015a, 2015c). Consensus on a policy, in turn, can potentially facilitate higher levels of compliance and, through that, increase policy effectiveness.

Furthermore, experimentation can be viewed as helping to lower the overall costs of policy (Abramowicz et al. 2010; Adkins and Ylöstalo 2018; Checkland et al. 2021; Farrelly 2008; Haynes et al. 2012; Hughes et al. 2020; Kvangraven 2020; Lee et al. 2009). Even though the running of experiments can entail data collection costs, such investments can pay off via allowing policy-makers to "weed out" programs that are not effective or avoid the costs of failure (Abramowicz et al. 2010; Bravo-Biosca 2020; Farrelly 2008; Haynes et al. 2012; Oakley 1998; Lee and Ma 2020; Lee et al. 2009). Experiments can also help to decide, in the context of limited resources, which policy options from a range of alternatives deliver the highest value and should be chosen over the others (Haynes et al. 2012). Given that changes in tax policy can entail considerable costs for the state budget (especially when they entail tax incentives), the financial considerations in opting for experiments can be particularly pertinent in the field.

## 4. The challenges and constraints of using experimental approaches

The more critical perspectives on using experiments in public policy point out several challenges and constraints. These derive from two sources – the problems inherent to experimental approaches and the political setting of public policies. Tax policy experiments are likely to be challenged by both.

There are challenges that are common to all types of policy experimentation, regardless of their specific design – whether they "slice" through space, time, scope or types of participants. For instance, an important challenge for all experiments is that due to their small scale, scope or limited duration they would not be able to capture some of the effects which could materialize when the same policy is offered at a large scale or over a longer period of time or in a different context (Bauknecht et al. 2020; Burtless 1995; Millo and Lezaun 2006; Werner and Riedl 2019). As Millo and Lezaun (2006, 181) put it, one can always point to "particular features of the world that the experiment failed to replicate". For example, in an experiment

that tests the impacts of tax credits for investments to lower the carbon footprint, if only a small fraction of businesses is enrolled, it may not be able to capture the kinds of peer-effects and isomorphic pressures that may emerge if the scheme applied to all businesses. Thus, in tax policy experiments, one should remain aware of the dangers of making "sweeping macro-level generalizations" by drawing on micro-level evidence (Stein et al. 2021, 64). Attribution of causality can be particularly challenging when "complex systems are involved in a difficult task" (Nair and Howlett 2016, 71). Thus, tax policy experiments would face challenges in assessing third-party effects (i.e. effects that go beyond the taxpayers themselves) and especially macroeconomic consequences of tax changes (Abramowicz 2019).

Furthermore, a key challenge for all experiments, including those in tax policy, is the Hawthorne effect, which occurs when individuals or businesses behave differently because they know that they are in an experiment (e.g. Levitt and List 2011). Also, if the participants know that the treatment is of limited duration, their reactions may be different from how they would react if the same policy was of enduring character (Burtless 1995). For example, taxpayers subject to an experiment may try to lower their tax payments by shifting their income and deductions from years in which they are subject to the experiment to years when the generally applicable tax rules apply to them or the other way around (Abramowicz 2019; Abramowicz et al. 2010). In such cases, a possible solution could be to have a longer time period for the experiment (spanning several years) (Abramowicz et al. 2010), which would limit the ability of individuals or companies to shift their income or expenses. However, the longer the experiment, the more serious the problem of attrition (Burtless 1995), which challenges the assessment of impacts. In addition to the Hawthorne effect, the results of an experiment may be influenced by spillovers to areas or individuals who were not supposed to be affected by the experiment. For example, if tax incentives for individuals or companies are offered in one jurisdiction (e.g. local government) and not in others, taxpayers may relocate to the experimental jurisdiction, which makes it difficult to compare the experimental jurisdiction with the non-experimental ones (Abramowicz 2008).

Next to the constraints related to the general nature of experiments, the *political setting* of public policies in a democratic context poses an additional range of challenges. The prevailing logics of political decision-making, policy and electoral cycles, electoral considerations, and value conflicts can render experimental approaches difficult for policy-makers to undertake (Bauknecht et al. 2020; Bravo-Biosca 2020; Burtless 1995; Voß and Simons 2018). The political setting entails four types of causes that contradict experimental logic and, consequently, may inhibit undertaking the experiments or shape their implementation.

The first challenge is related to the nature of politics per se. Politicians may reject policy experimentation as a general strategy if they regard it as a route for depoliticizing or de-democratizing policy decisions (Pearce and Raman 2014; Voß and Simons 2018; Strassheim 2020). Extensive literature about the obstacles to evidence-based policy-making has suggested that politicians often follow other logics than systematic evidence in adopting policy decisions. In a democratic system, competing societal values and preferences are translated into policy solutions (Kvangraven 2020). Thus, politicians may attribute higher importance to voter preferences, party agendas, and ideological considerations rather than to the kind of evidence that experiments promise to deliver. This is especially relevant for tax policy. Tax issues often constitute salient agenda points in electoral competition (Klitgaard

et al. 2015) and can define the core identity of some parties (Carmines et al. 2012; Ballard-Rosa et al. 2017; Osterloh and Debus 2012). As a result, tax policy issues may be regarded very ideologically and viewed as the core space in which ideological trade-offs (e.g. more state vs less state, equity vs freedom etc.) are struck (Carmines et al. 2012). This makes it particularly challenging to use experimental approaches in this domain. Driven by ideology, politicians may trust their gut feeling about the potential impacts of different tax policies – e.g. higher rates are "good" or "bad" and exemptions are "good" or "bad" depending on the ideological spectrum they are on – and hence do not feel the need to demand more rigorous evidence via experiments (Bravo-Biosca 2020). Politicians are likely to object to any tax experiments that they perceive to hurt their re-election chances (Bauknecht et al. 2020; Nair and Howlett 2016). Conversely, a tax policy option may be "so in tune with prevailing political values that subjecting it to tests is regarded by politicians as unnecessary or even unwelcome" (Pearce and Raman 2014, 393).

The second issue pertains to the need to avoid uncertainty. Beyond any ideological considerations, politicians may be concerned about how experimental decision-making as a policy-making style looks to the electorate. Experimentation may be problematic for politicians, since undertaking an experiment entails admitting uncertainty and lack of existing knowledge about the effects of policy interventions (Bauknecht et al. 2020; Farrelly 2008). Thus, if politicians were to opt for experimental tax policy at all, they may prefer to treat experiments as "demonstrations of effectiveness" or "justifications for decisions already taken" rather than answering open questions (Ettelt et al. 2015a, 294, 302; see also Zurbriggen and Lago 2019). These concerns may be reinforced by the culture of public sector organizations, which tend to be risk-averse and prefer stability, predictability and order, and exhibit limited tolerance of failure (Lewis et al. 2020; McGann et al. 2018; Zurbriggen and Lago 2019). The need to avoid uncertainty is likely to be particularly pronounced in case of a core state function like taxation. Instead of admitting uncertainty, elected representatives may prefer to be perceived as "decisive, energetic, and positive" (Farrelly 2008, 11) and emphasize "profound knowledge" about future developments (Bauknecht et al. 2020, 57). Experimenting strikes the kind of "chord of skepticism and indecision" that politicians seek to avoid (Peters 1998, 126) – even if the policy change in question entails using tax policy instruments in new technological spheres (e.g. taxing digital transactions) or changing societal context (e.g. digital nomads or platform work), all facing considerable uncertainties. Although the higher the uncertainty, the more learning opportunities experiments can offer, politicians may regard an open outcome as entailing a higher likelihood of failure – and this may be something they are motivated to avoid (Ettelt et al. 2015b; McFadgen and Huitema 2018).

The third issue concerns constrained timeframes. Some policy decisions may require immediate decisions, and hence policy-makers may not have the time needed for experimenting (Clarke and Craft 2019). For example, when tax policy is used for macroeconomic stabilization (i.e. cutting rates during recession and increasing them during a boom), timing is of pivotal importance and swift decisions crucial if undesirable lagged effects on the economic cycle are to be avoided. Furthermore, politicians may perceive that they are likely to be electorally rewarded for being swift rather than slow. As Stoker (2010, 53) explains, the demands of the experiment may clash with the policy cycles or political dynamics: "Experiments are a tool with a linear rhythm in a non-linear policy process and may as a result lose the battle for relevance by failing to produce results in a timely way." There might be political pressure

to go fully ahead with the policy owing to its perceived benefits rather than waiting for the results of the experiment (Stoker 2010) or hasten the evaluation of the pilot, especially if the policy deals with pressing social problems (Nair and Howlett 2016). Especially in the case of longer-lasting experiments spanning several years (like the social experiments in the US in the 1960s and 1970s), the political agendas may change, rendering the findings less relevant for policymakers (Burtless 1995; Oakley 1998).

Fourth, there may be lacking capacities. Experimental policy-making requires various types of analytical and collaborative capabilities from the policy-makers, which may be lacking (Bedard and Ouimet 2012; Bravo-Biosca 2020; Stoker 2010). Lacking collaborative capabilities can become a crucial hurdle to policy experiments that involve a large number of different organizations and the engagement of public officials at several levels of government (Stoker 2010, 51; Cotterill and Richardson 2010, 157). While some simpler tax policy experiments may be relatively straightforward and just involve the finance ministry and tax office, others (e.g. on using tax incentives for fostering innovation or environmental sustainability) can involve a considerably larger number of bodies. Furthermore, objections to experiments may emerge from the broader public, and considerable communicative efforts may be needed to explain the social benefits of an experiment to gain public acceptance (Bauknecht et al. 2020) – especially in domains like tax policy where the public is likely to have deep-rooted expectations of horizontal equity.

## 5. Promises and pitfalls of specific experimental designs

Policy experiments can take different forms. In this section, we will focus on the following designs: randomized controlled trials, non-randomized policy pilots, and design experiments. These are the main "ideal types" of experiments that have been discussed in the experimental policy-making literature. They follow different logics and hence entail different benefits and challenges from the perspective of tax policy. While sections 3 and 4 outlined the benefits and challenges that are common to *all* experimental approaches in tax policy, in this section, we zoom in on the promises and pitfalls of these three specific designs.

### *5.1. Randomized controlled trials*

Randomized controlled trials (RCT) hold the promise of allowing the experimenters to draw valid *causal conclusions* about the effects of a project, program or policy (Bravo-Biosca 2020; Burtless 1995; Cook 2002; Dalziel 2018; Ettelt and Mays 2015; Haynes et al. 2012; Pearce and Raman 2014). The random division of subjects into experimental and control groups can be expected to eliminate systematic differences between them and create equivalent groups, which are then subjected to different treatments (in the simplest design, intervention for the experimental group and nothing for the control group) (Burtless 1995; Cook 2002; Ettelt and Mays 2015; Farrelly 2008; Pearce and Raman 2012). In such a design, any observed differences between the groups are attributed to the "treatment" (i.e. the tested policy, program or project), assuming that the experimental group and control group operate in the same policy, social and economic environment (Bell and Peck 2016; Cook 2002). An RCT is expected to create a credible counterfactual (in the form of a control group), and this should enable

policy actors to assess the average treatment effect of the intervention (Bedard and Ouimet 2012; Bravo-Biosca 2020; Burtless 1995; Farrelly 2008; Haynes et al. 2012).

An RCT is widely regarded as "a gold standard method for measuring whether or not a particular intervention works better than doing something else or doing nothing" (Cotterill and Richardson 2010, 156). Given the uncertainties involved in many tax policy measures, RCTs could, in principle, offer opportunities to shed light on these questions in a systematic way.[3] An RCT could be used to assess the additional value (Bravo-Biosca 2020) generated by tax incentives, for example. It could offer the opportunity to test the assumptions or intuitions policy-makers have about a new tax deduction or exemption in terms of their behavioral effects or a new tax on previously untaxed objects or activities. As Haynes et al. (2012) emphasize, the untested intuitions of policy-makers may be wrong, even with policies that should be "guaranteed" to work.

Despite being regarded as the "gold standard" for causal inference (Bravo-Biosca 2020; Strassheim 2020; Webber and Prouse 2018), we should be aware of the dangers inherent in such "methodological triumphalism" (Barrett and Carter 2010, 516) and also pay attention to the key challenges of RCTs. RCTs suffer from major shortcomings – in light of which one could claim that, despite the glitter of recent Nobel prizes, it is "of baser metal than gold" (Barrett and Carter 2010, 516). Indeed, these may be the reasons behind the fact that although "tax law is a promising field in which the government might run randomized experiments", existing experiments only entail tax compliance and welfare (i.e. negative income tax experiments) rather than tax policy more broadly (Abramowicz 2019, 68). In the following, we will discuss the challenges of RCTs.

The RCT design works well under three assumptions: 1) the intervention has to be clearly delineated; 2) the expected outcomes of the intervention have to be measurable and identified in advance of the experiment, and 3) the causal mechanisms examined should be relatively simple (Bedecarrats et al. 2019; Jones and Whitehead 2018). Although such a setup allows for valid statistical conclusions about the average treatment effect (Dalziel 2018; de Leao and Eyal 2019; Kvangraven 2020), it also imposes considerable limitations on what kinds of policies we can test with such a design (Bedecarrats et al. 2019). All of these aspects may pose challenges for using RCTs in tax policy experimentation.

First, RCTs rely on a definition of policy as an "intervention" while actual conceptions of policy are usually more diffuse (Ettelt and Mays 2015, 380). Some aspects of tax policy – e.g. marginal tax rates – can be easily operationalized and quantified for the purposes of an RCT. For example, policy makers may test with an RCT whether lower tax rates for hiring disadvantaged workers have the intended effect. However, many other tax policy domains may be more diffuse. Tax policies, like most other policies, are often constituted by configurations of interacting activity setting, events, and technologies rather than by discrete interventions (Anderson 1975; Ettelt and Mays 2015). As Anderson (1975, 17-18) explains, even in a relatively straightforward negative income tax experiment[4], the "policy treatment" would be characterized by a host of variables such as rules for family size, what counts as a household, household benefits from other sources, accounting periods, and how windfall income is treated. All these are likely to affect the costs of the policy and its efficiency.

3 Indeed, some of the "classic" RCTs during the era of social experiments (1960-1980) in the US used the tax system as a policy instrument – most famously the experiments on negative income taxes (Burtless 1995; Oakley 1998).

4 Negative income tax means that the government sends money back to taxpayers whose income falls below a certain threshold (Moffitt 2003).

Second, RCTs are suitable when policy outcomes are easily measured but challenging when they are fuzzy (Bedecarrats et al. 2019; Bravo-Biosca 2020; de Leao and Eyal 2019). Thus, in the case of broad-aim tax policies, which entail wider societal impacts, general equilibrium effects or changes in economy-wide aggregates, the use of RCT would be difficult (Bedecarrats et al. 2019; Burtless 1995; Deaton 2010; Strassheim 2020). As Burtless (1995, 77) explains, a negative income tax experiment, for example, would be able to capture the effects on labor supply, but without knowing how the employers would alter the wages, it would be impossible to "forecast the full general equilibrium effect".

Furthermore, RCTs work well for assessing *average* effects but are challenging when the policy outcomes are skewed – e.g. when most projects fail and extreme successes are rare, as could be the case in tax-incentivized investment projects (Bravo-Biosca 2020; Bedecarrats et al. 2019; Deaton 2010). For instance, the take-up of tax credits for environmentally sustainable investments may be influenced by exceptional managerial capabilities (a characteristic that is difficult to measure), and this can introduce inaccuracies when estimating the average effect (Dalziel 2018). Similar objections are likely to be present in other tax policy measures that entail high variability in the target group. There might be considerable heterogeneity in how different groups respond to the tax policy treatment (Werner and Riedl 2019) and hence the average effect may conceal the fact that some groups are responding strongly and others are not. Furthermore, while some of the questions in tax policy may zoom in on *average* treatment effects, most of the issues policy-makers might be interested in concern *conditional* effects (e.g. the effects of tax incentives on the most innovative companies or the effects of tax credits on the poorest families). Policy-makers may also be interested in the *distribution* of positive and negative effects in different societal groups or types of firms, rather than just the aggregate average effect (Barrett and Carter 2010; Bedecarrats et al. 2019; Deaton 2010).

Third, "randomized experiments are best when a causal question is simple, sharply focused and easily justified" (Cook 2002, 179). If the policy intervention is targeted at a complicated phenomenon with a complex ecosystem (including unobserved interactions and linkages) it may be difficult to predict the impacts (Bravo-Biosca 2020; Bedecarrats et al. 2019). This is likely to be the case in experimenting with new forms of taxes on digital transactions or carbon footprints, for example. Also, those aspects of tax policy that seek to affect overall macroeconomic outcomes (e.g. stimulating the economy via lowering tax rates) depend on collective actions and interactions – and this clearly undermines the case of using RCTs for studying them.

RCTs are also challenging when the causal processes through which policies affect outcomes take *a long time* (Bravo-Biosca 2020). For reasons of cost and attrition, RCTs tend to be short in duration, which means that in reality only mid-point measurements rather than final indicators can be captured as outcomes (Bedecarrats et al. 2019; Farrelly 2008). Furthermore, the longer the time span of the experiment, the higher the likelihood that other factors besides the intervention start influencing the outcome(s) (de Leao and Eyal 2019; Farrelly 2008). For example, the effects of tax credits for investing in environmentally sustainable technologies may take a long time to materialize, and short-term evaluation may present an inaccurate picture of the eventual effects.

In addition to the aforementioned challenges, the design feature of using *randomization* can pose considerable obstacles to tax policy experiments. In the field of tax policy, randomization

may be difficult to justify ethically, politically, and legally.

*Ethically*, it may be difficult to argue why a policy measure should be denied to some potential recipients who would benefit from it (Bravo-Biosca 2020; Burtless 1995; Jones and Whitehead 2018; Pearce and Raman 2014). For example, if some companies have the opportunity to make use of tax credits for environmentally sustainable investments and others do not, this may be seen as unfair. Furthermore, randomization inherent in RCTs is especially likely to give rise to ethical objections when the experiment involves target groups that are vulnerable (Cotterill and Richardson 2010; Cook 2002). This may make it challenging, for example, to use an RCT to test the effects of giving tax incentives to employers for hiring disadvantaged workers.[5]

*Politically*, randomization may be difficult to justify to the wider public (Nair and Howlett 2016; Strassheim 2020). Policymakers may be concerned about whether it is fair for some people to receive help or benefits and others not if the experiment uses public resources (Nair and Howlett 2016; Strassheim 2020). It may be even more difficult for politicians to impose additional tax burdens on some individuals and businesses but not on others. For example, if the government wanted to try out a car tax on a smaller scale (e.g. applying it to randomly selected car owners) before implementing it on a large scale, in order to examine how it influences people's consumption choices, it may be next to impossible to justify it politically. Similarly, trying out the effects of a carbon tax (e.g. on the investment capacity of firms) in certain parts of the country before extending it to the whole country may run into similar difficulties. If randomization is difficult to justify, policy-makers are likely to fear negative public backlash to the trials (Bravo-Biosca 2020; Strassheim 2020) – and in the domain of tax policy, with clearly measurable costs and benefits, voters might be perceived to be particularly sensitive with regard to being treated unfairly.

*Legally*, challenges may arise from treating people differently when they should be treated equally (Adkins and Ylöstalo 2018; Burtless 1995; de Leao and Eyal 2019), especially if the intervention concerns rights or obligations of citizens (Abramowicz et al. 2010; Burtless 1995). In tax law, as Abramowicz (2019, 69) argues, the main hurdle to RCTs is the "core value of horizontal equity", which leads to "concerns that experiments necessarily produce unequal treatment of similarly situated individuals." However, Abramowicz (2019) argues that there might still be ample room for tax policy experiments where randomization could be legally justified with the fact that the experiment is revenue neutral. Revenue neutrality means that the "treatment group in such an experiment on average pays taxes as high as the control group" (70). In particular, such revenue-neutral designs would be suitable for assessing which combinations of tax deductions and tax rates would be most efficient (from the point of view of allocative efficiency). For example, if the current tax system entails specific deductions but higher (marginal) rates, policy-makers may want to test, with an RCT, whether taxpayers could in fact be better off with abolished (or reduced) deductions but lower rates. For instance, a government could undertake an experiment on abolishing entertainment deductions for businesses: businesses participating in the experiment would give up these deductions but face lower overall tax rates. Conversely, if the current system offers no deductions but lower marginal tax rates, policy-makers may want to experiment with offering deductions, combined

5 A solution proposed to make randomization more palatable to the public is in framing the experiment as "a lottery". For example, in an RCT that has the potential to yield valuable insights in the effects of different marginal tax rates on incentives to work or for entrepreneurship, it may be very difficult to justify to the public why some households are subjected to higher tax rates than others. Abramowicz et al. (2010, 999) suggest a solution according to which the government announces that it "is sponsoring a lottery, the winners of which receive a reduction in their tax rates".

with higher tax rates. Abramowicz (2019) suggests that in addition to small changes to tax systems (e.g. the effects of adding or removing deductions), revenue neutral randomized experiments could be used to test also more significant changes – for example replacing corporate income taxes with government equity in corporations' stock.

### *5.2. Non-randomized policy pilots*

Given the numerous challenges involved in RCTs, they may often not be feasible options for experimental tax policy. In that light, non-randomized policy pilots may offer an alternative route. A policy pilot seeks to test out a new policy approach in a confined setting, and/or on a small subset of the population or jurisdictions, and allows the introduction of a policy in a phase-wise manner (Bailey et al. 2017; Farrelly 2008; Ko and Shin 2017; Nair and Howlett 2016). Unlike RCTs, they do *not* entail randomization, although they may include comparisons with a control group (Philipsen et al. 2021). In order to be viewed as experiments, however, pilots should entail the establishment of concrete systems or processes to learn from them (e.g. Ko and Shin 2017; Lee et al. 2009; Philipsen et al. 2021).

Pilots enable policymakers to assess and adjust a solution before rolling it out nationally (Ko and Shin 2017; Nair and Howlett 2016). For example, a government may be interested in whether tax exemptions applied to cooperation agreements between universities and companies (e.g. creation of industrial professorships) or reducing the payroll tax of high-level researchers working at companies could facilitate the commercialization of basic research. Such exemptions may be first tested with some universities and companies before applying them to all. Special tax regimes for inbound workers could first be offered in some regions of the country to test whether this could be a useful instrument for regional development. In the EU context, tax policy pilots can also take the form of testing different tax solutions in the member states.[6]

Furthermore, owing to their small scale, pilots can foster policy innovations and aid the development of new policy designs (Nair and Howlett 2016). Policy pilots can also facilitate learning how to overcome implementation barriers and improve processes (Ettelt et al. 2015b; Ko and Shin 2017), which might be crucial in the case of taxing novel phenomena. In addition, pilots can catalyze the adoption of policy innovation through *demonstrating* how a new policy can be implemented successfully (Checkland et al. 2021; Ettelt et al. 2015b; Hughes et al. 2020).

Compared to RCTs, which require clearly defined input-output-mechanisms, non-randomized tax policy pilots may be more flexible in their setup and allow a more holistic assessment of a new measure. Also, RCTs tend to take longer time and hence policy-makers may prefer to undertake simpler policy pilots in order to get the information faster (and in line with the electoral cycles) (Ko and Shin 2017), which may be an important advantage in tax policy. There are, however, a number of challenges faced by non-randomized policy pilots.

First, in contrast to RCTs, the causal claims would be significantly weaker since we do not know "how the targeted population would have fared in the absence of treatment" (Farrelly 2008,

6 For example, Directive 1999/85/EC foresaw the possibility for member states to apply a reduced VAT rate on labor-intensive services to create jobs. The evaluation of the experiment by the Commission in 2003 revealed that the reduced tax rates were not translated into lower consumer prices and no clear impact on employment rates could be identified. Thus, it was concluded that "such measures were usually not very effective and the cost to the budget was high in relation to any impact the measures might have on the economy" (Van Gestel and Van Dijck 2011, 545).

8). Effects can be difficult to assess because "there may be alternative explanations for any observed changes" (Abramowicz 2008, 34). This means that in a politically contested policy domain like taxation, the findings of non-randomized pilots can be more easily attacked by political actors who are not pleased by the results of the pilot. Since there is no randomized control, policy makers cannot be confident whether in a policy pilot tax revenues have fallen because of the piloted measure or due to exogenous reasons, such as economic recession (Abramowicz 2019).

Second, replicating the success of a policy pilot may be challenged by differences in context (van der Heijden 2018; Farrelly 2008). The findings of policy pilots are likely to be influenced by various interdependent social, political, and economic factors, and this limits external validity. The groups, organizations or regions analyzed in a pilot may be systematically different from the rest of the population (Bailey et al. 2017). Furthermore, participants in the pilot may be motivated to ensure success and have incentives that may not be shared by the broader population (van der Heijden 2018). Van der Heijden (2018, 1385) refers to a frontrunner paradox, as experiments often "look for actors who want to be actively involved in solving a problem, who do not mind deviating from routines, and who are willing to take risks." These attitudes may not be characteristic to a broader population or a larger set of organizations, making scaling up challenging. For example, regions involved in a pilot testing a new tax regime for inbound workers may be particularly motivated to use that measure for promoting economic development and engage in extensive communicative efforts in spreading knowledge about that option.

Third, due to the limited geographical scope of policy pilots, spillovers to or from other regions may constitute an important challenge, especially in tax policy, where the incentives to move across jurisdictional boundaries are likely to be significant. As economic activity can move across jurisdictions, firms may shift its activities from a more highly taxed pilot location to one with a lower tax or the other way around (Abramowicz 2019), posing considerable challenges in making accurate assessments of the actual effects of the pilot.

Fourth, similarly to RCTs, an important question tax policy pilots have to wrestle with is the question of meaningful duration. If the designated time period is too short, we may not be able to capture the full impacts of the tested policy. Ranchordas (2015a, 912) suggests that this can be pertinent in experimental tax legislation that seeks to stimulate investment in renewable energy and advance clean-energy innovation. Since it takes a long time to develop a wind farm, for example, uncertainty with regard to the renewal of such tax credits can undermine long-term investment.

Finally, analogously to RCTs, tax policy pilots may give rise to ethical, legal and political challenges (Bauknecht 2019; Van Gestel and Van Dijck 2011). For example, offering a special tax regime in some geographical jurisdictions or to some organizations may be seen as violating the principles of horizontal equity, equal treatment and legal certainty (Huitema et al. 2018; Philipsen et al. 2021). However, it may be more feasible to use objectively defined and politically justifiable criteria for picking the subjects in a tax experiment in a non-randomized rather than a randomized way. For example, offering preferential tax regimes first to the least developed geographic regions or offering tax exemptions to the most promising cooperation networks of universities and businesses may be more palatable to politicians and the electorate than randomization.

### *5.3. Design experiments*

In a design experiment, "a solution concept (an idea, design, program, project, and so on) to a particular problem is created, and iteratively refined based on continuous feedback from stakeholders immersed in the experiment" (Zurbriggen and Lago 2019, 440). Design experiments draw on design thinking that emphasizes the importance of systems thinking, user centrism, regular iteration, and creativity (Clarke and Craft 2019, 6). Such an experiment adopts a "probe and learn" strategy, in order to understand the intervention (Ansell and Bartenberger 2016, 68) and to "re-specify and re-calibrate" the solution until it works (Stoker and John 2009, 358). Thus, the experiment would progress through iterative cycles of design, real-world testing and redesign based on lessons from earlier iterations (Stoker and John 2009, 256; van der Heijden and Hong 2021, 1119). Design experiments are less concerned with exploring causality than with manipulating an intervention in order to reach an acceptable outcome (Stoker and John 2009). Unlike in RCTs where policy experimenters can remain detached from the context, design experiments entail immersion in "thickly experiential policy contexts" (Lewis et al. 2020, 116).

A key feature of design experiments is taking into account the experiences of the persons affected by the issue. Design experiments assume that in order to address societal (especially wicked) problems, expertise from both professionals and members of the public are needed as their perceptions of the problems may diverge (Einfeld and Blomkamp 2021; Lewis et al. 2020). Hence, diversifying the sources of knowledge through the experiment can help policy-makers better understand and predict people's needs, perceptions and behaviors in real-life contexts (Clarke and Craft 2019; Einfeld and Blomkamp 2021; Lewis et al. 2020). In addition to refining policy solutions, stakeholders or end-users may also be involved in defining the policy *goals* and generating ideas for solutions (Clarke and Craft 2019; Einfeld and Blomkamp 2021; Lewis et al. 2020).

Design experiments share similarities with explorative pilots that are controlled only to a limited extent (Ansell and Bartenberger 2016). What makes the design experiments different is their explicitly *iterative* character. Thus, design experiments may provide dynamic and timely ways to change course during the experimentation process, therefore making them suitable for developing solutions for uncertain environments and complex problems. In tax policy, however, it is difficult to conceive of subjecting the stakeholders to varying tax rates, deductions, or exemptions in an iterative way. Given the importance of stable tax horizons in the investment and other decisions of individuals and businesses, such continuous changes and tweaks in the tax regime would be challenging.

Nevertheless, design experiments and the possibilities for iterative adjustments could be of value for developing novel tax policy solutions. For example, before introducing new taxes to quickly changing domains, the flexibility of the approach and insights from the stakeholders (Einfeld and Blomkamp 2021; Stoker and John 2009) may be crucial for the policymakers in assessing the feasibility of such taxes. As design experiments often seek to create prototypes and collect feedback on the potential responses to these policy measures (Clarke and Craft 2019; Einfeld and Blomkamp 2021), they could provide valuable insights about how the potential taxpayers, if they were subjected to such models, would react and adjust their behavior (Stoker and John 2009).

While RCTs require policy-makers to identify clear expectations about measurable impacts in order to commence the experiment, design experiments tend to have a "fuzzy front end", which allows for the "exploration of open-ended questions" (Clarke and Craft 2019, 9). That can be valuable in utilizing stakeholders' knowledge for making use of the tax system in solving new societal problems. By offering opportunities to test new approaches in an iterative way, design experiments offer a safe space for trying out novel solutions, reducing the fear of failure and hence promote innovation in policy design (Clarke and Craft 2019; Stoker and John 2009). Furthermore, by emphasizing the lived experiences of those affected by a policy design, such experiments can cater to the need to adapt the policy to different target populations as opposed to settling on a one-size-fits all approach (Clarke and Craft 2019, 7). Design experiments could be utilized in testing how potential tax payers would perceive information about new tax incentives before these changes are rolled out. The way tax incentives are presented can play a crucial role in how they are perceived by the taxpayers, and this, in turn can affect their potential tax behavior in the future (Werner and Riedl 2019). Complex tax regimes in particular can lead to weaker behavioral adjustments than expected by policy makers (e.g. Abeler and Jäger 2015). Design experiments can hence serve to assess the perceived complexity of a tax regime and help to mitigate potential distortions in the perception of the tax policy change.

In sum, design experiments could be valuable in understanding the reactions and opinions of policy target groups or developing new technological solutions that presume high inclusion of stakeholders in order to reach desired outcomes of tax policy. Nevertheless, design experiments also do face some crucial challenges in tax policy experimentation.

First, as continuous feedback and quick path adjustment is crucial during design experiments, they are not suitable in situations where effects of the action take a long time to appear or when there is a lack of control in different phases of experiment.

Second, design experiments tend to be applicable to small-scale policy issues rather than large-scale ones: they tend to be employed for "discrete service redesign projects" and in "exploratory work of scoping problems", rather than for the development of broader policy proposals or systemic reforms (Lewis et al. 2020, 113-114). It would be difficult to use them for assessing macroeconomic outcomes as the scope and intensity of design experiments (entailing immersion of experimenters and stakeholders) necessarily means a limited number of data points and the need to stay at the micro level.

Third, while the notions of "user centrism" inherent in design experiments may work well in the private sector context (where design thinking originates from), they may clash with the notions of rights and obligations of citizens, which prevail in tax policy. The stakeholders whose feedback is collected during the experiment may view taxes from their own narrow material point of view (with the goal to minimize their own tax burden), which may clash with broader policy goals. As Clarke and Craft (2019, 14) emphasize, user centrism may not be an easily applicable principle for contentious policy design choices, which include a broad range of users with conflicting needs and expectations, and thus require trade-offs between different values. The act of weighing these needs in tax policy is inherently political, subjective, normative, and ultimately falls to the accountable elected officials.

## Conclusion

In light of the growing importance of experimental approaches in public policy, we proposed a binary question in the title of this article: to experiment or not to experiment in the field of tax policy? As the discussion above shows, the answer to that question is much more nuanced than a simple "yes" or "no". Although Abramowicz (2019) in his thought-provoking study advocated an extensive use of randomized experimentation in tax policy, our claims are considerably more cautious. Using experiments in tax policy does have a range of promises but also a wide spectrum of pitfalls.

On the one hand, experimental approaches have the potential to increase both the substantive quality of tax policy and the policy-making process. Experimental tax policy can potentially help policy makers alleviate some of the uncertainties and information asymmetries with regard to the actual impacts of new tax policy measures, avoid adverse effects, build consensus, and foster attention to consequences rather than ideological hunches in tax policy debates. On the other hand, however, tax policy experiments may give rise to ethical, legal and political challenges. Most of all, experimenting may be seen to violate the principles of horizontal equity, equal treatment and legal certainty. Tax policy is a field characterized by contentious policy choices, which engage a broad range of stakeholders with conflicting perceptions and needs. The act of balancing these perceptions is inherently complex, political and demands trade-offs between different value considerations. Altogether, the application of experimental approaches in the field of tax policy can be characterized by two core challenges that experimenters need to acknowledge and address. We label them a political challenge and a methodological challenge.

First, experimenting in tax policy may clash starkly – perhaps even more so than in many other policy fields – with the political nature of democratic decision-making. Most importantly, tax policy questions are likely to be profoundly influenced by the ideological leanings of policy makers and even define the core identity of some parties, which makes it challenging to test tax policy questions in a genuinely open way. Also, given that experimenting in tax policy would often entail materially benefitting or burdening (in a very clearly measurable way) some groups of taxpayers at the expense of others, such experiments may be more vulnerable than experiments in many other domains to legal challenges. Despite the main promise of experimental approaches to provide new knowledge in complex environments where various actors, institutions and technologies interact, the low tolerance of the tax policy field towards uncertainty and unpredictability may render the experimental approaches politically, ethically or legally unfeasible.

Second, methodologically, different experimental designs entail different strengths and promises, and there is no universal recipe to follow. The experimenters need to decide whether to prioritize causal explanations facilitated by RCTs or exploratory and collaborative aims fostered by other designs. While RCTs enable stronger causal conclusions, non-randomized pilots allow the testing of more holistic policy solutions, and design experiments offer more open-ended approaches and a stronger focus on stakeholder experience. Randomization inherent in RCTs might be a particularly challenging "political sell" in the field of tax policy, since it concerns the core rights and obligations of citizens. Both RCTs and policy pilots could face the accusations of violating horizontal equity – a key principle in tax policy. The user-centrism inherent in design experiments may clash with the notions of taxes as obligations.

Although we discussed the various designs as distinct options, the lines between them can in reality be somewhat blurred, and they could be used in a sequential manner (Bravo-Biosca 2020; Pomeranz and Vila-Belda 2019). For example, a design experiment could first help take a fresh look at some tax policy goals, followed by a simple non-randomized policy pilot that tests variants of a solution offered in an exploratory way, and an RCT could then zoom in on specific causal questions. Our recommendation is that policy-makers should remain open to a diversity of possible designs of experiments in tax policy. In particular, they should be aware of the dangers of conceiving of experimentation very narrowly, only in terms of RCTs – which seems to be the default given the tendency of the evidence-based policy-making movement to regard RCTs as the "gold standard" (Adkins and Ylöstalo 2018; Barrett and Carter 2010; Bedecarrats et al. 2019; Dalziel 2018; de Leao and Eyal 2019; Pearce and Raman 2014; Strassheim 2020).

In sum, there are considerable constraints for the scope of tax policy experimentation. Thus, while in some disciplines and perhaps even policy-fields experimentation could be the prominent *Zeitgeist,* in tax policy it is likely to remain on the margins. We conclude that in the field of tax policy, experimentation could be feasible in carefully crafted revenue neutral experiments, in phase-wise introductions of larger programs where limited resources can offer a justifiable reasoning for benefitting some taxpayers before others, and fostering stakeholder discussions in very new tax policy domains.

## Acknowledgments

This work was supported by the Estonian Research Council grant PRG1125.

## References

Abeler, J. and S. Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy* 7(3), 1-28.

Abramowicz, M. 2008. "Randomized Legal Experimentation." *GWU Legal Studies Research Paper* 447. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1273875.

Abramowicz, M. 2019. "Tax Experimentation." *Florida Law Review* 71, 65-122.

Abramowicz, M., I. Ayres and Y. Listokin. 2010. "Randomizing Law." *University of Pennsylvania Law Review* 159(4), 929-1005.

Adkins, L. and H. Ylöstalo. 2018. "Experimental Policy, Price and the Provocative State." *Distinktion: Journal of Social Theory* 19(2), 152-169.

Akdogan, M.U. 2021. "Digital Service Taxes as the Fiscal Result of Digital Transformation." In U. Hacioglu and T. Aksoy (eds). *Financial Ecosystem and Strategy in the Digital Era*. Springer, 355-378.

Alm, J. and A. Malézieux. 2021. "40 Years of Tax Evasion Games: A Meta-Analysis." *Experimental Economics* 24(3), 699-750.

Anderson, A.B. 1975. "Policy Experiments: Selected Analytic Issues." *Sociological Methods & Research* 4(1), 13-30.

Ansell, C.K. and M. Bartenberger. 2016. "Varieties of Experimentalism." *Ecological Economics* 130, 64-73.

Bailey, S., K. Checkland, D. Hodgson, A. McBride, R. Elvey, S. Parkin, K. Rothwell and D. Pierides. 2017. "The Policy Work of Piloting: Mobilising and Managing Conflict and Ambiguity in the English NHS." *Social Science & Medicine* 179, 210-217.

Ballard-Rosa, C., L. Martin and K. Scheve. 2017. "The Structure of American Income Tax Policy Preferences." *The Journal of Politics* 79(1), 1-16.

Barrett, C.B. and M.R. Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections." *Applied Economic Perspectives and Policy* 32(4), 515-548.

Bauknecht, D., T.S. Bischoff, K. Bizer, M. Führ, P. Gailhofer, D.A. Heyen, T. Proeger, and K. von der Leyen. 2020. "Exploring the Pathways: Regulatory Experiments for Sustainable Development – An Interdisciplinary Approach." *Journal of Governance and Regulation* 9(3), 49-71.

Bedard, P.O. and M. Ouimet. 2012. "Cognizance and Consultation of Randomized Controlled Trials among Ministerial Policy Analysts." *Review of Policy Research* 29(5), 625-644.

Bedecarrats, F., I. Guerin and F. Roubaud. 2019. "All that Glitters is not Gold: The Political Economy of Randomized Evaluations in Development." *Development and Change* 50(3), 735-762.

Bell, S.H. and L.R. Peck. 2016. "On the Feasibility of Extending Social Experiments to Wider Applications." *Journal of MultiDisciplinary Evaluation* 12(27), 93-112.

Bravo-Biosca, A. 2020. "Experimental Innovation Policy." *Innovation Policy and the Economy* 20(1), 191-232.

Burtless, G. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* 9(2), 63-84.

Carmines, E.G., M.J. Ensley and M.W. Wagner. 2012. "Who Fits the Left-Right Divide? Partisan Polarization in the American Electorate." *American Behavioral Scientist* 56(12), 1631-1653.

Checkland, K., J. Hammond, A. Coleman, J. Macinnes, R. Mikelyte, S. Croke, J. Billings, S. Bailey, and P. Allen. 2021. "'Success' in Policy Piloting: Process, Programs, and Politics." *Public Administration*. Published ahead-of-print 1-18. https://doi.org/10.1111/padm.12790.

Clarke, A. and J. Craft. 2019. "The Twin Faces of Public Sector Design." *Governance* 32(1), 5-21.

Cook, T.D. 2002. "Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community Has Offered for Not Doing Them." *Educational Evaluation and Policy Analysis* 24(3), 175-199.

Cotterill, S. and L. Richardson. 2010. "Expanding the Use of Experiments on Civic Behavior: Experiments with Local Government as a Research Partner." *The Annals of the American Academy of Political and Social Science* 628(1), 148-164.

Dalziel, M. 2018. "Why are there (Almost) No Randomised Controlled Trial-Based evaluations of Business Support Programmes?" *Palgrave Communications* 4(1), 1-9.

Deaton, A. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48(2), 424-55.

Druckman, J.N., D.P. Green, J.H. Kuklinski and A. Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100(4), 627-635.

Einfeld, C. and E. Blomkamp. 2021. "Nudge and Co-Design: Complementary or Contradictory Approaches to Policy Innovation?" Published ahead-of-print in *Policy Studies*. https://doi.org/10.1080/01442872.2021.1879036.

Elkins, D. 2006. "Horizontal Equity as a Principle of Tax Theory." *Yale Law & Policy Review* 24, 43-90.

Ettelt, S. and Mays, N. 2015. "RCTs: How Compatible Are They with Policy-Making?" *British Journal of Healthcare Management* 21(8), 379-382.

Ettelt, S., N. Mays and P. Allen. 2015a. "Policy Experiments: Investigating Effectiveness or Confirming Direction?" *Evaluation* 21(3), 292-307.

Ettelt, S., N. Mays and P. Allen. 2015b. "The Multiple Purposes of Policy Piloting and Their Consequences: Three Examples from National Health and Social Care Policy in England." *Journal of Social Policy* 44(2), 319-337.

Farrelly, R. 2008. "Policy on Trial: Randomised Trials Are the Best Tool We Have for Finding Out if Policies Really Work." *Policy: A Journal of Public Policy and Ideas* 24(3), 7-12.

Hallsworth, M. 2014. "The Use of Field Experiments to Increase Tax Compliance." *Oxford Review of Economic Policy* 30(4), 658-679.

Haynes, L., B. Goldacre and D. Torgerson. 2012. "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials." Cabinet Office-Behavioural Insights Team. Available

at http://dx.doi.org/10.2139/ssrn.2131581.

Heilmann, S. 2008. "Policy Experimentation in China's Economic Rise." *Studies in Comparative International Development* 43(1), 1-26.

Heldeweg, A.M. 2015. "Experimental Legislation Concerning Technological & Governance Innovation: An Analytical Approach." *The Theory and Practice of Legislation* 3(2), 169-193.

Hughes, S., S. Yordi, L. and Besco. 2020. "The Role of Pilot Projects in Urban Climate Change Policy Innovation." *Policy Studies Journal* 48(2), 271-297.

Huitema, D., A. Jordan, S. Munaretto and M. Hilden. 2018. "Policy Experiments: Core Concepts, Political Dynamics and Impacts." *Policy Sciences* 51(2), 143-159.

John, P. 2014. "Policy Entrepreneurship in UK Central Government: The Behavioural Insights Team and the Use of Randomized Controlled Trials." *Public Policy and Administration* 29(3), 257-267.

Jones, R. and M. Whitehead. 2018. "'Politics Done like Science': Critical Perspectives on Psychological Governance and the Experimental State." *Environment and Planning D: Society and Space* 36(2), 313-330.

Karvonen, A. and B. Van Heur. 2014. "Urban Laboratories: Experiments in Reworking Cities." *International Journal of Urban and Regional Research* 38(2), 379-392.

Kay, J.A. 1990. "Tax Policy: A Survey." *The Economic Journal* 100(399), 18-75.

Klitgaard, M.B., G. Schumacher and M. Soentken. 2015. "The Partisan Politics of Institutional Welfare State Reform." *Journal of European Public Policy* 22(7), 948-966.

Ko, K. and K. Shin. 2017. "How Asian Countries Understand Policy Experiment as Policy Pilots?" *Asian Journal of Political Science* 25(3), 253-265.

Kvangraven, I.H. 2020. "Nobel Rebels in Disguise: Assessing the Rise and Rule of the Randomistas." *Review of Political Economy* 32(3), 305-341.

de Leao, L.S. and G. Eyal. 2019. "The Rise of Randomized Controlled Trials (RCTs) in International Development in Historical Perspective." *Theory and Society* 48(3), 383-418.

Lee, C. and L. Ma. 2020. "The Role of Policy Labs in Policy Experiment and Knowledge Transfer: A Comparison across the UK, Denmark, and Singapore." *Journal of Comparative Policy Analysis: Research and Practice* 22(4), 281-297.

Lee, S.-Y., Y.-R. Jung and E. Lee. 2009. "An Exploratory Study on Pilot Projects: The Ministry for Health, Welfare, and Family Affairs in Korea." *The Korean Journal of Public Administration* 47(4), 279-305.

Levitt, S.D. and J.A. List. 2011. "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments." *American Economic Journal: Applied Economics* 3(1), 224-238.

Lewis, J.M., M. McGann and E. Blomkamp. 2020. "When Design Meets Power: Design Thinking, Public Sector Innovation and the Politics of Policymaking." *Policy & Politics* 48(1), 111-130.

McFadgen, B.K. 2019. "Connecting Policy Change, Experimentation, and Entrepreneurs." *Ecology and Society* 24(1), 30.

McFadgen, B. and Huitema, D. 2017. "Are All Experiments Created Equal? A Framework for Analysis of the Learning Potential of Policy Experiments in Environmental Governance." *Journal of Environmental Planning and Management* 60(10), 1765-1784.

McFadgen, B. and D. Huitema. 2018. "Experimentation at the Interface of Science and Policy: A Multi-Case Analysis of How Policy Experiments Influence Political Decision-Makers." *Policy Sciences* 51(2), 161-187.

McGann, M., E. Blomkamp and J.M. Lewis. 2018. "The Rise of Public Sector Innovation Labs: Experiments in Design Thinking for Policy." *Policy Sciences* 51(3), 249-267.

Mascagni, G. 2018. "From the Lab to the Field: A Review of Tax Experiments." *Journal of Economic Surveys* 32(2), 273-301.

Millo, Y. and J. Lezaun. 2006. "Regulatory Experiments: Genetically Modified Crops and Financial Derivatives on Trial." *Science and Public Policy* 33(3), 179-190.

Moffitt, R.A. 2003. "The Negative Income Tax and the Evolution of U.S. Welfare Policy." *Journal of Economic Perspectives* 17(3), 119-140.

Musgrave, R.A. 1990. "Horizontal Equity, Once More." *National Tax Journal* 43(2), 113-122.

Nair, S. and M. Howlett. 2016. "Meaning and Power in the Design and Development of Policy Experiments." *Futures* 76, 67-74.

Oakley, A. 1998. "Public Policy Experimentation: Lessons from America." *Policy Studies* 19(2), 93-114.

Osterloh, S. and M. Debus. 2012. "Partisan Politics in Corporate Taxation." *European Journal of Political Economy* 28(2), 192-207.

Pearce, W. and S. Raman. 2014. "The New Randomised Controlled Trials (RCT) Movement in Public Policy: Challenges of Epistemic Governance." *Policy Sciences* 47(4), 387-402.

Peters, B.G. 1998. "The Experimenting Society and Policy Design." In W. Dunn (ed.). *The Experimenting Society. Essays in Honor of Donald T. Campbell*. New Brunswick, NJ: Transaction, 125-140.

Philipsen, S., E.F. Stamhuis and M. de Jong. 2021. "Legal Enclaves as a Test Environment for Innovative Products: Toward Legally Resilient Experimentation Policies." *Regulation & Governance* 15(4), 1128-1143.

Pomeranz, D. and J. Vila-Belda. 2019. "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities." *Annual Review of Economics* 11, 755-781.

Ranchordas, S. 2013. "The Whys and Woes of Experimental Legislation." *The Theory and Practice of Legislation* 1(3), 415-440.

Ranchordas, S. 2015a. "Innovation Experimentalism in the Age of the Sharing Economy." *Lewis & Clark Law Review* 19, 871-924.

Ranchordas, S. 2015b. "Time, Timing, and Experimental Legislation." *The Theory and Practice of Legislation* 3(2), 135-139.

Ranchordas, S. 2015c. "Innovation-Friendly Regulation: The Sunset of Regulation, the Sunrise of Innovation." *Jurimetrics* 5(2), 201-224.

Rangoni, B. and J. Zeitlin. 2021. "Is Experimentalist Governance Self-Limiting or Self-Reinforcing? Strategic Uncertainty and Recursive Rulemaking in European Union Electricity Regulation." *Regulation & Governance* 15(3), 822-839.

Shome, P. 2021. *Taxation History, Theory, Law and Administration*. Springer.

Stein, H., S. Cunningham and P. Carmody. 2021. "The Rise of 'Behavioral Man': Randomized Controlled Trials and the 'New' Development Agenda." *Human Geography* 14(1), 62-75.

Stoker, G. 2010. "Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and Policy Relevance." *Political Studies* 58(2), 300-319.

Stoker, G. and P. John. 2009. "Design Experiments: Engaging Policy Makers in the Search for Evidence about What Works." *Political Studies* 57(2), 356-373.

Strassheim, H. 2020. "De-Biasing Democracy. Behavioural Public Policy and the Post-Democratic Turn." *Democratization* 27(3), 461-476.

Tõnurist, P., R. Kattel and V. Lember. 2017. "Innovation Labs in the Public Sector: What They Are and What They Do?" *Public Management Review* 19(10), 1455-1479.

Van Gestel, R. and G. Van Dijck. 2011. "Better Regulation through Experimental Legislation." *European Public Law* 17(3), 539-533.

Van der Heijden, J. 2018. "From Leaders to Majority: A Frontrunner Paradox in Built-Environment Climate Governance Experimentation." *Journal of Environmental Planning and Management* 61(8), 1383-1401.

Van der Heijden, J. and S.H. Hong. 2021. "Urban Climate Governance Experimentation in Seoul: Science, Politics, or a Little of Both?" *Urban Affairs Review* 57(4), 1115-1148.

Voß, J.P. and A. Simons. 2018. "A Novel Understanding of Experimentation in Governance: Co-producing Innovations between 'Lab' and 'Field'". *Policy Sciences* 51(2), 213-229.

Webber, S. and C. Prouse. 2018. "The New Gold Standard: The Rise of Randomized Control Trials and Experimental Development." *Economic Geography* 94(2), 166-187.

Werner, P. and A. Riedl. 2019. "The Role of Experiments for Policy Design." In A. Schram and A. Ulu (eds.). *Handbook of Research Methods and Applications in Experimental Economics.* Cheltenham: Edward Elgar Publishing, 295-311.

Zurbriggen, C. and M.G. Lago. 2019. "An Experimental Evaluation Tool for the Public Innovation Lab of the Uruguayan Government." *Evidence & Policy* 15(3), 437-451.

**Ringa Raudla** (ringa.raudla@taltech.ee) is Professor of Fiscal Governance, Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Estonia. Her main research interests are fiscal policy, public budgeting, public administration reform, and institutional economics. Her recent publications include articles in *Public Administration Review*, *Governance*, *Public Administration*, *Policy Studies Journal*, *the American Review of Public Administration*, *Journal of Public Policy*, *Public Money & Management*, *Public Budgeting & Finance, Public Management Review, International Review of Administrative Sciences*, and *Journal of Comparative Policy Analysis*. She has also worked as a consultant for various governmental organizations, including the Ministry of Finance and the National Audit Office in Estonia, and international organizations, including the World Bank. Ringa Raudla received the Estonian national research award in social sciences in 2018. She is an editorial board member of seven different journals in the fields of public administration, public management, governance and economics.

**Külli Sarapuu** (kylli.sarapuu@taltech.ee) is Associate Professor of Public Sector Management and Organization, Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Estonia. Before joining TalTech, she worked for the Estonian Government Office and the Ministry of Research and Education. Külli's research focuses on public sector organization, civil service systems, public sector coordination, and the governance of small states. She has published numerous book chapters and articles on these topics, including in the journals *Public Management Review*, *International Review of Administrative Sciences* and *International Journal of Public Administration*.

**Egert Juuse** is a Research Fellow, Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Estonia. He is also the head of two Master level programs: Technology Governance and Digital Transformation (MA) and E-Governance and Public Sector Innovation (MSc). He has received a BA, MA and PhD in Technology Governance from Tallinn University of Technology. His research interests include various aspects related to economic development with a particular focus on global value chains; financial policies (incl. Europeanization); financialization and financial crises; innovation and industrial policies. He has also worked as an advisor to a Minister at the Ministry of Economic Affairs and Communications, Estonia.

**Nastassia Harbuzova** is a Doctoral Student and a Junior Research Fellow at the Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology. Her research concentrates on macroeconomic policymaking from both ideational and implementation perspectives. Currently, she is focused on studying how policy learning occurs in field experiments undertaken in different institutional contexts. Her educational background includes Economic Policy, Mathematical Methods in Economics and Intellectual Property Management. Previously she worked as a Lecturer at the Belarusian State University as well as co-led a student science club, CLIO BSU, focusing on research in hetero- and orthodox economics. In 2021 she published a taskbook, *Macroeconomics*, for undergraduates utilized by BSU.

**Kerli Onno** is an industrial PhD student at the Nurkse Department. Her research focuses on the value of data capabilities in public sector organizations. As she works as the Head of the Analysis Unit at the Estonian Tax and Customs Board, she is deeply interested in the ways data capabilities support public sector organizations in adapting to the changing environments and in developing dynamic ways of operating to provide long-term public value. In addition, she is exploring various experimental approaches in public sector organizations and is involved in a research project, *Experimental approaches and institutional innovations in the domains of fiscal and financial policy*. Kerli Onno holds an MA in Communication Management (*cum laude*) from Tallinn University and an MA in Internal Security from the Estonian Academy of Security Sciences with a focus on media studies and strategic communication in national defence.

**Johanna Vallistu** is a PhD student at the Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology. She also works as an expert at the Foresight Centre at the Parliament of Estonia. At the Foresight Centre, Johanna has led the foresight projects of Labor Market 2035, Financial Well-being of the Future Elderly and currently leads the project on the Future of Data Freedom. Johanna's main interests lie in the future social insurance system and how digitalization impacts governance.

**Aleksandrs Cepilovs** is a research fellow (part-time) at the Ragnar Nurkse Department of Innovation and Governance. He is also a research officer at the European Union Agency for the Operational Management of Large-Scale IT Systems in the Areas of Freedom, Justice and Security (eu-LISA). His research interests include innovation policy, policy transfer and policy learning, public sector innovation and e-Governance. Most recently, Aleksandrs has been involved in a research project focusing on experimental approaches and institutional innovations in the domains of fiscal and financial policy.

**Article II**

**Onno, K.** (2025). The Role of Data Capabilities in Developing Public Sector Dynamic Capabilities: Lessons from the COVID-19 Pandemic. International Journal of Public Administration, 48(11), 736–747.

INTERNATIONAL JOURNAL OF PUBLIC ADMINISTRATION
https://doi.org/10.1080/01900692.2024.2411580



# The Role of Data Capabilities in Developing Public Sector Dynamic Capabilities: Lessons from the COVID-19 Pandemic

Kerli Onno

Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology (TalTech), Tallinn, Estonia

**ABSTRACT**

For public sector organizations, adapting to unexpectedly changing circumstances has become a substantial challenge. The COVID-19 pandemic provides a fruitful ground for exploring how organizations operate and change under intense circumstances. This article draws on the case of Estonian public sector organizations to understand the way data capabilities were used to improve operational routines related to pandemic response. Certain data capabilities helped organizations to carry out tasks related to the pandemic response, which impelled rapid adaptation of routines. However, it remains a challenge to embrace long-term changes going beyond temporary fixes of the pandemic and to develop dynamic capabilities.

**KEYWORDS**

Public administration; public management; data capabilities; COVID-19 pandemic, crises

## Introduction

Public sector organizations operate in an increasingly uncertain environment, which has led to the continuous need to keep up with the expectations of their stakeholders and societies. This study focuses on the context of one of the most influential challenges for public sector organizations in recent years—the COVID-19 pandemic. Although the way organizations responded to the pandemic, coped with the related challenges and learned from the experience has already been discussed in academic literature (see, for example: Edwards & Ott, 2021; Klöckner et al., 2023; A. Sharma et al., 2021), much less is known about whether and how the pandemic as an exogenous turbulence impacted existing organizational routines, defined as "repetitive, recognizable patterns of interdependent actions" (Feldman & Pentland, 2003, p. 95) um

Since the 1990s, the dynamic capabilities framework (Teece et al., 1997) has been elaborated on to deepen our understanding about the factors that make some organizations capable of quickly adapting to changing environments. This framework has gained a lot of attention from researchers (e.g., Ambrosini et al., 2009; Eisenhardt & Martin, 2000; Haarhaus & Liening, 2020; Winter, 2003) and provides useful insights into the ways public sector organizations adapt, change, or innovate (Nuhu et al., 2019; Panagiotopoulos et al., 2019; Piening, 2013). It is argued that dynamic capabilities lead to organizational change through supporting an organization's ability to *sense* the surrounding environment to identify and assess opportunities, to *seize* those opportunities through mobilization of resources, and to implement *transformational* activities to ensure continuous renewal (Conboy et al., 2020; Teece, 2007).

The current study is centered around data capabilities that are closely related to what Mayne et al. (2020) understand as data-analytic capability: the ability to "collect, process, and analyse different types of information" to increase understanding of "external context, internal conditions, and performance" (Mayne et al., 2020, p. 40). Data capabilities are used to extract valuable knowledge from data about the surrounding environment, which provides input for decision-making. The current article explores three routines that require the use of data capabilities: allocating financial support to vulnerable groups in society; evaluating the economic impact of the pandemic; and providing management information to the ETCB board. These routines are regularly executed in public sector organizations but were modified in response to the pandemic. This study focuses on turbulence-induced changes in routines: it traces actions and patterns of action (Feldman, 2016, p. 29) related to pandemic response, explores the role of data capabilities in routine adaptation, and aims to understand the implications for dynamic capability building in the public sector.

From a theoretical perspective, this article conceptualizes dynamic capability building in turbulent

**CONTACT** Kerli Onno kerlionno@gmail.com Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology (TalTech), Ehitajate tee 5, Tallinn 19086, Estonia

Supplemental data for this article can be accessed online at https://doi.org/10.1080/01900692.2024.2411580

conditions. Although turbulent times may bring about the need to adjust organizational routines to cope with the challenges, and therefore, the need to employ dynamic capabilities, less is known about whether turbulence induces continuous and long-term changes in operational routines. Using the example of data capabilities, an analytical framework is proposed that contributes to understanding the ways in which dynamic capabilities are developed through improvement of organizational routines. This offers a good starting point for discussion on the organizational configurations and conditions that act as enablers in building dynamic capabilities, especially in turbulent conditions. Furthermore, tracing the way data capabilities influenced public sector routines during the pandemic provides practical insights in two respects. First, it expands our understanding of how public organizations responded to the turbulence and the ways the respective routines were adapted to rapidly changing conditions. Second, it helps to explore whether and how the knowledge and experience gained during the pandemic contributed to the long-term improvement of existing routines. Hence, two research questions are proposed: (1) How did data capabilities enhance the adaptation of routines during the COVID-19 pandemic? (2) Do data capabilities support the development of dynamic capabilities in the public sector, and if so, under what conditions?

Researchers have started to explore how dynamic capabilities influence and enhance the operational routines of organizations (e.g., Conboy et al., 2020; Jantunen et al., 2018; Mero & Haapio, 2022). However, such discussions have not yet been held in the public governance and management literature and in-depth studies regarding the ways dynamic capabilities are developed remain scarce.

To understand the ways dynamic capabilities in public sector are developed, this paper uses the case of the Estonian public sector, especially focusing on three routines that were influenced by the COVID-19 pandemic and dependent upon the data capabilities of the Estonian Tax and Customs Board (ETCB). The ETCB can be considered one of the most advanced Estonian public sector organizations in terms of data-analytics and data-based process automation (Lember et al., 2018). The case provides a good starting point to explore the impact of organizational data capabilities on inter-organizational routines and their role in developing dynamic capabilities.

This article is structured as follows. The second section introduces the analytical framework that was used to explain how data capabilities support the development of dynamic capabilities. The third section details the research method and is followed by an overview of the main results of the case study. Next, the discussion section combines the analytical framework and the results, highlighting the most relevant findings of the study. Finally, conclusions and suggestions for further research are outlined.

## Analytical framework

This section sets out an analytical framework to elucidate how data capabilities provide public sector organizations with means to change their existing routines and develop their dynamic capabilities in turbulent times. Routines are "repetitive, recognizable patterns of interdependent actions, carried out by multiple actors" (Feldman & Pentland, 2003, p. 95). In other words, routines are practices that are "enacted in order to do something in and for the organization, often to accomplish some task" (Feldman, 2016, p. 24). However, these routines may be modified or new routines created due to exogenous or endogenous causes that bring about changes in actions and patterns of action that constitute a routine (Feldman, 2016, p. 24).

Organisational routines are stored in rules and procedures (de Boer & Zandberg, 2012) or performed through action (Feldman, 2016, p. 30). In this study, the actions and patterns of action present an observable unit of analysis to understand routine dynamics (Feldman, 2016, pp. 26–27). As the COVID-19 pandemic was unexpected and disruptive in nature, this study is especially interested in how the exogenous turbulence influenced activities in practice and therefore induced adaptation of routines. In other words, the necessity to respond to the COVID-19 pandemic may have led to a change of focus in certain public sector organizations.

The study focuses on the performative aspect of routines and explores the actions that constituted organizational routines during the pandemic: how civil servants used data to provide a pandemic response. To analyze these changes, a framework (Figure 1) is proposed that synthesizes literature on dynamic capabilities, data capabilities, and their role in contributing to changes in routines. Building mainly on the work of Conboy et al. (2020), Teece (2007), and Piening (2013), it explores the impact of data capabilities on inter-organizational routines. In other words, the framework enables the description of data capabilities as an input that supports modifications of organizational routines through enabling changes in actions and patterns of action.

Some scholars refer to environmental dynamism as "the amount of uncertainty, complexity, and change

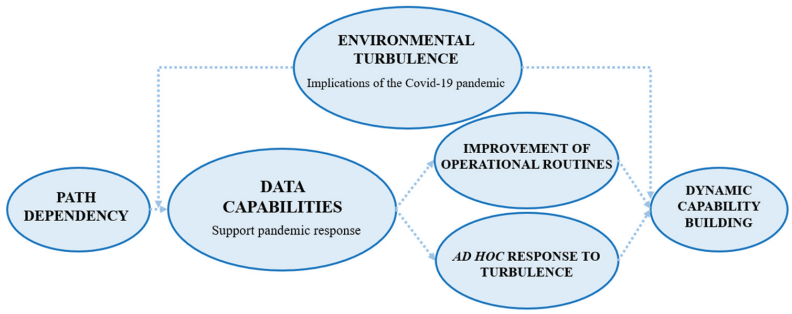


**Figure 1.** Framework for analysis (author, based on Conboy et al., 2020; Piening, 2013; Teece, 2007).

emanating from the external environment" (Chirico & Bau,' 2014, p. 5). **Environmental turbulence** in the analytical framework refers to the conditions brought about by the COVID-19 pandemic, also referred to as "the turbulence." Data, analytics, and forecasting support government policy and decision-making and facilitate understanding of the impact of their decisions (Ågerfalk et al., 2020). Additionally, such turbulent conditions demanded unusually fast decisions that had an impact on specific groups in society who were impacted by the pandemic and the accompanying restrictions (Fotiadis et al., 2021). Although the unexpected need to provide a response to the pandemic required public sector organizations to adapt new circumstances, the temporary nature of turbulence may not provide enough incentive to produce new routines in public sector organizations. Furthermore, given that established routines go through an adaptation process until they stabilize (Razak, 2017, p. 12), turbulence may not provide the necessary timeframe for new routines to become stable and repetitive.

The activities undertaken by public sector organizations are determined by existing routines that have been developed through **path dependency**. A wide range of studies are concerned with the antecedents and existing capabilities necessary for data-driven decision-making (Mikalef et al., 2017; R. Sharma et al., 2014). Organisational response to turbulence is dependent upon a combination of environmental influences and existing capacities to address related challenges.

To cope with a turbulent environment, **data capabilities** are considered to be higher-order routines (Winter, 2003) that serve as input for modifying operational routines. They contribute to an organizational response to exogenous turbulence through three factors of dynamic capabilities: *sensing* the surrounding environment to identify and assess opportunities, *seizing* those opportunities through mobilization of resources, and implementing *transformational* activities to ensure continuous renewal of the organization (Conboy et al., 2020; Teece, 2007). However, Wilhelm et al. (2022) argue that these three factors may differ in their routineness in organizations: the frequency of their execution and structuring (existence of rules, formal procedures, or standards). For example, they argue that while organizations may frequently apply dynamic capabilities to keep up with their surroundings, their sensing, seizing, and transforming activities may not be highly structured (Wilhelm et al., 2022). Instead, they may perform "experiential activities such as prototyping, real-time information, multiple options, and experimenting" to gain new information (Eisenhardt & Martin, 2000).

In this study, data capabilities are interpreted as follows:

- In turbulent conditions, such as the COVID-19 pandemic, *sensing* capabilities support the organization's ability to detect changes in the organization or its environment, and therefore, help them to adapt proactively (Duchek, 2014). Good sensing capabilities enhance understanding of how the turbulent environment could influence the organization's objectives, target groups, and other main stakeholders. Also, sensing capabilities can utilize internal and external data to gain new knowledge (Konopik et al., 2021) about the surrounding environment and include activities that are directed toward scanning the surroundings. For example, such data processes are related to understanding the conditions of target groups, spotting the changing needs of service users, or interpreting changes and uncertainties (Conboy et al., 2020).
- *Seizing* capabilities support finding and implementing a solution (Duchek, 2014) through reconfiguring resources to provide a response to unexpected events. In other words, organizations develop ways to respond to observed changes in

their environment. For example, underlying data processes are related to building competencies, decision-making, or mobilizing resources to address opportunities (Conboy et al., 2020), but also providing *ad hoc* solutions to emerging challenges.

- *Transforming* capabilities refer to continuous changes in operational routines. Following the distinction made by Zollo and Winter (2002), this analytical model proposes that data capabilities influence the performance of an organization indirectly by enabling more efficient, effective, or improved operational routines. Wenzel et al. (2021), p. 20) argue that "organisational routines . . . form a key component to explain how dynamic capabilities contribute to organisational change." This could be achieved through disrupting existing routines, e.g., with a new technology or work practice (Piening, 2013). Also, building dynamic capabilities is dependent upon agency: the intention to change organizational *status quo* and the level of influence of an individual who attempts to do so (Kurtmollaiev, 2020). In this phase, underlying processes are related to reconfiguring capabilities, re-engineering processes, or managing knowledge (Conboy et al., 2020).

In this study, **building dynamic** capabilities is understood to be an improvement in the way activities are carried out in public sector organizations, thus referring to an intrinsic change in existing routines. Stability and change of routines have been widely discussed by researchers (Feldman, 2016, pp. 26–27; Geiger, 2022). Research on routines has shifted from a characterization of routines as "objects" to a generative approach that argues that routines are continuously emerging and "characterised by internal structures and dynamics" (D'adderio, 2008). Exogeneous turbulences may influence the dynamics of routines through forcing change in usual patterns of action. However, not all changes of actions made under the pressure of the pandemic lead to transformation. To understand the extent and nature of the changes in routines studied, three aspects were considered: (1) changes in patterns of action compared to pre-COVID patterns; (2) the change in the frequency of using certain data and data capabilities as an input for organizational routines; and (3) the continuity of the routine adaptations induced by the turbulence. Understanding the extent and continuity of routine adaptations during turbulent conditions reveals data capabilities that have a potential to transform organizational routines for future resilience, as well as increase efficiency through small process adaptations to respond to changing environmental conditions.

However, whether dynamic capabilities are created in the long term is related to the costs and benefits of *ad hoc* problem solving (Winter, 2003) meaning that developing and maintaining dynamic capabilities is rather costly and providing temporary solutions to challenges may be more efficient. In addition, Wilhelm et al. (2015) have argued that when considering the costs of increased effectiveness, dynamic capabilities only pay off under high levels of environmental dynamism.

The proposed framework allows deeper insight into how data capabilities influence existing (inter-organizational) routines and whether the change they support during turbulence is continuous.

## Methodology

Eriksson (2013) argues that dynamic capabilities cannot be identified *ex ante*, which means that in empirical studies, careful sampling is necessary as it is difficult to spot organizations that would be interesting to study. Exploring the data capabilities of the Estonian Tax and Customs Board (ETCB) can be considered interesting for several reasons. Estonia is seen as being a successful e-state (Margetts & Naumann, 2017; Tammpuu & Masso, 2018) and a global leader in digital government with its electronic ID card system and secure data exchange architecture X-Road (Kattel et al., 2020). The digital revolution that has taken place over the recent decade has made it possible and affordable to capture large amounts of data (Shuradze & Wagner, 2016), indicating that high-level technological capacity supports deploying data capabilities. Being one of the most advanced Estonian public sector organizations in terms of data analytics, the ETCB provides a good starting point for discussion of the interrelations of data capabilities and dynamic capabilities in public sector organizations. The ETCB collects valuable data for assessing the economic conditions and well-being of citizens, such as information about incomes, turnovers, tax receipts, and tax arrears. This data provides public sector institutions with valuable insights into the labor market and was used for the response to the COVID-19 pandemic.

The current study uses the research strategy of the case study together with the explanatory case study method. A single-case design was chosen that explores the role of data capabilities in improving public sector operational routines. Two routines were inter-organizational, meaning that the data that the ETCB provided was used to perform tasks related to pandemic

response in other public sector organizations. Specifically, changes in the actions and patterns of action of three routines were examined:

- **Routine 1**. **Allocating financial support to vulnerable groups in society**. The ETCB provided data for institutions responsible for allocating financial support to businesses and employees that were impacted by the pandemic or providing loans to businesses in vulnerable conditions. The ETCB data was used when considering different alternatives of support measures, evaluating the applicant's compliance with the application criteria, and assessing risks and tax compliance of the applicants so that the financial support would go to the most vulnerable groups.
- **Routine 2**. **Evaluating the economic impact of the pandemic**. In order to develop supportive measures, legislative authorities (mainly the Ministry of Economic Affairs and Communications) required information that could help assess the conditions of different business sectors and to understand the economic impact of the pandemic, e.g., the impact of movement constraints and closed borders.
- **Routine 3**. **Providing management information to the board of the ETCB**. The board of the ETCB needed more immediate and systematic information to sense the environment: to evaluate the conditions of the economy and taxpayers to understand whether an abrupt decline in tax receipts could be expected. Before the pandemic, analysts prepared overviews of tax receipts and related indicators once a month, but the needs and expectations for the Management Information System changed during the pandemic. Also, the ETCB itself monitored tax arrears data to identify vulnerable taxpayers to decide whether there would be a need for decreased interest on tax arrears and how this decision would affect tax revenues. There were three main implications: the need for information to be provided closer to real-time (e.g., changes in turnovers, declared taxes, arrears), for information to be more predictive (e.g., macroeconomic prognosis), and more regular (e.g., weekly updates).

Two data collection methods were used. First, a *document analysis* was conducted of the strategic documents and data exchange contracts of the ETCB. Also, news media articles and web pages of related institutions were used to gather information about their role during the pandemic. Document analysis provided background information and deeper understanding of the extent and ways in which the COVID-19 pandemic influenced the processes under investigation but also offered insight into the outcomes and challenges of data-related processes during the pandemic.

Second, *qualitative interviews* were conducted with civil servants as the main data collection method (Table 1). First, interviews were carried out with the civil servants of the ETCB—including leaders concerned with data capabilities and specialists in positions that were closely related to resolving the challenges related to responding to the COVID-19 pandemic. After identifying the routines that were most influenced by the ETCB data and data capabilities (Table 2), civil servants from three other institutions were interviewed in order to understand the ways in which these capabilities influenced the work of partner organizations. Specifically, interviews were carried out with civil servants from the Estonian Unemployment Insurance Fund, Enterprise Estonia, and the Ministry of Economic Affairs and Communications, who were responsible for processes that used ETCB data during the pandemic. In total, interviews were conducted with nine civil servants. Interview guides were developed following the analytical framework proposed in Section 2. However, the interview guides were modified to fit the interviewees' specialties and areas of responsibility.

The interviews were recorded and transcribed. To reduce bias, multiple data collection methods were used (document analysis and interviews), and the transcripts of the interviews and the results of the study were discussed with the interviewees. Interview transcripts

**Table 1.** Overview of the interviews.

| No | Organisation | Date | Position of the interviewee | Interview code |
|---|---|---|---|---|
| 1 | Estonian Tax and Customs Board | July 14 2021 | Head of Risk Management | 1A |
| 2 | Estonian Tax and Customs Board | July 21 2021 | Unit Chief | 2A |
| 3 | Estonian Tax and Customs Board | July 23 2021 | Chief specialist | 3A |
| 4 | Estonian Tax and Customs Board | August 2 2021 | Chief specialist | 4A |
| 5 | Estonian Tax and Customs Board | August 2 2021 | Chief specialist | 5A |
| 6 | Estonian Tax and Customs Board | August 4 2021 | Chief specialist | 6A |
| 7 | Estonian Unemployment Insurance Fund | March 1 2022 | Deputy Unit Chief | 7B |
| 8 | Enterprise Estonia | March 3 2022 | Unit Chief | 8C |
| 9 | Ministry of Communications and Economic Affairs | March 23 2022 | Unit Chief | 9D |

Table 2. The routines under study and interviews conducted.

| No | Routine/process | Interviews related to the routine |
|---|---|---|
| 1 | Allocating financial support to vulnerable groups in society | 1A; 2A; 3A; 4A; 7B; 8C |
| 2 | Evaluating the potential impact of restrictions | 9D; 2A |
| 3 | Providing management information to the board of the ETCB | 2A; 5A; 6A |

were analyzed using qualitative content analysis and the NVivo program. The study used the deductive approach to coding as the categories were established based on the analytical framework proposed in the theoretical section of the article. Thus, six categories were developed: (1) environmental turbulence, (2) data capabilities, (3) *ad hoc* response, (4) improvement of operational routines, (5) path dependency, (6) background information. The transcripts of the interviews were coded, and the entire text was divided into those six categories.

## Results

This section gives an overview of the impact of the COVID-19 pandemic on the public sector routines studied. Also, the ways organizations responded to emerging challenges and implications of these modifications for dynamic capability building are highlighted. Table 3 (see Supplementray Materials Appendix 1) outlines the main results of the empirical study that are elaborated on in greater detail.

The changes indicated in the interviews could be divided into three types: *ad hoc*, long-term, and potential changes. *Ad hoc* refers to temporary changes to cope with the pandemic and no evidence was found that the solutions were used after the turbulence. Long-term changes imply continuous changes in the actions of operational routines. In addition, the interviewees pointed out the potential for long-term changes that were acknowledged but not yet acted upon.

### Implications of the COVID-19 pandemic for the routines studied

The interviewees pointed out that during the pandemic, they experienced a sudden increased need for data to make data-based decisions. From the ETCB's point of view, the need for data from three sources grew significantly and helped to evaluate the conditions of the workforce and labor market. These three sources were: (1) value added tax return forms; (2) declarations of income and social tax, unemployment insurance premiums, and contributions to mandatory funded pension; (3) employment registration system.

#### Routine 1

***Implication of exogenous turbulence: novel target groups.*** When it comes to providing financial support to vulnerable groups, two organizations faced the need to perform tasks in novel ways. For instance, the Estonian Unemployment Fund had not previously allocated financial support to businesses (previously their stakeholders were only employees) or used turnover data in their existing routines (Routine 1; Interview 7B).

***Implication of exogenous turbulence: volume of applicants.*** As described previously, public organizations faced a sudden increase in applicants for financial support. For example, while Enterprise Estonia (Routine 1, Interview 8C) had been examining individual financial aid applicants manually, they switched to mass decisions and handled hundreds of applicants simultaneously. This was made possible by increased inter-organizational data exchange and automatic data models that were developed in close collaboration with the ETCB.

In Routine 1, the capabilities of making mass decisions and providing services to novel target groups were temporary solutions to cope with the pandemic. Although the capacity to switch to making mass decisions and knowledge related to pandemic-related target groups exists in the long-term, it could be considered an *ad hoc* change as there was no need for related actions to continue afterward. Also, data exchange channels that were related to such automation (Routine 1) or were used to understand the implications of turbulence (Routine 2) did not continue to be used after pandemic-related tasks ended.

***Implication of exogenous turbulence: remoteness.*** Interviewees pointed out that during the pandemic, they were constantly adapting to the specific working conditions and influences of COVID-19, especially working from home, and promoting e-channels to reduce physical contact. Interviewees outlined that the pandemic promoted the usage of e-channels, and therefore, basic capabilities to provide e-services proved to be useful during the pandemic. When the volume of applicants grew significantly, the manual handling of applications would have been too slow and burdensome (Routine 1, Interview 7B).

One of the impacts of the pandemic was a leap toward working remotely and using e-channels in providing public services to customers (including applications for financial support or providing consultation) (Routine 1, Interview 7B). For example:

> *We have different services that we offer to our clients. Providing them remotely was quite rare* [before the pandemic]. *Now, we have actually reached a point where we have less eye-to-eye meetings and this is irreversible.* (Routine 1, Interview 7B)

The interviewees mentioned that tasks that needed to be performed during the pandemic were completed more easily when there was already an existing capacity (including experience and competencies) to perform similar tasks. Those include capabilities to integrate risk management into processes (Interview 1A) and provide e-services (Interview 7B). This indicates that the level of certain existing capabilities may define how challenging it is to provide pandemic response in organizations.

***Implication of exogenous turbulence: increased workload led to workload dispersion.*** To cope with the sudden increase in workload, one interviewee pointed to their actions toward dispersing workloads and simpler tasks among the civil servants of the organization (Routine 1, Interview 8C). An interviewee from another organization outlined the limitations of dispersing tasks and viewed it as one of the challenges they faced during that period (Routine 1, Interview 2A). Their usual approach of encouraging in-depth specialized knowledge among their specialists created a situation where the civil servants responsible for certain areas had an uneven and accelerated workload and were therefore more prone to burnout. Also, interviewees indicated that during the crisis, it could be very challenging to train new people to the extent that was often needed (Interview 5A).

Furthermore, the ability to disperse workloads was mentioned as crucial to operating in turbulent conditions. When the first wave of the pandemic hit, only a few experts of the ETCB data sources were involved in responding to data needs in the studied routines. This put a large amount of pressure on the civil servants responsible for interpreting this data, as well as dealing with the specifics of taxation and business. As time passed, some knowledge necessary for performing those tasks was dispersed and the need to continue with this approach was acknowledged (Routines 1, 2, Interview 2A).

### Routine 2

***Implication of exogenous turbulence: increased workload led to more open data.*** There were examples of using open data to cope with the increased enquiries and information requests from stakeholders. For example, one of the interviewees pointed out that to reduce the number of single enquiries, the data of fuel sales was made available on the ETCB homepage and was continuously used by stakeholders (Routine 2, Interview 2A). This data was continuously provided, and therefore, a significant change toward providing more open data could be detected.

***Implication of exogenous turbulence: the need for more detailed data led to deeper collaboration.*** The data that was usually provided to public sector organizations was not sufficiently detailed to meet the needs that arose during the pandemic which brought specific data needs. For example:

> *There was a need for data that we do not really get from Statistics Estonia with such precision, and then all this cooperation with the ETCB became important. To get such customised data and the data we needed to get a general overview and to justify the need for* [financial] *support . . . And second, we used data models to specifically calculate possible volumes of support under certain conditions, to develop possible support schemas, and assess their pros and cons and costs.* (Routine 2, Interview 9D)

Enterprise Estonia had to develop measures to support many small tourism-related enterprises that were not usually their target groups (Routine 1, Interview 8C). The main challenge was to find sufficiently detailed information for decision-making from the available data: the pandemic mostly influenced sectors related to tourism, which made knowledge about this sector especially necessary.

Therefore, the organizations studied faced the need to gain in-depth knowledge (e.g., technical specifics, data structure, combined with expert knowledge of tax legislation) in areas they had not previously encountered. This led to close collaboration between partners to exchange data and develop supportive measures (Routine 1) to combine the knowledge necessary for completing the task. In one case, data exchange with the ETCB had not occurred before the pandemic, which provides evidence of new data exchange channels motivated by the changed conditions (Routine 1, Interview 8C). This implies that although sharing data and providing input for decision-making can be described as a routine in ETCB, it has provided other public sector organization with novel means to improve their organizational routine.

***Implication of exogenous turbulence: the need for fast decision-making and policymaking.*** During the period of crisis, the extraction of information and knowledge from data was accelerated due to a need to make speedy decisions and an increased pressure to adapt to changed

conditions and priorities. However, this was the case only during the pandemic.

### Routine 3

***Implication of exogenous turbulence: monitoring economic conditions.*** One interviewee gave an example of the way increased workloads motivated them to automate an existing, non-crisis-related routine:

> *But it was during COVID that overviews of tax collection were needed more often, and I saw that I would be doing this for the whole month. . . . And I just took three days, and I made this thing more efficient for myself so that next time I would get this data more quickly. . . . I made my tasks more efficient, and I am still using it this way.* (Routine 3, Interview 4A)

***Implication of exogenous turbulence: increased interest in tax receipt indicators.*** To cope with the sudden need to understand economic conditions, more frequent presentations of data and simple dashboards were developed in the ETCB (Routine 3).

Moreover, the usual monthly overviews of tax receipts that included indicators with a two-month delay were not sufficient to make confident and up-to-date claims about the surrounding environment (Interview 5A) and constant monitoring of macroeconomic indicators became a part of the pandemic era routine (Interview 2A). As one interviewee put it:

> [When talking about the increased interest of the board of the ETCB in monitoring macroeconomic conditions during the pandemic] *I think it* [overview of macroeconomic indicators] *was* [usually] *made available when asked. It was not systematically monitored or observed by county. Occasionally, there were county tours, and the board was talking about them. But by default, it was not available at all times in the background.* (Routine 3, Interview 2A)

Automatic management reports were developed in Routine 3 so that the board of the ETCB could monitor in real time the necessary changes in relevant data. This indicates that the capability to automate management information provides organizations with a means to understand and respond quickly to changing circumstances. Although some of the automatic management reports are not used today, they could be activated on demand—leading to a long-term possibility of monitoring the economic situation should the need for rapid decision-making and monitoring of changing conditions arise again (Routine 3, Interview 5A). Therefore, even though the need for such information and dashboards was temporary, the capabilities to develop these on demand during the next turbulence may increase the organization's resilience.

In conclusion, the scale of data usage was diverse—from general information that was used to evaluate the economic conditions of certain sectors to detailed information that could be used in specific decisions and policymaking. Some data capabilities were outlined that helped organizations cope with the challenges of the COVID-19 pandemic, such as the capabilities to automate processes, provide open data, and build simple representations of data. Also, the ability to disperse data-related knowledge and workloads was one of the key solutions to cope with increased expectations.

## Discussion

This article explored whether and how data capabilities that were used for the COVID-19 response supported the adaptation of public sector routines during the pandemic resulting in the development of dynamic capabilities in the public sector. The results of the study indicate that during the pandemic, public sector organizations faced challenges that impelled rapid adaptation of routines. It was possible by applying certain data capabilities that enabled organizations to modify their existing routines.

The results section outlined some data capabilities that enabled public sector organizations to provide pandemic response. These capabilities serve as means to develop public sector dynamic capabilities that support rapid routine adaptations in turbulent conditions:

- *Ability to automate processes and dataflows that support decision-making.* During the pandemic, organizations used automation in different routines to monitor the environment, to gain understanding of the conditions of target groups, but also to evaluate different alternatives of action and support the implementation of the solutions chosen. In other words, changes in agency (from people to machines) proved to support routine adaptation. However, it was pointed out that automation presupposes high data quality and the availability of specific detailed data. In addition, this presumes timely and clear agreements on what is included in data sets, often requiring specific knowledge, which suggests the importance of collaborative capability (Mayne et al., 2020) for public sector organizations involved in the routine. The key is to build flexible systems that support the ability to develop quick solutions that offer deep insight into the areas influenced by the crisis, but also enable secure exchange of data between public institutions.
- *Ability to provide open data was one way to cope with the increased workload and high volume of*

*enquiries from various stakeholders*. Open data helped organizations to cope with rapidly changing and turbulent conditions. Although public organizations differ in their capacities and actions in providing open data, this ability could provide them with a means to focus on other important tasks made urgent by the turbulence, thus increasing their ability to provide a response and adapt to the implications of crises.

- Producing simple graphs and overviews of pandemic-related data was mentioned as one of the biggest challenges in providing data for decision-making. *The ability to develop simple representations of data and combine this knowledge with general macroeconomic trends* provides twofold value. First, it reduces complexity and encourages in-depth understanding among stakeholders that leads to informed decision-making. Second, given that the pandemic led public organizations to use data they had not used before, simplicity makes it easier to interpret data and increase knowledge about what it entails and represents.
- It was pointed out that the *ability to disperse workload* helped to cope with the pandemic. Building on the experience of the data analysts of the ETCB, the specific specialization of civil servants may become a hindrance during times of turbulence. Instead, the ability to disperse simpler tasks and data-related knowledge may reduce the probability of individual burnout and support the capacity to deal with increased workload. This could be achieved through well-designed knowledge-sharing processes and documentation about proprietary data and data models.
- In addition, the case demonstrated that the pandemic prompted some changes in organizational settings, such as a shift toward using e-services and working from home. However, this article does not explore these implications in more detail as they are only vaguely related to the data capabilities of organizations.

The study demonstrated that in the public sector, routine data capabilities of one organization may serve as temporary input for other organizations' routines in turbulent times, thereby determining the success of the pandemic response. However, routine adaptations were closely related to tasks brought about by the pandemic. This indicates that less evidence was found of routine adaptations being continuous after the turbulence. In other words, routine data capabilities enabled mostly temporary adaptations of routines. Routine performances are deeply related to the organizational context, which influences the flexibility of routines' use (Howard-Grenville, 2005). After turbulence, for change in public sector routines to continue when enabled through deploying dynamic capabilities could be considered complex and dependent upon three organizational factors: (1) legislative environment, (2) agency, and (3) path dependency.

(1) **Legislative environment surrounding public sector organizations**. As previously outlined, the data exchange channels that were developed during the pandemic were relatively *ad hoc* and the necessary legislation was developed quickly to enable additional inter-organizational use of data. This sets important limitations on the possibilities of using the data for non-pandemic-related tasks. Specifically, using this data in other routines presupposes the development of further legal grounds (Interviews 1A, 7B, 8C). For example, data used for modifying COVID-19 related routines could not be used in other operational routines without creating the necessary legislative basis, although the potential for improving surveillance processes was outlined (Routine 1, Interview 1A).

(2) Interviewees pointed out that activities related to pandemic response provided important and useful lessons. However, although the potential for improving other organizational routines based on the experience of pandemic response was outlined, the turbulence provided a certain "state of urgency" that dissolved afterward. Future changes are dependent upon **agency**: willingness, motivation, and incentives to change organizational routines. Howard-Grenville (2005) indicates that "individuals and groups approach routines with different intentions and orientations" that "shape particular routine performances."In addition, the results indicate that shifting tasks from people to machines provides considerable benefits, including developing automated solutions or using modern technologies to perform tasks.

(3) The current case study demonstrated that **path dependency** or "history" (Pentland et al., 2012) influenced the way pandemic response was provided in different organizations. The interviewees mentioned that tasks that were needed during the pandemic were more simply executed when there was already an existing capacity (including experience and competences) to execute similar tasks. Those included the capability to integrate risk management practices into

processes (Interview 1A) and provide e-services (Interview 7B). This indicates that the level of certain existing capabilities may define how challenging it is for an organization to provide a pandemic response. However, sometimes crisis creates conditions for overcoming intertia; in unexpected conditions, organizations may approach challenges through a learning-by-doing mind-set (Routine 1, Interview 3A).

The COVID-19 pandemic provided a fruitful ground to explore dynamic capability building in the public sector in turbulent conditions. Although the pandemic was specific in nature and may not provide a step-by-step guide to developing dynamic capabilities, it outlined some data capabilities that enable organizational routines to adapt in response to uncertain and unexpected conditions. Although applying dynamic capabilities during turbulence may determine the success of crisis response, the changes they support in routines may not be continued afterward due to the temporary nature of turbulence, limitations of legislative conditions, dependence on agency, or previous knowledge and expertise in organizations.

## Conclusion

This article explored the ways in which turbulent conditions motivate changes in public sector organizations' existing routines and the role of data capabilities in supporting routine adaptations. The study focused on the case of Estonian public administration, exploring the ways in which data capabilities contributed to the pandemic response through the improvement of three (inter-)organizational routines. As the case study demonstrated, the proposed analytical framework proved to be valuable in structuring the empirical analysis and tracing the mechanisms by which data capabilities support the development of public sector dynamic capabilities. Hence, the proposed framework could provide a fruitful starting point for future studies examining the ways in which dynamic capabilities are developed.

Two research questions were proposed. First, the study explored the value of ETCB data and data capabilities in responding to the COVID-19 pandemic and the ways they supported change in existing operational routines. Second, it aimed to understand whether and under what circumstances data capabilities support the development of public sector dynamic capabilities. The case study demonstrated that many of the changes that occurred in the studied routines were motivated by the need to provide specific responses to the pandemic. This means that when the pandemic was over, the "sense of urgency" ended and the need for such measures often dissolved as well, raising the question of the "continuity" of routine adaptations. However, even though some measures taken in response to the pandemic were inherently temporary, they provided the civil servants with new knowledge and expertise to develop similar solutions in the future. In other words, the turbulence increased the capacity to develop on-demand solutions. Considering that public sector organizations operate in increasingly unpredictable and turbulent conditions, investing in the development of data capabilities that enable successful *ad hoc* solutions to be offered in case of turbulence could entail a substantial contribution to building public sector dynamic capabilities.

There are several limitations to the current study. First, this is a small-scale case study that describes the experience of the public sector organizations of a small state. Further studies should explore whether the challenges and lessons from COVID-19 that the Estonian public sector organizations faced could be generalized to other (and larger) public administrations. Also, the impact of the COVID-19 pandemic on tax compliance and transparency through open data could be topics of wider interest in the future since they reflect the value created through the extensive use of data. Consequently, to understand whether the conclusions drawn in this paper function more broadly, further insight should be obtained through more extensive case studies.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Estonian Research Council grant PRG1125.

## References

Ågerfalk, P. J., Conboy, K., & Myers, M. D. (2020). Information systems in the age of pandemics: COVID-19 and beyond. *European Journal of Information Systems*, *29* (3), 203–207. https://doi.org/10.1080/0960085X.2020.1771968

Ambrosini, V., Bowman, C., & Collier, N. (2009). Dynamic capabilities: An exploration of how firms renew their resource base. *British Journal of Management*, *20*(s1), S9–S24. https://doi.org/10.1111/j.1467-8551.2008.00610.x

Chirico, F., & Bau', M. (2014). Is the family an "asset" or "liability" for firm performance? The moderating role of environmental dynamism. *Journal of Small Business Management*, *52*(2), 210–225. https://doi.org/10.1111/jsbm.12095

Conboy, K., Mikalef, P., Dennehy, D., & Krogstie, J. (2020). Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *European Journal of Operational Research*, *281*(3), 656–672. https://doi.org/10.1016/j.ejor.2019.06.051

D'adderio, L. (2008). The performativity of routines: Theorising the influence of artefacts and distributed agencies on routines dynamics. *Research Policy*, *37*(5), 769–789. https://doi.org/10.1016/j.respol.2007.12.012

de Boer, E. A., & Zandberg, T. (2012). Decision-making by organizational routines: The influence of agency and personality traits on deviating from formal organizational routines. *Journal of Quality Assurance in Hospitality & Tourism*, *13*(4), 316–325. https://doi.org/10.1080/1528008X.2012.692287

Duchek, S. (2014). Growth in the face of crisis: The role of organizational resilience capabilities. *Academy of Management Proceedings* (No. 1, p. 13487). Academy of Management, Briarcliff Manor, NY 10510.

Edwards, F. L., & Ott, J. S. (2021). Governments' responses to the COVID-19 pandemic. *International Journal of Public Administration*, *44*(11–12), 879–884. https://doi.org/10.1080/01900692.2021.1936964

Eisenhardt, K., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, *21*(10–11), 1105–1121. https://doi.org/10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E

Eriksson, T. (2013). Methodological issues in dynamic capabilities research – a critical review. *Baltic Journal of Management*, *8*(3), 306–327. https://doi.org/10.1108/BJOM-Jul-2011-0072

Feldman, M. (2016). Routines as process: Past, present, and future. In J. A. Howard-Grenville, C. Rerup, A. Langley, & H. Tsoukas (Eds.), *Organizational routines: How they are created, maintained, and changed* (pp. 23–46). Oxford University Press.

Feldman, M., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*(1), 94–118. https://doi.org/10.2307/3556620

Fotiadis, A., Polyzos, S., & Huan, T. C. T. (2021). The good, the bad and the ugly on COVID-19 tourism recovery. *Annals of Tourism Research*, *87*, 103117. https://doi.org/10.1016/j.annals.2020.103117

Geiger, D. (2022). Routine dynamics and paradox: A revised research agenda. *Scandinavian Journal of Management*, *38*(4), 101244. https://doi.org/10.1016/j.scaman.2022.101244

Haarhaus, T., & Liening, A. (2020). Building dynamic capabilities to cope with environmental uncertainty: The role of strategic foresight. *Technological Forecasting & Social Change*, *155*, 120033. https://doi.org/10.1016/j.techfore.2020.120033

Howard-Grenville, J. A. (2005). The persistence of flexible organizational routines: The role of agency and organizational context. *Organization Science*, *16*(6), 618–636. https://doi.org/10.1287/orsc.1050.0150

Jantunen, A., Tarkiainen, A., Chari, S., & Oghazi, P. (2018). Dynamic capabilities, operational changes, and performance outcomes in the media industry. *Journal of Business Research*, *89*, 251–257. https://doi.org/10.1016/j.jbusres.2018.01.037

Kattel, R., Lember, V., & Tõnurist, P. (2020). Collaborative innovation and human-machine networks. *Public Management Review*, *22*(11), 1652–1673. https://doi.org/10.1080/14719037.2019.1645873

Klöckner, M., Schmidt, C. G., Wagner, S. M., & Swink, M. (2023). Firms' responses to the COVID-19 pandemic. *Journal of Business Research*, *158*, 113664. https://doi.org/10.1016/j.jbusres.2023.113664

Konopik, J., Jahn, C., Schuster, T., Hoßbach, N., & Pflaum, A. (2021). Mastering the digital transformation through organizational capabilities: A conceptual framework. *Digital Business*, *2*(2), 100019. https://doi.org/10.1016/j.digbus.2021.100019

Kurtmollaiev, S. (2020). Dynamic capabilities and where to find them. *Journal of Management Inquiry*, *29*(1), 3–16. https://doi.org/10.1177/1056492617730126

Lember, V., Kattel, R., & Tõnurist, P. (2018). Technological capacity in the public sector: The case of Estonia. *International Review of Administrative Sciences*, *84*(2), 214–230. https://doi.org/10.1177/0020852317735164

Margetts, H., & Naumann, A. (2017). *Government as a platform: What can Estonia show the world. Research paper. University of Oxford.* https://www.politics.ox.ac.uk/publications/government-as-a-platform-what-can-estonia-show-the-world.html (Accessed 01.08.2021.

Mayne, Q., De Jong, J., & Fernandez-Monge, F. (2020). State capabilities for problem-oriented governance. *Perspectives on Public Management and Governance*, *3*(1), 33–44. https://doi.org/10.1093/ppmgov/gvz023

Mero, J., & Haapio, H. (2022). An effectual approach to executing dynamic capabilities under unexpected uncertainty. *Industrial Marketing Management*, *107*, 82–91. https://doi.org/10.1016/j.indmarman.2022.09.021

Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J., & Olsen, D. (2017). *Big data analytics capability: Antecedents and business value.*

Nuhu, N. A., Baird, K., & Appuhami, R. (2019). The impact of management control systems on organisational change and performance in the public sector: The role of organisational dynamic capabilities. *Journal of Accounting & Organizational Change*, *15*(3), 473–495. https://doi.org/10.1108/JAOC-08-2018-0084

Panagiotopoulos, P., Klievink, B., & Cordella, A. (2019). Public value creation in digital government. *Government Information Quarterly*, *36*(4), 101421. https://doi.org/10.1016/j.giq.2019.101421

Pentland, B. T., Feldman, M. S., Becker, M. C., & Liu, P. (2012). Dynamics of organizational routines: A generative model. *Journal of Management Studies*, *49*(8), 1484–1508. https://doi.org/10.1111/j.1467-6486.2012.01064.x

Piening, E. P. (2013). Dynamic capabilities in public organizations: A literature review and research agenda. *Public Management Review*, *15*(2), 209–245. https://doi.org/10.1080/14719037.2012.708358

Razak, A. R. A. (2017). *An investigation into the micro-dynamics of routine flexibility*. Lancaster University (United Kingdom).

Sharma, A., Borah, S. B., & Moses, A. C. (2021). Responses to COVID-19: The role of governance, healthcare infrastructure, and learning from past pandemics. *Journal of Business Research*, *122*, 597–607. https://doi.org/10.1016/j.jbusres.2020.09.011

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), 433–441. https://doi.org/10.1057/ejis.2014.17

Shuradze, G., & Wagner, H. T. (2016). Towards a conceptualization of data analytics capabilities. In T. X. Bui & R. H. Sprague, Jr. (Eds.), *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 5052–5064). IEEE.

Tammpuu, P., & Masso, A. (2018). 'Welcome to the virtual state': Estonian e-residency and the digitalised state as a commodity. *European Journal of Cultural Studies*, *21*(5), 543–560. https://doi.org/10.1177/1367549417751148

Teece, D. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, *28*(13), 1319–1350. https://doi.org/10.1002/smj.640

Teece, D., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

Wenzel, M., Danner-Schröder, A., & Spee, A. P. (2021). Dynamic capabilities? Unleashing their dynamics through a practice perspective on organizational routines. *Journal of Management Inquiry*, *30*(4), 395–406. https://doi.org/10.1177/1056492620916549

Wilhelm, H., Maurer, I., & Ebers, M. (2022). (When) are dynamic capabilities routine? A Mixed-Methods configurational analysis. *Journal of Management Studies*, *59*(6), 1531–1562. https://doi.org/10.1111/joms.12789

Wilhelm, H., Schlömer, M., & Maurer, I. (2015). How dynamic capabilities affect the effectiveness and efficiency of operating routines under high and low levels of environmental dynamism. *British Journal of Management*, *26*(2), 327–345. https://doi.org/10.1111/1467-8551.12085

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, *24*(10), 991–995. https://doi.org/10.1002/smj.318

Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science*, *13*(3), 339–351. https://doi.org/10.1287/orsc.13.3.339.2780

**Article III**

Raudla, R., Sarapuu, K., Vallistu, J., **Onno, K.**, & Harbuzova, N. (2025). The politics of experimental policymaking: the influence of blame avoidance and credit claiming. Policy Sciences, 58(2), 245–266.

Policy Sciences
https://doi.org/10.1007/s11077-025-09568-7

RESEARCH ARTICLE

# The politics of experimental policymaking: the influence of blame avoidance and credit claiming

**Ringa Raudla**[1] · **Külli Sarapuu**[1] · **Johanna Vallistu**[1] · **Kerli Onno**[1] · **Nastassia Harbuzova**[1]

Accepted: 24 January 2025


## Abstract

Policy experimentation has been proposed as a key strategy for coping with increasingly complex policy challenges. Despite considerable academic discussion on public policy experiments, there is a lack of systematic analyses of the political dimensions of policy experimentation. In this paper, we advance the understanding of politics of experimentation by analysing how policy actors' perceptions of blame avoidance and credit claiming influence experimental policymaking. As a theoretical contribution, we outline expectations about how the mechanisms of blame avoidance and credit claiming can influence policymakers' attitudes towards experiments and which contextual factors are likely to shape these dynamics. In the empirical part, we probe the plausibility of the theoretical propositions by using a comparative case study of Estonia and Finland. We draw upon policy documents and semi-structured interviews conducted with 66 public officials in Estonia and Finland in 2022–2023. Our empirical analysis demonstrates that the mechanisms of blame avoidance and credit claiming play a significant role in politicians' decisions to launch large-scale policy experiments. Our study also shows that these impacts are mediated by contextual factors like the urgency of policy problems, expected media reactions, public trust, and cumulative experience with policy experimentation.

**Keywords** Policy experiments · Blame avoidance · Credit claiming · Estonia · Finland · Experimental policymaking

✉ Ringa Raudla
ringa.raudla@gmail.com

Külli Sarapuu
kylli.sarapuu@taltech.ee

Johanna Vallistu
johanna.vallistu@taltech.ee

Kerli Onno
kerli.onno@taltech.ee

Nastassia Harbuzova
nastassia.harbuzova@taltech.ee

[1] Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Akadeemia 3, 12618 Tallinn, Estonia

Published online: 09 February 2025

## Introduction

Faced with increasingly complex economic, societal, and environmental problems, governments pursue novel ways to design and adapt public policies. Experimentation has been suggested as a key strategy for addressing complexity and gaining a better understanding of policy problems that entail high levels of uncertainty (e.g. Ansell & Bartenberger, 2016; Bravo-Biosca, 2020; Huitema et al., 2018; Mattocks, 2021; Nair & Howlett, 2016). Various studies refer to the growing use of experimentation in contemporary policymaking (e.g. Ansell & Bartenberger, 2016; Huitema et al., 2018; Lee & Ma, 2020; Mattocks, 2021; McFadgen & Huitema, 2018a, 2018b; Pearce & Raman, 2014).

The concept of public policy experiments has been defined in various ways. We rely on a definition of experiments as *time-limited tests of novel policy solutions that offer information for further policy decisions* (Bravo-Biosca, 2020; McFadgen & Huitema, 2018a, 2018b; Nair & Howlett, 2016; Raudla et al., 2023). By novel policy solutions we understand both those that are completely new (in the sense that they have not been tested elsewhere) but also solutions that are new for the specific country in terms of space and time. We focus on field experiments launched by policy actors and not on laboratory experiments conducted by researchers. In essence, such experiments constitute temporary policy interventions initiated by government organizations to test new ideas, instruments, solutions and to gain knowledge about their functioning in practice. Our focus is on large-scale public policy experiments that involve political decision-making.

Despite extensive academic debate and the use of different experimental approaches by governments, the existing literature on policy experiments has paid only limited attention to the *politics* of experimental policymaking and the role of politicians in launching the experiments (Huitema et al., 2018; Mattocks, 2021). Huitema et al., (2018, p. 148) argue that in the policy sciences, experiments tend to be conceptualized as *a research method*, and viewed "mostly in quite narrow, methodological terms". However, since, in practice, policy experiments take place in a political context, they are not neutral evidence-creating research activities (Brodkin & Kaufman, 2000; Huitema et al., 2018; Mattocks, 2021; Nair & Howlett, 2016). Political motives are likely to affect whether experiments get launched and how they are conducted and interpreted (Brodkin & Kaufman, 2000; Nair & Howlett, 2016; Sanderson, 2002). Despite the expectation that the political setting is likely to affect experimental policymaking, we currently have very limited knowledge on how political motivations affect the use of public policy experiments. While political motivations have been studied extensively in the literature dealing with the diffusion of policy innovations (see e.g. Berry & Berry, 2018 for an overview), they have received only limited attention in existing research on policy experimentation. In this article, we explore this important aspect of policy experimentation.

In order to contribute to a more systematic understanding of the *politics* of experimental policymaking, we use the theoretical lenses of blame avoidance and credit claiming to shed light on the motives of political actors regarding the use of policy experiments. In the literature on policy sciences, blame avoidance and credit claiming are regarded to be important drivers of politicians' decisions and behaviour (Bueno, 2023; Cruz & Schneider, 2017; Hinterleitner, 2017; Hinterleitner & Sager, 2015; Hood, 2002, 2010; Howlett, 2012; Leong & Howlett, 2017; Leong et al., 2023; McConnell, 2010; Weaver, 1986; Wenzelburger, 2014). The desire to avoid blame has been considered so influential in policymakers' behaviours that Hood (2010, p. 24) has called it an "imperative". Leong and Howlett (2017, p. 600) also observe that "the idea that credit claiming and blame avoidance are key

motivating forces of politicians and other policy-makers has become ubiquitous". Furthermore, as Hinterleitner and Sager (2015, p. 139) note, since policymaking is increasingly characterized by "the media induced politicization and scandalization of events, the pressure for public actors to engage in blame avoidance behaviour is on the rise".

Given that blame avoidance and credit claiming play an increasingly important role in policymakers' behaviours (Hood, 2002, 2010; Howlett, 2012; Leong & Howlett, 2017; Leong et al., 2023; Weaver, 1986), we would expect these motivations to influence experimental policymaking. So far, however, studies of experimental policymaking have *not* employed the theoretical lenses of blame avoidance and credit claiming for understanding how politics influences the use of policy experiments. Our paper addresses this gap. Our theoretical contribution to the literature on experimental policymaking is to juxtapose the theoretical insights of blame avoidance and credit claiming literature with insights from experimental policymaking literature and to discuss how the considerations of avoiding blame and claiming credit can influence experimental policymaking.

We probe the plausibility of the theoretical conjectures by examining experimental policymaking in two cases: Finland and Estonia. While both countries have similar political systems and embrace public sector innovation, their experiences with policy experiments vary considerably. Finland has become world famous for launching large-scale RCTs during the last decade (e.g. Leino & Akerman, 2022), while in Estonia, no large-scale policy experiments have been undertaken yet. Our research questions are: (1) How have the motivations of blame avoidance and credit claiming influenced experimental policymaking in those two countries? (2) Which contextual factors have influenced the blame avoidance and credit claiming motivations in public policy experiments? As sources of data, we use 66 interviews conducted with public officials in Estonia and Finland in 2022–2023. In both countries, we focus on the central (rather than subnational) government level. As existing studies have indicated, policy experimentation tends to be more prevalent in federal systems and on a subnational level (e.g. Oakley, 1998). Thus, our focus on the central government level in Finland and Estonia allows us to expand the understanding of how political motivations affect experimental policymaking on the national level in unitary systems.

The paper is structured as follows. Section two outlines the theoretical discussion, followed by the description of methods in section three and empirical analysis in section four. Section five provides a concluding discussion.

## Theoretical discussion

### The essence of public policy experiments

Experimental policymaking has been viewed as a strategic approach governments can use to meet the growing expectations to be more innovative and agile (e.g., Ansell & Bartenberger, 2016; Ettelt et al., 2015; Lee & Ma, 2020; McGann et al., 2018; Raudla et al., 2023). Experimentalist approaches to policymaking promise various benefits. Testing the viability of new policy solutions in a real-world setting before launching them on a larger scale is expected to increase the quality of policies and facilitate adaptations to change in an increasingly uncertain and complex policy environment (Ansell & Bartenberger, 2016; Bailey et al., 2017; Bravo-Biosca, 2020; Ettelt et al., 2015; Farrelly, 2008; Hughes et al., 2020; Lee & Ma, 2020; Lee et al., 2009; Millo & Lezaun, 2006; Nair & Howlett, 2016).

Governments can employ different types of policy experiments to fulfil these aims: randomized controlled trials (RCTs), non-randomized policy pilots, and design experiments (Raudla et al., 2023). RCTs entail random assignment of subjects into two or more groups and treating them differently in the hope of drawing valid causal conclusions about the treatment(s) being tested (Bravo-Biosca, 2020; Pearce & Raman, 2014). Non-randomized policy pilots try out a new policy solution on a small subset of a target population or jurisdictions before it is scaled up (Farrelly, 2008; Nair & Howlett, 2016). In design experiments, a policy solution is refined in an iterative way by gathering feedback from the affected individuals, until satisfactory results emerge (Ansell & Bartenberger, 2016; Stoker & John, 2009).

As such, public policy experiments hold the promise of a rational approach to policymaking (Checkland et al., 2023). In practice, they are always influenced by a specific political context (e.g. Bédécarrats et al., 2019; Brodkin & Kaufman, 2000; Mattocks, 2021; McGann et al., 2018). So far, however, scholarly discussions on experimental policymaking have paid only limited attention to the role of *politics* in experimental policymaking (Mattocks, 2021). We draw on the theoretical lenses of blame avoidance and credit claiming in order to contribute to a more systematic understanding of the *politics* of experimental policymaking.

## Key insights of the theory of blame avoidance and credit claiming in politics

A considerable body of literature in policy sciences has discussed how policy actors' decisions and behaviours are influenced by the wish to claim credit and avoid blame (e.g. Bueno, 2023; Cruz & Schneider, 2017; Hinterleitner, 2017; Hinterleitner & Sager, 2015; Hood, 2002, 2010; Howlett, 2012, 2014; Leong & Howlett, 2017; Leong et al., 2023; McConnell, 2010; McGraw, 1990; 1991; Weaver, 1986; Wenzelburger, 2014). It is expected that political actors believe that claiming credit earns them votes, while receiving blame hurts their electoral chances (Hansson, 2018; Hood, 2010; Howlett, 2012, 2014; Leong et al., 2023; Weaver, 1986). Hence, "they seek to avoid failures for which they can plausibly be held responsible and will claim credit for policy successes" (Leong & Howlett, 2017, p. 603).

The attribution of blame occurs if policy actors are perceived to be responsible for a decision or a policy that caused avoidable harm (Hood, 2002, 2010; Howlett, 2014; Leong & Howlett, 2017). Blame avoidance tactics include selecting policies that minimize the risk of blame, choosing institutional arrangements that help distribute responsibility and diffuse the blame, and employing rhetorical approaches and arguments as shields against blame (Hinterleitner, 2017; Hinterleitner & Sager, 2015; Hood, 2002, 2010; Leong & Howlett, 2017; Rodriquez, 2022). In claiming credit, politicians seek to declare a policy a success and portray themselves as being responsible for that success, rendering them worthy of re-election (Hood, 2010; Howlett, 2014; Leong & Howlett, 2017; Weaver, 1986). Credit claiming tactics include ceremonies, messages, appearances, and other means through which politicians attempt to ensure that voters associate them with positive policy initiatives (Bueno, 2023; Cruz & Schneider, 2017).

The literature on blame avoidance and credit claiming suggests that policymakers value the avoidance of blame more highly than garnering credit (Balla et al., 2012; Hood, 2002, 2010; Howlett, 2012, 2014; Leong & Howlett, 2017; Twight, 1991; Weaver, 1986). In other words, they are more inclined to minimize blame than to maximize credit. They are willing to forego potential credit-claiming opportunities to avoid blame (Hood, 2002; Weaver,

1986). These insights have been derived from the pioneering works on cognitive psychology by Kahneman and Tversky (1979, 1984). As Kahneman and Tversky (1979, 1984) showed, people attribute higher weight to losses compared to equivalent gains, and, in a situation of uncertainty, people overestimate negative consequences. These tendencies, in turn, are linked to negativity bias (Hood, 2010; Rozin & Royzman, 2001; Weaver, 1986). "Negativity bias denotes a commonly observed cognitive tendency for more attention to be paid to negative than to positive information and for losses to be valued more highly than gains of an equivalent amount" (Hood, 2010, p. 9). Since the nature of their positions in the power hierarchy is precarious, elected officials tend to be blame averse (Hood, 2002, 2010; Leong & Howlett, 2017; McGraw 1990, 1991). Decision-makers' desire to avoid blame for failures leads them to "attempt to avoid any action at all and then only when forced to do so by the threat of blame for inaction to undertake as little action as possible" (Howlett, 2014, p. 396).

Leong and Howlett (2017) emphasize that the attribution of blame and credit take place within broader aspirations of legitimacy by policymakers and their perceptions of whether they would be held accountable by the electorate for their policy choices. Thus, credit claiming and blame avoidance are behavioural concepts "but also capture components of key structural and institutional issues surrounding the accountability of officials and politicians for their actions" (Ibid., p. 600). Leong and Howlett (2017, p. 602) suggest that in order to assess the role of blame avoidance and credit claiming in policymaking, it would be insightful to examine "the role of the public in casting blame and ascribing credit, how decision-makers assess this, the reasons why decision-makers might choose to avoid blame or garner credit".

## Implications of blame avoidance and credit claiming for experimental policymaking

### Credit claiming and experimental policymaking

The attribution of credit would require "a judgment as to the extraordinary nature of policy-makers efforts without which a laudable policy or outcome would not have occurred" (Leong & Howlett, 2017, p. 601). There are various ways how policymakers can score points by launching policy experiments.

First, policy experiments are a way for decision-makers to show that they are using evidence-based policymaking (e.g. Sanderson, 2002; Stoker & John, 2009) and to glean credit for taking a rational approach to policymaking. Policymakers can garner such credit *ex ante* by announcing that they experiment before a policy solution is scaled up. They can also reap credit *ex post*, after an experiment has been conducted, by showing that a policy solution they are planning to launch is supported with experimental evidence. As Corduneanu-Huci et al. (2021, p. 4) put it, experiments allow politicians to signal voters their "intentions to implement policies based on rigorous evidence, address criticism, and establish objectivity in politically contentious issues". Policymakers can expect to garner such credit from policy experiments when evidence-based policymaking is regarded as legitimate and beneficial by the public, media, and political opponents.

Second, the testing of a policy on a small scale *before* launching a reform enables policymakers to argue that an experiment helps to deal with uncertainty and avoid potential large-scale failures (Heilmann, 2008; Howlett, 2012; Makse & Volden, 2011; Nair & Howlett, 2016; Rogers, 2004). By experimenting, policymakers can demonstrate to the electorate that they are cautious in the face of uncertainty and want to shield the public from

larger harm. Credit can be garnered by pointing to improved knowledge about the policy effects that the experiment or pilot offer (Nair & Howlett, 2016). Further, if the experiment fails and the solution does not deliver the expected results, policymakers can point to the fact that the experiment allowed them to avoid a larger failure (Heilmann, 2008). They can, for example, reap credit by referring to the costs that were spared by not launching a reform prone to failure and how that saved taxpayers' money (Bravo-Biosca, 2020; Farrelly, 2008; Lee & Ma, 2020; Raudla et al., 2023).

Third, launching an experiment allows policymakers to demonstrate that they are taking steps towards addressing a societally relevant policy problem while actually avoiding an initiation of a comprehensive reform. If addressing a particular problem entails painful and unpopular decisions, politicians may use an experiment as a "substitute" for large-scale policy action (Brodkin & Kaufman, 2000; Howlett, 2014; Nair & Howlett, 2016). In other words, pilots can serve "as an excuse for policymakers to delay large-scale policy reforms beyond their term in office" (Nair & Howlett, 2016, p. 71) and serve predominantly symbolic purposes (Teets & Hasmath, 2020). In that way, the elected officials can demonstrate that they are "dealing" with a problem – and reap credit for taking action – but at the same time avoid retribution from the electorate for politically difficult decisions.

### Blame avoidance and experimental policymaking

Although policy experiments entail possibilities for garnering credit, they can also give rise to blame and hence make policymakers hesitant to engage in experimental policymaking.

First, by launching a policy experiment, policymakers admit uncertainty and insufficient knowledge about the effects of the policy solution or intervention considered (Bauknecht et al., 2020; Farrelly, 2008). Instead of acknowledging uncertainty, however, elected representatives may prefer to be perceived as "decisive, energetic, and positive" (Farrelly, 2008, p. 11) and emphasize "profound knowledge" about future developments (Bauknecht et al., 2020, p. 57). Experimenting strikes the "chord of skepticism and indecision" that politicians seek to avoid (Peters, 1998, p. 126). Thus, in order to avoid blame for appearing indecisive and uncertain, policymakers may prefer to refrain from policy experiments.

Second, policy experimentation may be perceived by the public as a postponement or delay in solving a salient problem. If policymakers were to respond to public pressure to address an urgent societal problem with an experiment instead of a large-scale policy, they might be viewed as doing "too little" and hence generate blame for wasting precious time in addressing a problem (Hood, 2010; Howlett, 2014). When faced with an urgent societal problem or a crisis, elected officials are likely to assume that voters would give credit to leaders who take charge boldly and are "not dithering on the sidelines blaming everyone else" (Hood, 2011, p. 71). Thus, in order to avoid the blame for a delay, policymakers would prefer to avoid experimentation.

Third, due to inherent uncertainty, experiments entail the possibility of failure. Failures, in turn, can be used to blame politicians (Howlett, 2012, 2014). Like with policy innovations more generally (e.g. Berry & Berry, 2018), conducting policy experiments can entail considerable budgetary costs (Heilmann, 2008). Thus, policymakers may fear that if an experiment they initiated fails, they could be blamed for wasting public resources (Raudla et al., 2023). Furthermore, policymakers may fear that a failed experiment could generate quicker and more visible blame than a large-scale policy reform where the failure could take a longer time to materialize and be more diffuse. If a policy failure is gradual, rather than acute, it is likely to generate less blame (Bovens & t'Hart, 1996; Howlett,

2014). Visible failures, in contrast, have more potential for blame generation (Howlett, 2012, 2014). As experiments usually involve concrete metrics to measure "success", often by assessing a change in quantitative indicators, it may be possible to highlight "failure" more clearly and visibly than would be the case with large-scale reforms where indicators are more ambiguous. Consequently, in order to avoid such visible blame for failed experiments, policymakers may be disinclined to engage in experimental policymaking.

### Implications of negativity bias for experimental policymaking

Drawing on insights about negativity bias in policymakers' motivations concerning blame avoidance and credit claiming (Hood, 2010; Howlett, 2014; Leong & Howlett, 2017), we can postulate the following general expectations about experimental policymaking.

First, *ceteris paribus*, when faced with the option of conducting a policy experiment or doing nothing, the blame-minimization tendency should lead policymakers to prefer the latter. The political credit garnered from a successful experiment is likely to be relatively small given the limited scale. At the same time, due to uncertainty, experiments entail a significant chance of failure. Drawing on Kahneman and Tversky's (1979, 1984) argument that people attribute higher weight to losses compared to equivalent gains and in the context of uncertainty overestimate the negative consequences, we can expect that politicians prefer avoiding the blame associated with failure and hence avoid undertaking the experiment. Furthermore, as experiments usually entail the explicit pronouncements of results and hence visible labelling of "failure", policymakers may view experiments as unnecessary liabilities in their portfolio of decisions. Furthermore, judgments of intentionality and avoidability are important considerations in blame attribution (Hood, 2002, 2010; Howlett, 2012, 2014; Leong & Howlett, 2017; Leong et al., 2023; Weaver, 1986). The potential for blame attribution is higher when policy actors are seen as having intentionally undertaken a course of action that could have been avoidable (Hood, 2010; Howlett, 2014; Leong & Howlett, 2017). Hence, in the case of experiments, policymakers may fear that if the experiment fails and inflicts harm, the blame-generators may easily claim that the experiment should have been avoided.

Second, *ceteris paribus*, when faced with the choice of launching a large-scale reform or conducting an experiment, the blame-minimization imperative can be expected to lead policymakers to prefer the latter. A failed reform would generate more extensive blame than a failed experiment due to the differences in scope and the number of people affected (Howlett, 2014). The more extensive the perceived harm, the stronger the attribution of blame is likely to be (Hood, 2011). Since in a situation of uncertainty, people overestimate negative consequences (Kahneman & Tversky, 1979, 1984), we can expect that conducting an experiment will be perceived as entailing more limited negative consequences compared to a large-scale reform. These choices may be influenced by various contextual factors.

### Contextual factors

Several authors have called for investigating how different contextual factors influence blame avoidance and credit claiming behaviours (e.g. Cruz & Schneider, 2017; Hinterleitner et al., 2022; Leong & Howlett, 2017; Leong et al., 2023). As Leong and Howlett (2017, p. 601) emphasize, "ascribing blame, or credit, therefore is not just a matter of evaluation of the character and value of a policy but is also a matter of determining agency and culpability within a set of social and political norms". Experimental policymaking offers fruitful

opportunities to explore situations where the meaning of blame and credit are *ambiguous* and influenced by contextual factors. In particular, various contextual factors are likely to influence whether the *failure* of an experiment is perceived to generate "blame" or "credit". If policy actors think that a failed experiment would be predominantly regarded as a "failure", they would be hesitant to launch experiments. In contrast, if policy actors regard a failed experiment as a possibility for claiming credit (as it offers evidence of learning or an averted "larger failure"), they would be more inclined to support policy experiments. Drawing on the existing literature, we would expect the following contextual factors to influence the blame-credit calculus in experimental policymaking.

First, the *saliency* of the policy field or the policy problem and the perceived *urgency* of the problem are likely to affect policymakers' blame-credit calculations (Howlett, 2014; Leong & Howlett, 2017) when deciding whether to undertake an experiment. The discussions on blame avoidance have pointed out that if the policy problem is not visible and has a limited scope, "it may not necessarily require any ameliorative or corrective action in order for a government to escape the blame for failing to deal with it" (Leong & Howlett, 2017, p. 608). In such a context, policymakers may expect that taking small steps (like an experiment) would be seen as "doing something" and through that they can reap credit or avoid blame (Howlett, 2014; Leong & Howlett, 2017; Maor, 2017). Conversely, when the problem is perceived as salient and urgent, and voters view policymakers as responsible for resolving it, large-scale policy action would be preferred over small steps (Howlett, 2014) as it would entail the possibility of garnering larger credit. Studies about motivations behind policy innovation suggest that the saliency of the policy problem fosters the adoption of policy innovations (Berry & Berry, 2018; Jordan & Huitema, 2014; Krause, 2011).

Second, *the scope of the experiment* is likely to affect the blame-credit calculus of policymakers. If the policy experiment is large-scale (e.g. pilots or RCTs affecting hundreds or thousands of people), it is likely to attract public attention (Oakley, 1998). Hence, the expected public attention would amplify the size of the projected credit or blame of an experiment and influence how policymakers perceive the expediency of undertaking a large-scale experiment. Given that in smaller experiments the failure is more limited – and hence the blame potential circumscribed (Howlett, 2014) – policymakers may favour small-scale experiments.

Third, drawing on the work of Kahneman and Tversky (1979, 1984), we can expect that the degree *of uncertainty* regarding the outcomes of the experiment can play a role in the blame-credit calculus. On the one hand, the higher the level of uncertainty of the outcomes, the greater the opportunities for policy learning (Ansell & Bartenberger, 2016; Bravo-Biosca, 2020; Nair & Howlett, 2016) and the larger the potential size of the "credit" policymakers can claim. On the other hand, the higher the uncertainty, the higher the perceived probability that the experiment could fail, which may induce risk aversion (Kahneman & Tversky, 1979, 1984) and hence a disinclination to experiment (Ettelt et al., 2015; McFadgen & Huitema, 2018a, 2018b; Teets & Hasmath, 2020).

Finally, in addition to the features of the policy problem and the experiment itself, various aspects of the broader political setting within which experimental policymaking takes place can influence the credit-blame calculus. The existing literature on blame avoidance and credit claiming has pointed to the role of the media and public trust as potentially relevant factors. The media tends to reinforce the negativity bias by foregrounding blame and downplaying credit (Hinterleitner, 2017; Hood, 2011). The extent to which the media is willing to ascribe blame to individual politicians is likely to affect their blame-credit calculus (Hinterleitner & Sager, 2015) in experimental policy making. As trust in government is expected to influence blame avoidance behaviour (Boin et al., 2009; Hinterleitner

& Sager, 2015; McGraw, 1990), it is likely to influence policy-makers attitudes towards policy experiments as well. A government that enjoys high trust is likely to be more successful in advancing its own narrative concerning the policy action in question (Hinterleitner & Sager, 2015); thus, it is more likely to be positively predisposed to experimental policymaking.

## Methods

To probe the plausibility of the theoretical expectations outlined in section two, we draw on interviews carried out in Finland and Estonia between July 2022 and July 2023. The semi-structured interviews (N = 32 in Estonia and N = 34 in Finland) were carried out in both countries according to the same interview protocol. They engaged central government officials from the Government Office / Prime Minister's Office, financial and fiscal policy institutions, and other public organizations that have had experiences with experimental policymaking. The list of interviewees is provided in Appendix 1. To maintain the anonymity of the interviewees, they are identified by a short code in the analysis (e.g. Est1, Fin1).

Our case selection proceeded from the following considerations. We wanted to examine countries that are similar with regard to the key features of their political institutions but vary in terms of how much their central governments use policy experiments. Estonia and Finland have very similar political institutions. They both have parliamentary systems, proportional electoral systems, and coalition governments. Both are unitary (rather than federal) systems. Despite having similar political institutions and being geographically close, the two countries represent different cases in terms of their historical-social background and experience with policy experimentation at the central government level. Finland has had extensive experience with policy experiments, going back several decades (Antikainen et al., 2017). Experimental policymaking was re-emphasized during the Prime Minister (PM) Juha Sipilä's government in 2015–2019 via launching the program "Experimental Finland" (Leino & Akerman, 2022). Within the framework of this program, alongside numerous smaller pilots, six large-scale policy experiments were launched, including the universal basic income (UBI) experiment (Kangas et al., 2021). The subsequent government led by PM Sanna Marin launched various large-scale RCTs and non-randomized pilots as well, focusing on the fields of education and employment. In Estonia, policy experimentation has emerged only recently and has tended to be bottom-up. Experimentation has been facilitated by public sector innovation programs financed by the European Union funds and by "innovation sprints" and "experiment design sprints" organized by the Public Sector Innovation Team located in the Government Office. The experiments conducted in Estonia so far have mostly been small-scale design experiments and policy pilots initiated by individual organizations (Riigikantselei, 2023).

Such differences between the cases allow us to examine how perceptions related to credit claiming and blame avoidance in experimental policymaking vary in different settings and to explore contextual factors affecting the credit-blame calculus. The interviews explored the perceptions of Estonian and Finnish public officials regarding the benefits and pitfalls of policy experiments as well as their own experience with experimenting. The average interview length was one hour. All the interviews were recorded and transcribed. In coding the interviews with MAXQDA, all the statements of the interviewees concerning the role and perceptions of politicians and political institutions towards policy experiments

and experimentation as an approach as well as all the statements regarding the risks and benefits of experimenting were coded with respective codes. We combined deductive and inductive approaches to coding. In the deductive phase, we sought to identify to what extent the factors outlined in the theoretical discussion could be observed in the interviews. In the inductive phase, we sought to identify factors that were not covered by theoretical expectations.

## Empirical analysis

### Perceived blame and credit in experimental policymaking

Our interviews reveal a clear contrast between the perceptions of public officials in Estonia and Finland regarding the credit and blame associated with public policy experiments. In Finland, public policy experiments are largely regarded as a cautious approach to policymaking − a way to avoid blame. In contrast, in Estonia, policy experiments are predominantly perceived to be "risky" and entail considerable potential for blame.

#### Finland: experiments perceived as credit-generators

Our interviews indicate that in Finland, politicians support an experimental approach to policymaking. As one of the interviewees put it, "When a new policy is considered, often the automatic reaction of politicians is: 'Let's pilot it first'" (Interview Fin2). Many interviewees observed that politicians regard experiments as opportunities to garner political credit. For example, it was observed that "politicians want to conduct experiments because they want to be known for initiating these" (Interview Fin10). Another interviewee noted that "politicians always want to initiate something new, and a pilot is a very clever way of starting something new and because it costs less than a full-scale reform and, in that way, you can start a larger number of new things" (Interview Fin2). One of the interviewees also suggested that "undertaking an experiment is a way for a politician to signal that they are serious about the reform or change" (Interview Fin29).

Furthermore, in Finland, policy experiments are perceived as an instrument for alleviating uncertainty before launching major policy decisions. Overwhelmingly, the interviewees said that it is better to fail in an experiment than in a major policy change itself. According to many interviewees, when the experiment fails − in the sense that it does not achieve the expected results − it still offers opportunities for learning and avoiding larger harm. As the interviewees explained, "When the results of the experiment are bad, politicians are like: 'Isn't it good that we started by piloting? Now the negative outcomes are much more restricted.' … Experimenting is a careful person's policy: you take small steps forward and at each step you have the possibility to rethink the approach. It would actually be risky not to use experiments" (Interview Fin2). "Failure is something that gives us more information. And if we experimented with something and found it didn't work, it is a good result as well, because now we know it didn't work and now we can do something else. And it's better than implementing the solution on a large scale and then finding it doesn't work" (Interview Fin8). For example, an experiment testing the effects of the abolition of employers' social security contributions in Northern Finland on employment showed that it did not lead to increased employment. Subsequently, the government decided not to continue with this policy solution (Interview Fin12).

Furthermore, in Finland, politicians have used experiments to introduce "radically" innovative ideas. As one of the interviewees explained, "I do see policy experiments as a productive way of introducing radically innovative policy approaches. For example, the Minister of Labor publicly stated that she has asked the ministry to consider different alternative ways of implementing a four-day work week experiment. She wouldn't be bold enough to say that she is advocating for a four-day work week in Finland for the next governmental term since that would be politically quite risky. But, when she says that she's willing to push for an experiment within the next governmental term, it is a way of mitigating political risk" (Interview Fin34).

The Finnish case also indicates that in addition to garnering "domestic credit" from the experiments (i.e. from the local electorate), policy experiments allow policymakers to glean reputational benefits internationally. It was observed that one of the motivations behind launching the Experimental Finland programme was to receive international attention and recognition for embracing experimental policymaking (e.g. Interview Fin7).

In section two, we conjectured that experiments could be employed as a strategy for avoiding larger "pain" – and also blame – associated with unpopular policy decisions. Our interviewees suggest that policy experiments in Finland have indeed been utilized in such a way. Several interviewees argued that since launching pilots requires less time, money, and effort than a full-scale policy change, politicians often prefer it as a first step in addressing a problem (Interviews Fin2, Fin4, Fin8, Fin16, Fin20). As one of the interviewees put it, "If politicians can't decide on tough things then they say: 'first we can experiment?' But it can be just a way of pushing tough decisions forward" (Interview Fin8).

Also, it emerged from the Finnish interviews that, at times, politicians use experiments to reach compromise in situations where deciding on a "permanent" policy solution may be too difficult (Interviews Fin4, Fin8, Fin12, Fin16, Fin20, Fin29). This often happens when the coalition parties in the government cannot agree on a policy. In the words of an interviewee, "If the coalition partners in the government cannot agree on a measure, then one solution is to make it temporary" (Interview Fin20). For example, reaching an agreement on offering tax deductions for household services (that constitute green investments) was facilitated by the fact that these were decided to be temporary and regarded as experiments (Interviews Fin4, Fin12). As another example, a job alternation leave trial came about because the government had difficulties reaching a compromise with the employers' unions and labor unions. They agreed on conducting a trial rather than launching a full-scale policy (Interview Fin12).

### Estonia: experiments mostly perceived as potential blame-generators

Our interviews indicate that in contrast to Finland, an "experimental mindset" is not part of the Estonian politics yet. Although politicians have not interfered in or stopped small-scale experiments undertaken by civil servants, there have been no large-scale strategic policy experiments (comparable to the Finnish UBI experiment) in Estonia. The interviewees, for the most part, observed that politicians in Estonia do *not* view policy experiments favourably. As one of the interviewees put it, "I really cannot imagine politicians in Estonia launching anything like the Finnish UBI experiment" (Interview Est2). Another interviewee observed, "When I was trying to convince politicians of the value of experiments, they told me that I have a naive view of how policies are really made" (Interview Est19). It was also noted that "it seems to be difficult to convince politicians that if they plan a

large-scale reform, first, a small sum could be foreseen for testing whether the solution works" (Interview Est18).

Interviews with Estonian officials revealed that politicians do not regard experiments as possibilities to garner credit. Public policy experiments are viewed to offer considerably fewer credit-claiming opportunities than launching large-scale policy solutions. As one of the interviewees aptly put it, "When politicians have given a promise (e.g. in the government programme or some action plan), then at the end of the year they want to beat their chest and show what they have actually accomplished. If they had to say that no, this promise could not be implemented because it is only being experimented on, that would not be enough" (Interview Est18). Another interviewee noted that for politicians, "experimentation sounds like the problem is not regarded seriously enough" (Interview Est6).

The Estonian interviewees observed that although politicians embrace innovation, they view conducting experiments as politically risky, and experiments are regarded as potential political liabilities rather than assets. In the words of the interviewees: "To the politicians' ears, experimenting sounds risky" (Interview Est24). "Politicians don't have the courage to experiment" (Interview Est2). According to our interviews, the perceived riskiness of experiments is among the key reasons why large-scale public policy experiments akin to the Finnish UBI experiments have not been launched in Estonia. Such strategic experiments would require political commitment and politicians would need to be involved to initiate the trial, provide the legislative mandate, and secure funding (Interviews Est2, Est7, Est11, Est12, Est15, Est17, Est18, Est19, Est20).

According to our interviews, politicians fear that experiments could generate considerable blame when they fail. A failed experiment is perceived to bring negative public backlash and subsequent negative electoral consequences (e.g. Interviews Est2, Est11, Est13, Est17, Est18, Est20, Est21, Est24, Est28). As one of the interviewees put it, "It is very difficult to communicate when the experiment fails. And that makes it politically risky" (Interview Est18). Another official explained: "Politicians' pain threshold for failure is very low. If a minister is in office for 1.5 years and then is facing elections, they wouldn't want to report that they failed in several experiments. They wouldn't want to say that they tried several things but failed in all of them. That would not get them any votes at the elections" (Interview Est17). Several interviewees pointed out that if an experiment were to fail, it would generate blame for wasting public resources (Interviews Est2, Est5, Est13, Est21, Est32).

As suggested in section two, one of the "blames" associated with experimentation is that it entails admitting uncertainty by policymakers. Our interviews indicate that this is indeed one of the obstacles to experimental policymaking in Estonia (Interviews Est4, Est18, Est32). As one of the Estonian interviewees summarized it: "The biggest challenge for experimentation is that politicians find it difficult to deal with uncertainty. They tend to think that the public expects black and white, yes/no answers from them and that admitting uncertainty – a probability distribution of different outcomes – is somehow politically unacceptable" (Interview Est4). Another interviewee observed, "If the experiment fails, the public reaction is likely to be: why did you waste resources on an experiment? Why did you not know immediately what the right thing to do is?" (Interview Est32).

Overall, negativity bias (Hood, 2002, 2011) can help to explain officials' sceptical stance about the feasibility of launching strategic experiments in Estonia. Such experiments are viewed as generating "too little" political credit, while their potential failure is perceived to give rise to political blame. As one of the interviewees put it, "Politicians have very low risk tolerance: their key motivation is to avoid any kind of mess or fuss that they would then have to mop up" (Interview Est2).

## Factors influencing the perceptions of blame and credit in public policy experimentation

In the theoretical section, we expected blame avoidance and credit claiming to be mediated by contextual factors. In particular, we expected the nature of the policy problem, the scope of the experiment, the expected media reactions, and public trust to influence the perceptions of blame and credit related to experimental policymaking. In addition to corroborating the relevance of these factors (discussed in Sects. "Perceived urgency of policy problems"-"The level of public trust in government"), our comparative study also uncovered the accumulated experience with experimenting as an important aspect in shaping the political calculus (discussed in Sect. "Cumulative experience with experiments").

### Perceived urgency of policy problems

In section two, we conjectured that the perceived urgency of policy problems is likely to hinder policy experimentation since delay is viewed as generating blame. The Estonian interviews do indeed indicate that delays in solving societal issues are often perceived as problematic, which, in turn, constrains experimental policymaking (e.g. Interviews Est1, Est4, Est13, Est15, Est17, Est27). As one of the interviewed officials observed, "Politicians feel that they have to put out the fire immediately. They want to put the seed into the ground today and take out the carrot tomorrow. This makes it challenging to introduce policy experiments" (Interview Est17). Thus, the perceived urgency of having to solve problems makes politicians reluctant to experiment, as this would delay the resolution of the problem. In addition, it was observed that "especially in domains where the target groups need help or assistance from the government, it would be difficult for the politicians to say that they are going to postpone the solution because they are experimenting" (Interview Est18).

Although the Finnish government has – to some extent – been able to overcome the short-termism pressures imposed by the perceived urgency and launched large-scale policy experiments, the interviewees did mention that time pressures have also undermined experimentation. As one of the interviewees summarized it: "We are currently living in social media driven political reality where politicians are so stuck to minute-based communication that it seriously harms visionary politics of which experimental policymaking could be part of" (Interview Fin29). It was also observed that the perception of a crisis creates pressures to offer immediate solutions, which, in turn, undermines experimentation. As one of the interviewees put it, "During the COVID-19 crisis, nobody in their right mind could suggest experimenting with the solutions before scaling them up. We had to act immediately" (Interview Fin34).

### The scope of the experiment

As conjectured in Sect. "Theoretical discussion", the scope of the experiment is likely to affect the blame-credit calculus of politicians and that policy actors would favour smaller-scale policy experiments. In the Finnish context, we can see that politicians have not shied away from launching large-scale experiments that have high visibility.

In the Estonian context, although politicians have not launched large-scale experiments, there have been some smaller-scale experiments where politicians have been involved. For

example, the comprehensive performance budgeting reform undertaken in Estonia in 2020 led to a considerably shorter budget law and created dissatisfaction among some legislators and media who argued that the budget was not transparent enough. In 2022, design experiments were conducted to improve the structure and content of the draft budget and the ways of presenting budget information to the public. There was a political demand for these experiments since the performance budget format had become strongly criticized and the finance minister was involved in the experiment (Interviews Est8, Est11, Est15). As the government politicians did not want to discard the major budgeting reform itself, they hoped to improve the presentation of budget information via the design experiments. The design experiments (involving critical users from the media who had been complaining about the new budget format) allowed the minister of finance to demonstrate that efforts are made to address the lack of transparency in the budget and through that garner some political credit.

### Expected media reaction to experiments

In the Estonian context, the anticipated media attention to an experiment and its failure was considered to be an important mediator of blame (Interviews Est4, Est13, Est17, Est18, Est29, Est32). Policymakers expect the media to react negatively to a failed experiment and amplify the critical attention to it. As one of the interviewees put it, "A failed experiment would elicit negative media attention, and this failure would not be forgiven" (Interview Est29). Several interviewees mentioned that an important concern of politicians is to avoid scandals and unfavourable attention in the media. An interviewed official observed that "in the era of 24/7 news media and the internet, the media dictates what politicians do" (Interview Est4). It was also noted that "if an experiment were to fail, you really need a strong communication department" (Interview Est17). It was argued that "it is difficult to communicate to the public that even if an experiment failed, it provided new information" (Interview Est18).

The interviews with Finnish public officials indicate, however, that fears about negative media attention have not deterred policy experiments. As one of the interviewees explained, "Of course, the media can say that '20 million euros was used up and nothing was achieved', but I cannot see that anyone would be afraid of running experiments because of potential media attacks" (Interview Fin9). Another interviewee observed that in response to the media attention, "You can always say that 'if you don't try anything, you don't get anything'" (Interview Fin1).

### The level of public trust in government

In Sect. "Theoretical discussion", it was proposed that trust is another aspect of the political setting next to media attention that may shape the blame-credit calculus and explain politicians' stances towards experimentation. Trust emerges as an important factor from the comparative study of Estonia and Finland, indicating the relevance of the socio-political context for experimental policymaking. A high level of public trust in government helps to explain the favourable perception of experimental policymaking in Finland and hence the corresponding expectations of blame and credit associated with experiments. High trust translates into public confidence that the government would not abuse an experiment but would launch it for a useful purpose. High public trust can give policymakers the sense that they are allowed to fail with the experiment,

without getting the blame. As one of the Finnish interviewees explained, "In those societies where the trust is high, it is always easier to experiment, because people don't think that the government is harming their interests, but they are testing something new, which is supposed to improve the current system. If people do not trust the public authorities, it would be almost impossible to conduct policy experiments." (Interview Fin29) Another interviewee observed, "Societal trust has been of very high level in Finland. That is why we could conduct the pre-primary education trial at all" (Interview Fin16).

In the Estonian context, in contrast, the interviews pointed to the lack of trust towards the government and concerns that experiments would elicit negative reactions from the public (Interviews Est2, Est17, Est20, Est21). As one of the interviewees put it, "If they hear the word "experiment", the public is likely to think that it is another stupidity that is likely to make their life worse" (Interview Est2). Another interviewee argued that the public may react negatively to a policy experiment because it might sound like the government is "treating the public as lab rats" (Interview Est20). It was observed that "in the Estonian society, if the solution that was experimented with does not work, it is seen as a failure and a waste of time and resources − instead of seeing it as a learning opportunity" (Interview Est21). As one of the interviewees explained it, "The public wants every taxpayers' cent to be used rationally and that is why it is difficult to undertake such 'nice to have' type things like experiments. Experimentation sounds like an unnecessary luxury" (Interview Est13). The Estonian interviews suggest that the lack of societal trust probably amplifies the negativity bias of policymakers and, consequently, creates a more discouraging environment for policy experimentation.

### Cumulative experience with experiments

Our comparative case study points to another contextual factor that can play a role in affecting the blame-credit calculus of policy experiments: past accumulated experience with experiments. If experiments have become an integral part of the policymaking process and legitimized in the eyes of the public, launching any given experiment may feel less risky for policymakers.

Our interviews with Finnish officials suggest that the cumulative experience with past experiments in Finland and in particular the Sipilä Government's "Experimenting Finland" programme legitimized the use of experiments in the public sector and made them less risky in the eyes of the policymakers (Interviews Fin3, Fin4, Fin5, Fin6, Fin7, Fin16, Fin18, Fin21, Fin28, Fin29, Fin32, Fin33). It allowed the spread of the "it is OK to try and fail if failure leads to learning" narrative throughout the public sector organizations (Interviews Fin5, Fin21). Furthermore, the "Experimental Finland" program improved politicians' knowledge about what experiments are and "how these could be used to introduce radical policy suggestions" (Interview Fin34).

In Estonia, the lack of experience with large-scale experiments helps to explain the prevailing hesitancy among politicians to launch a large-scale experiment. One of the interviewees observed that, "We would need good examples of where experiments have helped to save money and averted larger-scale failure" (Interview Est18). Another noted that, "there is still a lot of work to be done in explaining to politicians that experiments are a useful policymaking tool and to alleviate their fears" (Interview Est17).

## Concluding discussion

Part of the appeal of policy experiments is that they hold the promise of "rational and depoliticized policy making" but, in reality, experiments take place in a "messy world of politics" (Checkland et al., 2023, p. 464). Although existing studies on experimental policymaking have acknowledged the importance of the political setting in which policy experiments take place (Brodkin & Kaufman, 2000; Huitema et al., 2018; Mattocks, 2021; Nair & Howlett, 2016), we have lacked systematic knowledge on how various political dimensions affect experimental policymaking. In this paper we draw on the literature of blame avoidance and credit claiming in policymaking to shed light on how political motivations can affect the use of public policy experiments. Although the considerations of blame and credit are regarded as key imperatives of policymakers (Hood, 2002, 2010; Leong & Howlett, 2017; Weaver, 1986), none of the existing studies on experimental policymaking have employed the theoretical lenses of blame avoidance/credit claiming. Our study addresses this gap in the existing literature.

The novel theoretical contribution of our paper is synthesizing insights from the blame avoidance/credit claiming literature and studies on experimental policymaking, and outlining how those considerations can affect policymakers' attitudes towards policy experimentation. The empirical contribution of our paper is to probe the plausibility of the theoretical expectations in two different countries: Finland and Estonia.

As our comparative case study demonstrates, the theoretical lenses of blame avoidance and credit claiming can indeed help to advance our understanding of political motivations in experimental policymaking. Examining the blame/credit calculus of policymakers can help to explain why politicians can be inclined or disinclined to undertake policy experiments. Furthermore, observations from the two cases reinforce the insight of Leong and Howlett (2017) who suggest that the attributions of blame and credit are subjective and dependent on the specific context. The diverging perceptions of blame and credit associated with large-scale policy experiments in Finland and Estonia reveal how the same policymaking instrument can be viewed very differently depending on the political setting. In Finland, policy actors predominantly regard experiments as potential sources of credit: a way to show that politicians are taking steps to address problems and proceed in a cautious way. In Estonia, negativity bias helps to explain why experiments are viewed as politically risky: they are perceived to generate too little credit (compared to large-scale reforms), while the potential failure of an experiment is feared to generate considerable blame. While in both countries policy actors embrace public sector innovation, it seems that in Estonia policymakers perceive that the public expects the government to innovate without experimenting, whereas in Finland, innovation with experimenting is more acceptable and policymakers perceive they have leeway to fail in experiments.

The inclination to consider experiments "risky" in the Estonian context seems somewhat paradoxical, given that undertaking large-scale reforms without testing the novel solution should be perceived as being "even riskier". Indeed, from a "rational" perspective, we would assume that experimenting with a policy solution before scaling it should look riskier to politicians than testing it before launching it. One possible explanation that the blame avoidance/credit claiming literature can offer for this paradox is that experiments may entail faster, visible, measured failures. While these are considered to be the "price of learning" in Finland, in Estonia politicians are perceived to be reluctant to accumulate such portfolios of visible and measurable failure.

Probing the different contexts of Estonia and Finland helps to shed light on why policymakers in the two countries view policy experiments through different blame/credit calculus. The finding of contrasting perceptions and the fear of failure in Estonia may look surprising given the international reputation that the country has built over the past decades in the field of administrative reforms, public sector innovation, and, in particular, advancements in e-government (e.g. Espinosa & Pino, 2024). However, while the development of the digital state in Estonia is indeed noteworthy, the Estonian government is lagging behind in using the ICT infrastructure for digital democracy, citizen engagement, and transforming public welfare services (Kattel & Raudla, 2022). That, in turn, has left at least certain parts of the public disillusioned about what the government can deliver and sceptical towards new initiatives. OECD data also shows that public trust in government is considerably lower in Estonia compared to Finland (OECD, 2023). The lower trust in government in Estonia likely reflects the long-term impact of the neo-liberal ideology adopted by the prevailing political forces in the 1990s seeking a minimal state. Drechsler (1995, p. 112) has argued that despite the very successful initial transformation, newly independent Estonia's most serious problem was "historical distrust of the state" resulting from the Soviet past. This legacy of low trust has been compounded by insufficient policies targeting social welfare and rising inequality (Kattel & Raudla, 2022). In contrast, over the past decades, public organizations in Finland have enjoyed high levels trust from the public (e.g. Grönlund & Setälä, 2012; OECD, 2023; Salminen & Ikola-Norrbacka, 2010), which can be attributed to perceived competencies, equality in the treatment of citizens (Salminen et al., 2012), and satisfaction with policy outputs (Grönlund & Setälä, 2012). Distrust of the state is reflected in the media's approach to politics and governance in Estonia. As Estonia is a poorer country in terms of GDP per capita compared to Finland, it makes the Estonian media particularly keen to attack any perceived "waste" of public resources, especially in the generally prevailing context of austerity and the predisposition of the media to prefer "lean" government (e.g. Mikecz, 2023; Kattel & Raudla, 2022). Such an approach further amplifies the blame potential of failed experiments and hence discourage politicians from undertaking them. Furthermore, having a small population (1.3 million) where the limited distance between citizens and policy-makers allows personification of decisions and direct blame attribution (Sarapuu & Randma-Liiv, 2020) also hinders experimentation.

An additional explanatory factor emerging from our empirical research is that in Finland, major efforts have been undertaken in the past to legitimize experimental policymaking and that has led to cumulative experiences with policy experiments. The use of public policy experiments was actively promoted by the "Experimental Finland" program during the PM Sipilä's government. By offering a strong political mandate for experiments and encouraging experimentation through the Finnish public sector, the perception among policymakers that it is acceptable to fail in an experiment has been able to take root. Such a visible and comprehensive approach to experimental policymaking instilled the narrative that a failure in an experiment is a source of learning and hence, a "failed" experiment can still be viewed as a "success". In Estonia, in contrast, no comparable reform programme promoting experimentation has been undertaken, and experimenting has mostly been a bottom-up initiative. Thus, in Estonia the predominant perception of experiments is that if they fail, they bring blame for wasting public resources. In sum, while in Finland, strategic and systematic experimentation is likely to have contributed to a shared understanding of the nature, aims, and importance of experiments among politicians but also by the public, in the Estonian public discourse, the meaning of "experimentation" is more ambivalent and has been often used as a term to describe failed policy decisions (e.g. Palts, 2019; Tooming, 2024).

Our empirical observations confirm the insight by Leong and Howlett (2017, p. 601), stating that attributions of both blame and credit occur "within broader considerations of legitimacy". The study corroborates the theoretical expectations that the urgency of the policy problems, the scope of experiments, expected media reactions, and public trust are likely to influence blame avoidance and credit claiming dynamics in experimental policymaking. In addition, our empirical study indicates that the cumulative experience with experimental policy making also interacts with the other contextual variables. If a government accrues experiences with policy experiments over time, the public and the media become familiar with them and as a result, it becomes easier to launch large-scale experiments. In that light, the theoretical framework we have proposed in this paper can be used as a starting point to explore experimental policymaking in other countries.

Our comparative empirical study does have several limitations that could be addressed in future research. First, as we used our two cases to probe the plausibility of the theoretical conjectures we put forth, the ambition of the comparative study was not to offer a complete explanation of the differences between two countries. Thus, country-focused idiographic case studies could pursue that in the future. Second, we only examined two countries. Future studies could explore how considerations of credit and blame influence experimental policymaking in a larger set of countries. Furthermore, we focused on two countries that have similar political institutions (unitary parliamentary systems with proportional electoral rules and coalition governments). In order to develop a more comprehensive understanding of the blame-credit calculus in policy experiments, it would be useful to analyze and compare countries with different political structures: presidential, majoritarian, and federal. Fourth, our study only focused on the central government level. However, as the literature on policy experiments has indicated, experimentation can also take place at the subnational level and involve non-governmental organizations. Thus, in future studies, it would be useful to explore whether the credit-blame calculus plays out in a similar way in such settings and how such arrangements facilitate blame-shifting. Fifth, our study focused on political motivations in democratic multi-party settings. Given, however, that extensive policy experimentation can also take place in single-party systems (see, e.g. Heilmann, 2008; Teets & Hasmath, 2020; Zhang & Zhu, 2020), future studies could contrast and compare how blame avoidance and credit claiming influence political motivations related to policy experiments in such different contexts.

Sixth, in this paper, we focused on the overall attitudes of policymakers toward experimental policymaking and whether they are inclined to undertake large-scale policy experiments. Future studies could zoom in on specific phases of policy experiments and examine how the blame/credit calculus affects the design of experiments (e.g. the choice between RCTs and non-randomized pilots), their evaluation, and scaling up. Finally, in our paper, we were able to cover only a limited set of contextual factors that are likely to affect the perceptions of credit and blame involved in public policy experiments. In future research, additional factors that influence the blame/credit calculus of experimental policymaking could be investigated. In particular, it would be insightful to examine whether this calculus varies systematically across different policy domains. Also, future studies could examine how the context of crises and resource scarcity affects how politicians view policy experiments through the credit/blame lens. In our paper we have focused on the political level, but future studies could examine whether politicians' and civil servants' perceptions of experimental policy experiments vary with regard to their blame and credit potential and how such dynamics influences the implementation of experiments.

## Appendix 1

| Reference(s) | Institution | Interview time(s) |
|---|---|---|
| *Estonian Interviewees* | | |
| Est1-Est15 | Ministry of Finance | 06.06.2022–06.01.2023 |
| Est16 | Ministry of Economic Affairs and Communications | 29.11.2022 |
| Est17-Est20 | Government Office | 10.10.2022–01.12.2022 |
| Est21-Est22 | Bank of Estonia | 15.12.2022–05.03.2023 |
| Est23-Est29 | Estonian Tax and Customs Board | 05.07.2022–02.02.2023 |
| Est30-Est31 | Financial Supervision and Resolution Authority | 28.10.2022–09.11.2022 |
| Est32 | Enterprise and Innovation Foundation | 07.12.2022 |
| *Finnish Interviewees* | | |
| Fin1-Fin11 | Ministry of Finance | 15.02.2023–24.04.2023 |
| Fin12-Fin15 | Ministry of Economic Affairs and Employment | 21.03.2023–05.07.2023 |
| Fin16 | Ministry of Education and Culture | 15.03.2023 |
| Fin17-Fin18 | Prime Minister's Office | 14.03.2023–08.05.2023 |
| Fin19 | Bank of Finland | 09.03.2023 |
| Fin20-Fin21 | National Audit Office | 20.02.2023–07.03.2023 |
| Fin22-Fin25 | Finnish Tax Administration | 20.03.2023–05.04.2023 |
| Fin26 | Finnish Financial Supervisory Authority | 20.02.2023 |
| Fin27-Fin28 | KELA | 28.03.2023–08.03.2023 |
| Fin29-Fin32 | SITRA | 24.02.2023–12.06.2023 |
| Fin33 | Motiva | 17.03.2023 |
| Fin34 | Demos Helsinki | 31.03.2023 |

**Acknowledgements** This work was supported by the Estonian Research Council grant PRG1125

**Funding** Eesti Teadusagentuur, PRG1125, Ringa Raudla

## Declarations

**Conflict of interest** The authors have no conflicts of interest to disclose.

## References

Ansell, C. K., & Bartenberger, M. (2016). Varieties of experimentalism. *Ecological Economics, 130*, 64–73.

Antikainen, R., Alhola, K., & Jääskeläinen, T. (2017). Experiments as a means towards sustainable societies – lessons learnt and future outlooks from a Finnish perspective. *Journal of Cleaner Production, 169*, 216–224.

Bailey, S., Checkland, K., Hodgson, D., McBride, A., Elvey, R., Parkin, S., Rothwell, K., & Pierides, D. (2017). The policy work of piloting: Mobilising and managing conflict and ambiguity in the English NHS. *Social Science & Medicine, 179*, 210–217.

Balla, S. J., Lawrence, E. D., Maltzman, F., & Sigelman, L. (2012). Partisanship, blame avoidance, and the distribution of legislative pork. *American Journal of Political Science, 46*(3), 515–525.

Bauknecht, D., Bischoff, T. S., Bizer, K., Führ, M., Gailhofer, P., Heyen, D. A., Proeger, T., & von der Leyen, K. (2020). Exploring the pathways: Regulatory experiments for sustainable development – an interdisciplinary approach. *Journal of Governance and Regulation, 9*(3), 49–71.

Bédécarrats, F., Guérin, I., & Roubaud, F. (2019). All that glitters is not gold. The political economy of randomized evaluations in development. *Development and Change, 50*(3), 735–762.
Berry, F. S., & Berry, W. D. (2018). Innovation and diffusion models in policy research. In C. M. Weible & P. A. Sabatier (Eds.), *Theories of the Policy Process* (pp. 253–297). Routledge.
Boin, A., 't hart, P., & McConnell, A. (2009). Crisis exploitation: Political and policy impacts of framing contests. *Journal of European Public Policy, 16*(1), 81–106.
Bovens, M., & t'Hart, P. (1996). *Understanding policy fiascos*. Transaction Press
Bravo-Biosca, A. (2020). Experimental innovation policy. *Innovation Policy and the Economy, 20*, 191–232.
Brodkin, E. Z., & Kaufman, A. (2000). Policy experiments and poverty politics. *Social Service Review, 74*(4), 507–532.
Bueno, N. S. (2023). The timing of public policies: Political budget cycles and credit claiming. *American Journal of Political Science, 67*(4), 996–1011.
Checkland, K., et al. (2023). 'Success' in policy piloting: Process, programs, and politics. *Public Administration, 101*(2), 463–480.
Corduneanu-Huci, C., Dorsch, M.T., & Maarek, P. (2021). The politics of experimentation: Political competition and randomized controlled trials. *Journal of Comparative Economics, 49*(1), 1–21.
Cruz, C., & Schneider, C. J. (2017). Foreign aid and undeserved credit claiming. *American Journal of Political Science, 61*(2), 396–408.
Drechsler, W. (1995). Estonia in transition. *World Affairs, 157*(3), 111–118.
Espinosa, V.I.,& Pino, A., 2024. E-Government as a development strategy: The case of Estonia. *International Journal of Public Administration*, 1–14
Ettelt, S., Mays, N., & Allen, P. (2015). Policy experiments: Investigating effectiveness or confirming direction? *Evaluation, 21*(3), 292–307.
Farrelly, R. (2008). Policy on trial: Randomised trials are the best tool we have for finding out if policies really work. *Policy: A Journal of Public Policy and Ideas, 24*(3), 7–12.
Grönlund, K., & Setälä, M. (2012). In honest officials we trust: Institutional confidence in Europe. *The American Review of Public Administration, 42*(5), 523–542.
Hansson, S. (2018). The discursive micro-politics of blame avoidance: Unpacking the language of government blame games. *Policy Sciences, 51*(4), 545–564.
Heilmann, S. (2008). Policy experimentation in China's economic rise. *Studies in Comparative International Development, 43*(1), 1–26.
Hinterleitner, M. (2017). Reconciling perspectives on blame avoidance behaviour. *Political Studies Review, 15*(2), 243–254.
Hinterleitner, M., Honegger, C., & Sager, F. (2022). Blame avoidance in hard times: Complex governance structures and the COVID-19 pandemic. *West European Politics, 46*(2), 324–346.
Hinterleitner, M., & Sager, F. (2015). Avoiding blame—a comprehensive framework and the Australian home insulation program fiasco. *Policy Studies Journal, 43*(1), 139–161.
Hood, C. (2002). The risk game and the blame game. *Government and Opposition, 37*(1), 15–37.
Hood, C. (2010). *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton University Press.
Howlett, M. (2012). The lessons of failure: Learning and blame avoidance in public policy-making. *International Political Science Review, 33*(5), 539–555.
Howlett, M. (2014). Why are policy innovations rare and so often negative? Blame avoidance and problem denial in climate change policy-making. *Global Environmental Change, 29*, 395–403.
Hughes, S., Yordi, S., & Besco, L. (2020). The role of pilot projects in urban climate change policy innovation. *Policy Studies Journal, 48*(2), 271–297.
Huitema, D., Jordan, A., Munaretto, S., & Hildén, M. (2018). Policy experimentation: Core concepts, political dynamics, governance and impacts. *Policy Sciences, 51*(2), 143–159.
Jordan, A., & Huitema, D. (2014). Innovations in climate policy: The politics of invention, diffusion, and evaluation. *Environmental Politics, 23*(5), 715–734.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47*, 263–291.
Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist, 39*(4), 341–350.
Kangas, O., Jauhiainen, S., Simanainen, M. and Ylikännö, M. (Eds.) (2021). *Experimenting with unconditional basic income: Lessons from the Finnish BI experiment 2017–2018.* Edward Elgar
Kattel, R., & Raudla, R. (2022). Estonia's radical transformation: successes and failures of 'crazy ideas'. Accessible at: https://economy2030.resolutionfoundation.org/reports/estonias-radical-transformation/
Krause, R. M. (2011). Policy innovation, intergovernmental relations, and the adoption of climate protection initiatives by US cities. *Journal of Urban Affairs, 33*(1), 45–60.

Lee, C., & Ma, L. (2020). The role of policy labs in policy experiment and knowledge transfer: A comparison across the UK, Denmark, and Singapore. *Journal of Comparative Policy Analysis: Research and Practice, 22*(4), 281–297.
Lee, S.-Y., Jung, Y.-R., & Lee, E. (2009). An exploratory study on pilot projects: The ministry for health, welfare, and family affairs in Korea. *The Korean Journal of Public Administration, 47*(4), 279–305.
Leino, H., & Akerman, M. (2022). The politics of making Finland an experimenting nation. *Critical Policy Studies, 16*(4), 441–459.
Leong, C., & Howlett, M. (2017). On credit and blame: Disentangling the motivations of public policy decision-making behaviour. *Policy Sciences, 50*, 599–618.
Leong, C., Howlett, M., & Safaei, M. (2023). Blame avoidance and credit-claiming dynamics in government policy communications: Evidence from leadership tweets in four OECD countries during the 2020–2022 COVID-19 pandemic. *Policy and Society, 42*(4), 564–585.
Makse, T., & Volden, C. (2011). The role of policy attributes in the diffusion of innovations. *The Journal of Politics, 73*(1), 108–124.
Maor, M. (2017). The implications of the emerging disproportionate policy perspective for the new policy design studies. *Policy Sciences, 50*(3), 383–398.
Mattocks, K. (2021). Policy experimentation and policy learning in Canadian cultural policy. *Policy Sciences, 54*(4), 891–909.
McConnell, A. (2010). Policy success, policy failure and grey areas in-between. *Journal of Public Policy, 30*(3), 345–362.
McFadgen, B., & Huitema, D. (2018a). Experimentation at the interface of science and policy: A multi-case analysis of how policy experiments influence political decision-makers. *Policy Sciences, 51*(2), 161–187.
McFadgen, B., & Huitema, D. (2018b). Experimentation at the interface of science and policy: A multi-case analysis of how policy experiments influence political decision-makers. *Policy Sciences, 51*, 161–187.
McGann, M., Blomkamp, E., & Lewis, J. M. (2018). The rise of public sector innovation labs: Experiments in design thinking for policy. *Policy Sciences, 51*(3), 249–267.
McGraw, K. M. (1990). Avoiding blame: An experimental investigation of political excuses and justifications. *British Journal of Political Science, 20*(1), 99–142.
McGraw, K. M. (1991). Managing blame: An experimental test of the effects of political accounts. *The American Political Science Review, 85*(4), 1133–1157.
Mikecz, R. (2023). Neoliberalism ingrained. The rise and consolidation of neoliberalism in Estonia. *Europe-Asia Studies, 75*(4), 606–624.
Millo, Y., & Lezaun, J. (2006). Regulatory experiments: Genetically modified crops and financial derivatives on trial. *Science and Public Policy, 33*(3), 179–190.
Nair, S., & Howlett, M. (2016). Meaning and power in the design and development of policy experiments. *Futures, 76*, 67–74.
Oakley, A. (1998). Public policy experimentation: Lessons from America. *Policy Studies, 19*(2), 93–114.
OECD. (2023). Government at a glance. OECD Publishing Paris
Palts, T. (2019). Kuidas lüüa laiaks 4.5 miljardit [How to waste 4.5 billion]. Postimees, 24 September https://arvamus.postimees.ee/6785660/tonis-palts-kuidas-luua-laiaks-4-5-miljardit
Pearce, W., & Raman, S. (2014). The new randomised controlled trials movement in public policy: Challenges of epistemic governance. *Policy Sciences, 47*(4), 387–402.
Peters, B.G. (1998). The Experimenting Society and Policy Design. In W. Dunn (ed.). The Experimenting Society. Essays in Honor of Donald T. Campbell. New Brunswick, NJ: Transaction, pp 125–140.
Raudla, R., Sarapuu, K., Juuse, E., Harbuzova, N., Onno, K., Vallistu, J., & Cepilovs, A. (2023). To experiment or not to experiment in tax policy? *Halduskultuur – the Estonian Journal of Administrative Culture and Digital Governance, 23*(1), 27–48.
Riigikantselei. (2023). Katsetamise juhend avalikule sektorile [Government Office. (2023). Experimenting guide for the public sector]. Available at: https://katsetamine.riigikantselei.ee/. Accessed 04 Nov 2024
Rodríguez R (2022) From anticipatory strategies to reactive blame games in multi-level settings: the role of structure and politics in stability and policy change. *Journal of Public Policy, 42*(4), 802–826.
Rogers, E. M. (2004). *Diffusion of innovations* (5th ed.). The Free Press.
Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320.
Salminen, A., & Ikola-Norrbacka, R. (2010). Trust, good governance and unethical actions in Finnish public administration. *International Journal of Public Sector Management, 23*(7), 647–668.
Salminen, A., Lähdesmäki, K., & Ikola-Norrbacka, R. (2012). Decent citizenship, justice and trust as cornerstones of legitimation: Tensions between generations in Finland. *International Review of Administrative Sciences, 78*(3), 447–473.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policymaking. *Public Administration, 80*(1), 1–22.
Sarapuu, K., & Randma-Liiv, T. (2020). Small states: Public management and policy-making. In G. Baldacchino & A. Wivel (Eds.), *Handbook on the politics of small states* (pp. 55–69). Edward Elgar.
Stoker, G., & John, P. (2009). Design experiments: Engaging policy makers in the search for evidence about what works. *Political Studies, 57*(2), 356–373.
Teets, J. C., & Hasmath, R. (2020). The evolution of policy experimentation in China. *Journal of Asian Public Policy, 13*(1), 49–59.
Tooming, M. (2024). Rõivas tegevuspõhisest eelarvest: see eksperiment on läbi kukkunud. [Rõivas about performance budgeting: this experiment has failed]. ERR 19 July https://www.err.ee/1609398427/roivas-tegevuspohisest-riigieelarvest-see-eksperiment-on-labi-kukkunud
Twight, C. (1991). From claiming credit to avoiding blame: The evolution of congressional strategy for asbestos management. *Journal of Public Policy, 11*(2), 153–186.
Weaver, K. (1986). The politics of blame avoidance. *Journal of Public Policy, 6*(4), 371–398.
Wenzelburger, G. (2014). Blame avoidance, electoral punishment and the perceptions of risk. *Journal of European Social Policy, 24*(1), 80–91.
Zhu, X., & Zhang, Y. (2020). Comparison of policy experiments: Practices in Asia. *Journal of Comparative Policy Analysis: Research and Practice, 22*(4), 275–280.

# Curriculum Vitae

**Personal data**

| | |
|---|---|
| Name: | Kerli Onno |
| Date of birth: | 25.09.1990 |
| Place of birth: | Estonia |
| Citizenship: | Estonian |

**Contact data**

| | |
|---|---|
| E-mail: | kerlionno@gmail.com |

**Education**

| | |
|---|---|
| 2020–2025 | Tallinn University of Technology, Ragnar Nurkse department of innovation and governance, PhD in Public Administration |
| 2017–2019 | Estonian Academy of Security Sciences, MSC in Internal Security |
| 2012–2014 | Tallinn University, MSC in Communication Management |
| 2009–2012 | Tallinn University, BSC in Information Science |

**Language competence**

| | |
|---|---|
| Estonian | Native |
| English | Fluent |
| Russian | Basic |

**Professional employment**

| | |
|---|---|
| 2025–... | Estonian Tax and Customs Board, head of revenue department |
| 2018–2025 | Estonian Tax and Customs Board, head of analysis unit in the intelligence department; strategy advisor |
| 2019 | Estonian Social Security Board, head of service (psychosocial crisis aid) |
| 2016–2018 | Tallinn University, research communication specialist |
| 2014–2015 | Special Committee on State Budget Control of the Parliament, consultant |
| 2013–2014 | Elion Enterprise, private customer consultant |
| 2009–2012 | Library of Estonian Business School, information specialist |

# Elulookirjeldus

**Isikuandmed**

| | |
|---|---|
| Nimi: | Kerli Onno |
| Sünniaeg: | 25.09.1990 |
| Sünnikoht: | Eesti |
| Kodakondsus: | Eesti |

**Kontaktandmed**

| | |
|---|---|
| E-post: | kerlionno@gmail.com |

**Hariduskäik**

| | |
|---|---|
| 2020–2025 | Tallinna Tehnikaülikool, Ragnar Nurkse innovatsiooni ja valitsemise instituut, avalik haldus PhD |
| 2017–2019 | Sisekaitseakadeemia, sisejulgeoleku magistrikraad |
| 2012–2014 | Tallinna Ülikool, kommunikatsioonijuhtimise magistrikraad |
| 2009–2012 | Tallinna Ülikool, infoteaduse bakalaureusekraad |

**Keelteoskus**

| | |
|---|---|
| Eesti keel | emakeel |
| Inglise keel | kõrgtase |

**Teenistuskäik**

| | |
|---|---|
| 2025–… | Maksu- ja Tolliamet, tulude osakonna juhataja |
| 2018–2025 | Maksu- ja Tolliamet, teabeosakonna analüüsi valdkonnajuht, strateegianõunik |
| 2019 | Sotsiaalkindlustusamet, teenuse juht (psühhosotsiaalne kriisiabi) |
| 2016–2018 | Tallinna Ülikool, teaduskommunikatsiooni spetsialist |
| 2014–2015 | Riigikogu riigieelarve kontrolli erikomisjon, konsultant |
| 2013–2014 | Elion Ettevõtted AS, erakliendi konsultant |
| 2009–2012 | Estonian Business Schooli raamatukogu, infospetsialist |

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-489-6 (PDF)