DOCTORAL THESIS

The Politics of Experimental Policymaking: How Strategic, Ideological, and Temporal Motivations Shape Policy Experimentation

Nastassia Harbuzova

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2026

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
61/2026

The Politics of Experimental Policymaking: How Strategic, Ideological, and Temporal Motivations Shape Policy Experimentation

NASTASSIA HARBUZOVA

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 25/08/2026

Supervisor: Prof. Ringa Raudla
School of Business and Governance
Tallinn University of Technology
Tallinn, Estonia

Co-supervisor: Associate Prof. Külli Sarapuu
School of Business and Governance
Tallinn University of Technology
Tallinn, Estonia

Opponents: Prof. Jaroslav Dvorak
Faculty of Social Sciences and Humanities
Dep. of Public Administration and Political Sciences
Klaipėda University
Klaipėda, Lithuania

Dr. Pekka Kettunen
Migration Institute of Finland
Turku, Finland

Defence of the thesis: 07/10/2026, 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.

Nastassia Harbuzova

---
signature

ISSN 2585-6898 (paperback)
ISBN 978-9916-80-552-7 (paperback)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-553-4 (PDF)
https://doi.org/10.23658/taltech.61/2026

Harbuzova, N. (2026). *The Politics of Experimental Policymaking: How Strategic, Ideological, and Temporal Motivations Shape Policy Experimentation* [TalTech Press]. https://doi.org/10.23658/taltech.61/2026

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
61/2026

Katsetava poliitikakujundamise poliitilised mõjutegurid: strateegiliste, ideoloogiliste ja ajaliste motivatsioonide roll

NASTASSIA HARBUZOVA

TAL TECH
KIRJASTUS

Contents

List of Publications

The list of author's publications, on the basis of which the thesis has been prepared:

I **Harbuzova, N.** (2026). Why successful policy pilots are (not) scaled up: The case of European employment-creation schemes. *Halduskultuur: The Estonian Journal of Administrative Culture and Digital Governance*, 25(1), 24–49. **ETIS 1.1**.

II 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. **ETIS 1.1**.

III Raudla, R., Sarapuu, K., Vallistu, J., **Harbuzova, N.** (2024). It is about time! Exploring the clashing timeframes of politics and public policy experiments. *Perspectives on Public Management and Governance,* 7(4), 137–148. **ETIS 1.1**.

IV **Harbuzova, N.** (2025). Choix de conception et perspectives de passage à l'échelle dans le pilotage des politiques: exploration des relations dans le domaine des politiques de création d'emplois (Original: Design choices and scaling prospects in policy piloting: exploring the relationship in the domain of job creation). *In:* Duverger, T. et al. (eds). *Territoires zéro chômeur de longue durée: de l'expérimentation à l'institutionnalisation*. Le Bord de L'eau, 45–72 (Ch. 3; tr. in French). **ETIS 3.2**.

Author's Contribution to the Publications

Contributions to the papers of this thesis are:

I A single-authored article. The author was responsible for all the research and writing steps required to prepare the manuscript: from designing and managing the study, collecting and analysing data (analytical literature synthesis, interviews, follow-up discussions, desk research), to describing findings and preparing the paper for respective publication.

II An article with the author's input. The author was responsible for transcribing and editing empirical data (interviews), took part in shaping the theoretical discussion section, reviewed and edited the final manuscript, and helped improve it further during the journal peer-review round.

III An article with the author's input. The author was responsible for conducting a targeted literature review to develop the theoretical framework, contributed to the concluding discussion, helped with preparing the empirical data (interview transcription; preparing the data-consolidating appendix) and improving the manuscript during the journal peer-review round.

IV A single-authored book chapter. The author was responsible for all the research and writing steps required to prepare the manuscript.

Introduction

Our increasingly turbulent world forces governments to face a quandary where conventional policy solutions no longer work against complex societal challenges, while their viable alternatives have yet to be developed. In this context, policy experimentation emerges as a prudent approach for developing effective public responses in a timely and cost-effective manner (e.g., Varazzani et al. 2023; OECD 2020). Public policy experiments, according to their rationale, refer to well-structured evidence-producing interventions of limited time and scale that enable policymakers to test, evaluate, and modify novel policy solutions before deciding on their future: whether and how to scale them up (e.g., Huitema et al. 2018; McFadgen & Huitema 2017, 2018; Al-Ubaydli et al. 2021). Experimental policymaking thus spans different types of policy interventions – from randomised controlled trials (RCTs) and non-randomised pilot projects to experimental legislation – that all aim at evaluating the performance of novel policy solutions to ensure that (only) the ones 'that work'[1] will be adopted further, i.e., expanded, prolonged, or both (e.g., Huitema et al. 2018; Ranchordas 2013; OECD 2020).

And indeed, the experience of recent decades demonstrates that policy experimentation is gaining momentum. Diverse literatures indicate a growing use of experimental approaches by governments across the globe (e.g., Nesti et al. 2025; Markowitsch & Scharle 2024; Al-Ubaydli et al. 2021). Moreover, many of these studies promote experimentation further, often framing it within a technocratic narrative where it manifests as an ideologically neutral practice with outcomes that can (and are assumed to) *directly* inform policy decisions (e.g., Laczik et al. 2023; Duflo 2020; Parth & Nyby 2022; Ruggeri et al. 2020). However, a remarkable discrepancy is observed between what policy experiments are expected to be and what policymaking practice shows they *are* (e.g., Capano & Lippi 2017; Checkland et al. 2021; Mattocks 2026). Along with the growing reliance on experimentation in developing policy innovations, there is also growing evidence that it often does not unfold in line with its rationale, that the novel solutions evaluated as 'successful' remain unscaled, and that experimental findings ultimately have little, if any, bearing on policymakers' decisions (Ettelt et al. 2015a; Raudla et al. 2023; Reymond et al. 2020; Zhu & Zhao 2021; Wispelaere et al. 2018). In real-world contexts, public policy experiments are inherently *political* undertakings, as the actors involved in them are not neutral decision-makers. Yet surprisingly, the political dimension of experimental policymaking lacks recognition in academic and practitioner-oriented literatures (e.g., Duflo 2020; Duflo & Bertrand 2016; Varazzani et al. 2023; OECD 2020). Specifically, a two-fold research gap characterises the domain. On the one hand, the studies exploring the effects of politics on policy experimentation remain limited and scattered. And although in recent years the research landscape has been gradually changing (e.g., see Nesti et al. 2025; Roosma 2022; Mattocks 2026, 2025, 2021 investigating the matter), the emerging publications do not provide a holistic view of the matter and remain inconsistent with one another. As a result, much more work is needed to reconcile the knowledge they offer and advance it further. On the other hand, the domain lacks analytical frameworks that would facilitate such internal research coherence and enable the development of a more systematic view on the variety of political influences involved in experimenting with novel policy solutions.

[1] In the thesis, single quotation marks ('') are utilised for coined terms, metaphorical meanings, or meanings with caveats; double quotation marks ("") are used for citations, labels, and when introducing regular terms.

The doctoral thesis aims to address this gap. The overarching purpose of the work is to advance understanding of the political dimension of experimental policymaking by offering a novel analytical lens that renders it more tractable and presenting the findings this lens yields. To explore how politics is involved in and influences public policy experimentation, the author adopts the lens of *political motivations of* policymakers – where the latter denotes actors authorised to initiate policy experiments *and/or* decide on their further adoption and scaling prospects. In the thesis, the term "political motivations" refers to policymakers' *intrinsic* aspirations to maximise their political advantage (such as available authority, time-budget, or electoral appeal) from the decisions made, as well as to align these decisions with specific party interests, favoured ideas, and normative beliefs (in line with Callander 2008; Dunlop et al. 2024; Jones & Whitehead 2018). While multiple considerations beyond the technocratic argument may motivate the choices of policymakers – from socio-economic and financial to legal and ethical factors – political motivations stand out as their *primary* drivers in the policy sciences scholarship, in particular in the literatures on public policymaking and policy change, policy design and innovation governance as well as recent studies on evidence-based policymaking and policy experiments (e.g., Baumgartner & Jones 1993; Hallsworth 2011; Béland & Cox 2016; Schönhage et al. 2024; Garcia-Hombrados et al., 2024; Corduneanu-Huci et al. 2021). However, while signalling the importance of political motivations for understanding how (experimental) interventions operate in real-world contexts, these studies typically do not refer to this conceptual category as such or do not employ it as a self-sufficient framework to structure the analysis. That is, they typically investigate the effects of political power, reputation, electoral expediency, ideology, and normative pre-conceptions *separately from one another*. The thesis, by contrast, pioneers this approach in the domain of experimental policymaking – illustrating how the lens of political motivations can help delineate and explore the variety of political effects that public policy experiments are exposed to.

The doctoral analysis, thus, builds on the three research questions as follows:

1. *What are the key political motivations that influence public policy experimentation?*
2. *Which aspects of experimentation do they influence?*
3. *How do they influence those aspects?*

The thesis is an outcome of the research project "*Experimental approaches and institutional innovations in the domains of fiscal and financial policy*" (Grant PRG1125, funded by the Estonian Research Council). It builds on the findings of two comparative independent studies conducted within the project. Each of the studies employs *a small-N comparative case-study design* ($N_1 = 3$, $N_2 = 2$; N – number of cases compared) yet pursues a different approach to analysing the political dimension of policy experiments, thereby facilitating its more nuanced understanding.

The first study focuses on a specific policy innovation – the *Job Guarantee*. The Job Guarantee refers to an active labour market policy aimed at eradicating (long-term) unemployment by creating voluntary, individually tailored, and locally beneficial job opportunities that do not compete with the regular economy and provide at least a living wage (e.g., Tcherneva 2024). Theoretically, this policy solution is rooted in Post-Keynesian and MMT (Modern Monetary Theory) economics and promoted as a rational, normatively justified, and effective policy in the context of twin transition (e.g., Markowitsch & Scharle 2024; Antonopoulos 2023; Tcherneva & Lalucq 2022). The study

compares the experiences of three ideationally similar pilots that exemplify this policy concept in real-world contexts, including: *Zero Long-Term Unemployment Territories (TZCLD*, France)*, MAGMA Job Guarantee* (*MAGMA*, Austria), and *Kinofelis Ergasia* (*Kinofelis*, Greece). The three employment-creation schemes defy current policymaking approaches to (long-term) unemployment. Rather than managing unemployment and mitigating its negative societal implications, the schemes aim to directly *eliminate* it by creating job opportunities from the bottom up. All three programs were officially labelled as policy pilots, involved experimental evaluation, yet varied in institutional set-up (covering governance and implementation settings of the programs) and primary purposes. Accordingly, the aim of the study was to understand the role of politics in shaping the prospects of this experimental solution and why this role varies in different institutional set-ups.

The second study compares experiences with policy experimentation in two neighbouring countries, Estonia and Finland. The two cases were selected to probe the plausibility of theoretical ideas in two countries with similar political institutions (both countries are unitary parliamentary democracies with proportional electoral systems and multi-party coalition governments) but different experiences with experimental policymaking. In particular, the papers prepared within the study shed light on how specific political drivers – such as *blame-credit calculus* and political timeframes – influence experimental policymaking in the two countries.

The outlined studies have resulted in a series of complementary publications, four of which form the main argument of this thesis. The four doctoral publications – three academic articles and a book chapter – investigate the political dimension of public policy experimentation from different angles, in different contexts, and at different analytical levels.

Article I (builds on the study of employment-creation schemes) demonstrates that political motivations largely affect the (up)scaling dynamics of experimental programs. The study reveals that to be deemed *successful* and qualify for scaling-up, a pilot policy should be compatible with the ideological predispositions of policymakers at different governance levels and not jeopardise their current political advantage.

Article II (draws on the Estonian-Finnish study) takes a closer look into how blame-avoidance and credit-claiming interests shape policymakers' decisions on policy experimentation. The article suggests that the *blame-credit calculus* can largely explain why policy actors may be (dis)inclined to conduct policy experiments or prefer one experimental design over another. The work also illustrates that the calculus *per se* should be studied as a *dependent variable* – being shaped by such contextual factors as the urgency of policy problems, expected media reactions, public trust, and cumulative experience with policy experimentation (perceived legitimacy of the latter).

Article III (Estonian-Finnish study) analyses the causes and consequences of clashes between political and experimental timeframes, explaining how these clashes influence the launching chances of policy experiments as well as such aspects of their research design as the duration, scope, and pace of interventions. In particular, the article indicates that the short-termism of electoral cycles challenges the launch of experimental interventions with transformative policy solutions, as such experiments typically require longer timeframes and larger scales, and thus risk being incompatible with the time-budgets of policymakers.

Book Chapter IV (employment-creation schemes study) focuses on the interplay of politics and the institutionalisation of employment-creation schemes. The work identifies

principal differences between the schemes' institutional arrangements and analyses how they shape the interference of politics in the programs' adoption prospects and (up)scaling dynamics. The analysis concludes with a series of recommendations for practitioners on how to 'design better' – enhancing the adoption prospects of employment-creation schemes. The recommendations span and distinguish between the issues of both scaling-up successful experimental solutions and sustaining them at scale against the backdrop of short-term political cycles. The Chapter concludes that the political risks and opportunities involved in piloting novel policies are conditioned by such characteristics of the pilots as their modes of launch and funding schemes, the types of institutional networks operating the policies, and the learning frameworks adopted in experimentation.

Altogether, the findings of the four publications, combined with a broader literature analysis, allowed the author to address the research questions posed, operationalising the political dimension of public policy experimentation via three interplaying, yet analytically distinct categories of political motivations: *strategic*, *ideological*, and *temporal*. The thesis puts forth these groups of motivations as crucial driving forces that shape policy experimentation. The groups originate in policymakers' aspirations of different kinds. Strategic motivations refer to their aspirations to maximise *private* political advantage while making *public-oriented* policy decisions, encompassing both power- and reputation-maximising aspirations. The aspects of electoral reputation and available authority as factors of policymaking have been prominently studied within the political economy and policy sciences literatures on blame avoidance, credit claiming, and power-seeking (e.g., Capano & Lippi 2017; Britz et al. 2022; Giger & Nelson 2011; Leong et al. 2023). Ideological motivations capture policymakers' inclinations to act in line with their party affiliations, ideational preferences, and normative convictions. Such commitments have been examined within the literatures on policy ideas, partisanship, and evidence-based policymaking (e.g., Wolak 2020; Strøm & Müller 1999; Cairney 2022; Garcia-Hombrados et al. 2024). Temporal motivations cover aspirations of policymakers to align policy choices with the dynamics of political timeframes they operate within. Such dynamics encompass both political cycles per se – electoral cycles, office terms, and leadership turnovers occurring at the level of experimenting agencies – and anticipated policy and political changes at the 'junctures' of these cycles. Policymakers' orientation toward political time has been analysed within the electoral budget cycles and political time literatures (e.g., Bueno 2023; Goetz 2014; Linz 1998; Mink & De Haan 2006). Bringing these diverse forces together via the lens of political motivations enables, for the first time in the domain of experimental policymaking, a systematic and relational view of the variety of political influences that policy experimentation is exposed to.

Taken together, the strategic, ideological, and temporal motivations prove to interact with one another and the broader settings of policy experimentation – affecting whether the latter will be conducted (chances for launch), how it will be conducted (research design; evaluation purpose), and what outcomes it will deliver (evaluation results, including recognition of policy success/failure; adoption prospects, and scaling dynamics). With that, the thesis illustrates that the effects of political motivations are not fixed but largely depend on the specific context of experimentation. The context mediates and moderates the effects of political motivations, explaining why and how the motivations become activated (mediation) and defining their significance, 'sign' (i.e., whether the motivations facilitate or hinder experimentation), and the extent to which policymakers can act on them (moderation). Specifically, such contextual factors as the

institutional set-up of experimental interventions (modes of launch, types of institutional networks), the policy domain they pertain to, and experimentation legitimacy (both its general electoral endorsement and the legal frameworks enabling different policy actors to experiment) prove to shape the interplay of strategic, ideological, and temporal motivations. Altogether, the outlined motivations and the framework they form together – which captures how they interact with one another and with the broader 'operational environment' of experimentation – constitute the author's original contribution to existing scholarship.

The doctoral thesis offers a two-fold contribution to existing scholarship. First, it aids in collecting, systematising, and advancing existing knowledge on the politics of experimental policymaking. Second, it provides a novel analytical framework that is useful for delineating multiple political influences that experimentation with public policy solutions is exposed to. Furthermore, the framework could serve as an analytical map to guide further studies on the matter – helping not only structure their research inquiries but also situate them within a broader and more systematic context. For practitioners, the thesis provides some guidance on navigating the diversity of political interests involved in complex experimentation networks, in order to ultimately negotiate better trade-offs for reconciling these interests. Furthermore, policy actors interested in (piloting) employment-creation schemes could benefit from the comparative findings of the study on European 'job-guarantees', thus advancing understanding of how politics affect scaling-up of such solutions in different settings. The findings are particularly relevant given the growing, yet highly politically contentious attention towards (piloting) employment-creation schemes, as observed both across (European) countries and at the international level (including the movement for the launch of the European Job Guarantee; e.g., Antonopoulos 2023; Tcherneva & Lalucq 2022; Markowitsch & Scharle 2024).

The rest of the thesis is organised as follows. Section 2 introduces research methodology adopted in the studies conducted. Section 3 develops the main argument of the work – analysing how the strategic, ideological, and temporal motivations influence public policy experimentation. Section 4 provides a concluding discussion of the analysis in line with the research questions posed – reflecting on the role of specific experimentation settings in shaping the effects of political motivations. Section 5 finalises the thesis with propositions for further research, including systematised avenues for future studies and a conceptual revision that could strengthen the analysis of political interference in (up)scaling dynamics of experimental policies.

1 Methodology

The thesis draws on two comparative qualitative studies, where each constituted a separate research track within a larger project (PRG1125). Both studies employ a small-N comparative case-study design ($N_1 = 3$, $N_2 = 2$), utilising in-depth semi-structured interviews and desk research to collect data. The author's ambition to develop an empirically grounded and more systematic understanding of the roles and influences that political motivations have in experimental policymaking justifies the choice of this methodological approach. Accordingly, consistent with Creswell & Poth (2023), Yin (2018) and Van Thiel (2014), the small-N case-study design is especially well suited for developing a scientifically reliable, holistic view of complex social phenomena of intricate and deeply contextualised nature through the synthesis of multiple first-hand perspectives on it.

In terms of general research strategy, the thesis combines exploratory and explanatory analytical lenses – seeking to understand *which* political motivations influence policy experimentation and *how* they influence it.

The two studies that the thesis draws on vary in terms of the author's role and contributions. The first study is the author's independently developed research project where she was responsible for all the steps required to manage, write, and publish the study (**I**, **IV**): from defining research inquiries and analytical frameworks to collecting and analysing empirical data, and preparing manuscripts for publication. The second study is based on team research in which the thesis' author was one of seven investigators and contributed to collecting, preparing, and analysing empirical data, developing the analytical frameworks for the studies (**II**, **III**), reviewing and suggesting improvements to the manuscripts. Each study draws on a comparative case-study where desk research (secondary data analysis, including existing scholarship, practitioner-oriented reports, official websites and related documents, and media releases) complemented primary data analysis collected through in-depth semi-structured interviews (18 and 66 respectively) and follow-up (group) discussions (in the first study).

The first study analyses experiences with piloting a specific active labour-market policy – the *Job Guarantee.* The Job Guarantee refers to a disruptive policy innovation that provides fundamental support for achieving full and meaningful employment by creating voluntary, individually-tailored, and locally beneficial job opportunities that do not compete with the regular economy and provide at least a living wage (e.g., Tcherneva 2019). Conceptually, this policy proposal is understood simultaneously as a structural labour-market reform, an automatic macroeconomic stabiliser, and a wage and working-conditions benchmark for the job market (Tcherneva 2019, 2024). With that, the Job Guarantee challenges current policymaking approaches to tackling (long-term) unemployment implying its *direct elimination* rather than containing its level and mitigating negative implications. The conducted study compares three employment-creation schemes exemplifying the Job Guarantee in real-world contexts, namely: *Territoires zéro chômeur de longue durée* (Zero long-term unemployment territories) undertaken in France (hereinafter: TZCLD); *Modellprojekt Arbeitsplatzgarantie Marienthal* (MAGMA Job Guarantee) in Austria (hereinafter: MAGMA); and *ΚΟΙΝΩΦΕΛΗΣ ΕΡΓΑΣΙΑ* (Of Common Benefit) in Greece (hereinafter: Kinofelis). TZCLD, MAGMA, and Kinofelis were officially labelled "policy pilots", involved experimental evaluation, yet varied in their institutional set-up (governance and implementation of the schemes) and primary purposes. Respectively, the pilots were intended to promote

the approach and reinvent local economies through cooperative public management (TZCLD); to rigorously test the innovation (MAGMA); to mitigate a pressing issue in response to a severe financial crisis (Kinofelis). While at the international level all three programs are being promoted as "successful" (e.g., Markowitsch & Scharle 2024; Tcherneva & Lalucq 2022), the outcomes of their domestic evaluations and adoption dynamics vary and remain ambiguous. Namely, the sample includes an ongoing extended experimentation (TZCLD; 2016-now); a pilot that remained unscaled (MAGMA; 2020-2024, finished); and a pilot that has been scaled up into a regular national program, yet reduced in scope later on (Kinofelis; 2011/6-now, with a limited scope). Therefore, one of the primary aims of the conducted research was to closely examine why this is the case and how political factors influence the adoption prospects and scaling dynamics of this policy solution.

Throughout the thesis, this research undertaking will be referred to as the 'job guarantee' study. The single quotation marks indicate that the programs compared are closely informed by the eponymous concept yet not identical to it. While TZCLD, MAGMA, and Kinofelis all share the key characteristics of the Job Guarantee – such as bottom-up creation of meaningful, locally beneficial, and non-competitive job opportunities ensuring at least a subsistence-level income – some of their features differ from the conceptual 'requirements', including the schemes' operation particularities, funding mechanisms, and temporary nature.

The choice of the 'job guarantees' for the case-study was driven by two complementary considerations. First, methodology- and data-wise, the case allows for an insightful analysis of the 'potentials' that *disruptive* experimental interventions – i.e., those that trial self-contained policy alternatives rather than incremental alterations to conventional policy measures – have for policy learning and policy change. Given the limited practice of piloting radical economic policies in general and labour-market innovations in particular at the time of the study inception (early 2022), it has been highly advantageous to find three ideationally similar experimental programs that are comparable in space and time. Notably, the three programs were the only cases available in the EU context during research inception, whereas by 2026, around a dozen similar pilots have been or are currently undertaken across EU member states, including those that resulted from the ESF+ call for piloting labour-market innovations (March 2024). Along with that, the comparative characteristics of the three pilots (ideationally similar yet different in terms of underlying motivations, institutional set-up, domestic evaluations, and scaling-up scenarios) have been utilised strategically to understand how and why the political effects on the experiments vary. Second, the 'job guarantee' schemes studied exemplify a policy solution with a strong ideological and theoretical background. The solution challenges conventional policymaking approaches to tackling unemployment (e.g., Tcherneva 2019, 2024) and raises strenuous debates regarding the piloting and adoption of the schemes between policy actors of different partisan predispositions as well as mainstream- and heterodox-economics academic societies (e.g., Antonopoulos 2023). Thus, both the relevance of the political dimension and the urgent need to better comprehend it substantiate the choice of the policy case.

The second study compares Estonian and Finnish public-sector officials' experiences with policy experiments. The intention behind case selection was to investigate what (political) factors may explain *why* these experiences differ in two neighbouring countries with similar governance institutions (Estonia and Finland are unitary parliamentary democracies with proportional electoral systems and multi-party coalition

governments). On the one hand, Finland has established the 'experimental culture' during the last decade (Leino & Åkerman 2021), enabling multiple policy experiments to be undertaken – including top-down and strategic/large-scale interventions (in particular, the well-known universal basic income, 2017-2018 – e.g., Wispelaere et al. 2023; Kangas 2021). On the other hand, the experience of Estonia so far has been limited to a number of predominantly small-scale interventions conducted bottom-up (e.g., design experiments on improving the explanatory note of the state budget, small RCTs concerning recycling of packaging).

While pursuing different foci, the studies complement each other in terms of the research contribution offered, enabling the author to address different research gaps. Namely, the 'job guarantee' study analyses how political motivations influence the *already launched* experimental solutions that are disruptive and highly politically contentious – focusing on the evaluation and scaling-up stages of policy experimentation. The study examines such issues as: the role of ideology and strong normative pre-conceptions in experimental policymaking; the role, meaning, and (unresolvable?) ambiguity of experimental policy 'success', and the links between politics and the institutional set-up of experimental interventions. The Estonian-Finnish study, by contrast, allowed the author to explore *ex-ante* variables of policy experimentation by examining the effects of political motivations on its launch prospects and research design (although political effects on policy learning and scaling-up were studied as well). Drawing on a high number of observations (66 interviews), the study explored the opportunities and risks that politics poses for novel policy solutions of a more incremental nature (such as job alternation leave, pre-primary education reform, and tax deductions for household services representing green investments) as well as for radical policy proposals (such as basic income provisioning). As a result, the evidence collected enabled examination of how the clashing timeframes (electoral vs experimental) and the electoral expediency shape different types of experiments in the public sector.

In both studies, semi-structured interviews were employed to collect in-depth empirical data. The 'job guarantee' study draws on 18 interviews with the programs' key stakeholders (5 for *TZCLD*, 7 for *MAGMA*, 6 for *Kinofelis*; 2023-2024) as well as follow-up discussions (2024-2025). The sample includes the initiators, managers, key evaluators, and direct executors of the three programs. To select the interviewees and ensure the representativeness of data, criterion-specific sampling (officials from different public institutions / roles from the pilot) was mixed with the snowball technique (in line with Patton 2015). The analysis has also benefited from the follow-up (group) discussions with the leadership and evaluators of all three programs. The discussions were held in-person (mostly) and via email correspondence[2], enabling the author to verify and advance the inferences. In the Estonian-Finnish case-study, a total of 66 interviews were conducted (32 in Estonia, 34 in Finland; 2022-2023). The sample consists of ministry- and agency-level public officials of different hierarchical positions who were either directly involved in policy experimentation or witnessed the respective experiences (first-person accounts). The interviewees include the representatives of the Prime Minister's Office in Finland, the Government Office in Estonia, fiscal, financial and other public institutions.

[2] In connection with the author's follow-up engagements, including a contribution to the TZCLD book (**IV**) and presentations of tentative and finalised findings of the study on TZCLD, MAGMA, and Kinofelis (2024-2025).

In both studies, the interviews lasted for one hour on average, were recorded and transcribed (*Otter.ai* software). The interview protocols were iteratively adjusted in line with the interviewee's position/role in the experimental intervention and the point of time when the interview was conducted (i.e., in accordance with previously collected data). The interview protocol of the 'job guarantee' study was focused on gathering details regarding the launch, implementation, and scaling dynamics of each of the employment-creation schemes, including the political challenges and risks involved, outcomes of experimental evaluations, scaling-up factors, and expected/fulfilled adoption scenarios. In the Estonian-Finnish study, the questionnaire (a similar interview protocol was designed for both countries) covered the interviewees' personal experiences with policy experimentation as well as their perceptions of experiments as a policymaking tool, attitudes on the role of politics and policymakers in conducting experiments, opinions on the opportunities, pitfalls, and risks involved in the experimental process.

In terms of data analysis, both studies pursue an iterative and interpretive approach consistent with the small-N, constructivist research design applied (Creswell & Poth 2023). Specifically, data collection and analysis were conducted in parallel, each informing the other: the studies combined deductive and inductive strategies – *probing the* factors and inferences identified throughout existing scholarship with empirical data (deductive phase) and discovering *additional* factors drawn from the empirical analysis (inductive phase). To facilitate the analysis of large amounts of data in the Estonian-Finnish case, the respective interviews were coded using *MAXQDA* software.

As a key weakness of the small-N case-study design, literature on research methodology typically **notes** limited opportunities for extrapolating and generalising from the findings of such studies (Creswell & Poth 2023; Yin 2018). The thesis acknowledges this limitation and seeks to address it to the extent possible by emphasising and examining the crucial role of a broader experimentation setting in shaping the effects of the subject matter studied – political motivations. The qualitative nature of the conducted studies constrains findings' generalisability as well – which is linked to the subjective and interpretive character of the empirical data they build on, as well as lack of access to certain interviewees. In particular, in the Estonian-Finnish study, interviews were conducted exclusively with civil servants (rather than elected officials, i.e., higher-level policymakers). In the 'job-guarantee' study, on the contrary, the interviewed stakeholders of Kinofelis did not include public officials directly involved in the implementation of the program at the central and local levels. As a result, the study captured the perspectives of the initiator and evaluators of Kinofelis, as well as ILO contributors involved in the program's operationalisation and upscaling.

Altogether, the studies have produced a series of manuscripts and publications, of which four form the basis of this thesis. In terms of the central research inquiry and the related analytical framework, these studies are structured as follows. **Article I** probes the plausibility of four explanations for why *successful* pilot policies may not be scaled up in practice, synthesised from the existing scholarship (literature on policy-piloting and social innovation, policy experimentation, and evidence-based policymaking in general; multiple policy domains). With that, the study explores which factors shape the scaling-up of *successful* policy pilots and how. Empirically, the study draws on the 'job guarantee' study. **Articles II** and **III** pursue a similar (to **Article I** and each other) analytical strategy while drawing on the data of the Estonian-Finnish case-study. The studies probe the assumptions developed based on the literature (respectively: blame avoidance and

credit claiming and experimental policymaking studies; literature on public policy, electoral politics, mediatisation of politics and experimental policymaking), investigating how motivations of blame avoidance and credit claiming (**II**) and political timeframes (**III**) influence experimental policymaking, and which contextual factors shape them (both). **Book Chapter IV** ('job guarantee' case-study) utilises the adapted framework of 'governance design' to study how the way policy pilots are operationalised affects their scaling-up trajectories. It first identifies the key dimensions of *design choices* that differentiate the programs compared and then traces how these *choices* affect(ed) the experimentation and scaling-up in each case.

The overall methodological strategy of the thesis is summarised in Table 1, itemised by publication.

Table 1. Methodological approaches adopted in doctoral publications

Publication	Research type and focus	Data collection	Data analysis	Research inquiry	Analytical approach
Article I	Comparative case-study (N = 3); 'job guarantees'	Semi-structured interviews (18), follow-up discussions, desk research ('white' and 'grey' literatures, broader secondary data)	Iterative and interpretive approach (in parallel with data collection); deductive-inductive strategies	To explore what factors shape scaling prospects for *successful* experimental policies and how	Probing plausibility of integral theoretical expectations developed based on the scholarship
Article II	Comparative case-study (N = 2); public officials' experiences with policy experimentation in Estonia and Finland	Semi-structured interviews (66), desk research ('white' and 'grey' literatures)	Iterative and interpretive; deductive - inductive; coding in MAXQDA	To explore how and why blame avoidance and credit claiming motivations influence experimental policymaking	Probing plausibility of theoretical expectations synthesized from the scholarship
Article III	Comparative case-study (N = 2); public officials' experiences with policy experimentation in Estonia and Finland	Semi-structured interviews (66), desk research ('white' and 'grey' literatures)	Iterative and interpretive; deductive - inductive; coding in MAXQDA	To explore how political timeframes influence experimental policymaking	Probing plausibility of theoretical expectations synthesized from the scholarship

Publication	Research type and focus	Data collection	Data analysis	Research inquiry	Analytical approach
Book Chapter IV	Comparative case-study (N = 3); 'job guarantees'	Semi-structured interviews (18), follow-up discussions, desk research ('white' and 'grey' literatures, broader secondary data)	Iterative and interpretive; deductive - inductive	To explore how the institutional set-up of policy experiments influences their adoption prospects and scaling dynamics	Using the framework of 'governance design' to compare the risks and opportunities involved in institutionally different policy experiments

Source: Author

2 Analysis: How Political Motivations Shape Public Policy Experimentation

This section explores how the key political motivations identified throughout the conducted studies – strategic, ideological, and temporal – influence public policy experimentation across its different stages. The three groups draw on the synthesis of empirical findings from the two comparative case studies (as introduced in Section 2) and insights from existing policy science literature. Strategic, ideological, and temporal motivations all originate in policymakers' intrinsic aspirations (Deci & Ryan 2000; Besley 2007; Strøm & Müller 1999; Callander 2008) yet differ in the kind of aspirations each represents. Namely, policymakers' aspirations to maximise *private* political advantage in terms of power and reputation are gathered under the umbrella of strategic motivations. Aspirations to act in line with partisan commitments and normative beliefs make up the group of ideological motivations. Aspirations to align decisions with political timeframes and their anticipated disruptions – policy and political changes – fall under temporal motivations. Subsections 3.1-3.3 zoom in on the effects each group has on policy experimentation, while Subsection 3.4 examines the groups' interplay, elucidating how the broader experimentation context mediates and moderates these effects. Unlike political motivations which are intrinsic to policymakers, the context of policy experimentation encompasses the factors of a specific 'operational environment' where it is or can be undertaken. The contextual factors determine whether and how political motivations become activated, how much weight they carry among the decision-making grounds of policymakers, and which will dominate among these grounds.

2.1 Strategic Motivations: Reputation- and Power-Maximisation

Private gains and losses are inherent in public-oriented decisions, and there exists ample evidence that policy actors weigh one over another when choosing between policy alternatives (e.g., Esaiasson & de Fine Licht 2024; Leong et al. 2023; Capano & Lippi 2017; Twight 1991). Policymakers' strategic[3] interests – i.e., their natural aspirations to maximise net political advantage from the decisions taken – are well-recognised among the key factors that shape both the process and the outcomes of public policy formation (e.g., Britz et al. 2022; Xi et al. 2018; Gersbach 2004). Yet, there remains a lack of evidence on their role in public policy experimentation. The thesis helps to address this gap. By synthesizing the inferences from existing scholarship with the findings offered by the case-studies, the paragraphs below elucidate how the strategic interests of policymakers inform their choices *for* or *against* experimenting with novel policy solutions (**II**, **I**, **IV**).

To make the analysis more precise, the thesis distinguishes between reputation- and power- maximisation as two separate strategic interests that inform policymakers choices regarding policy experiments. On the one hand, policy sciences literature characterises policymakers as actors who seek to maintain or enhance their 'electoral appeal' by receiving praise for the actions taken, refraining from unpopular decisions, or minimising the risk of public accusations for potential policy failures (Schönhage et al. 2024; Wenzelburger 2014; Giger & Nelson 2011). To analyse the effects of the

[3] The term is close in meaning to 'opportunistic interests', yet it does not contain a negative connotation of the latter: the issue of whether policymakers' aspirations to maximise private advantage work *in line with* or *against* public value maximisation goes beyond the research inquiry of this thesis.

reputational interests, the thesis utilises the lens of *blame avoidance-credit claiming calculus* referring to the 'ratio' of perceived reputational gains (opportunities for ascribing credit) and losses (risks of being blamed) that policymakers associate with a certain policy choice, as developed in **II**. This conceptual frame helps to understand why policymakers do (or do not) support experimentation or favour one experimental intervention over another. On the other hand, policymakers tend to be interested in maintaining or enhancing their political power in terms of available authority and budgets, by preserving the status quo, consolidating disposable funds, or expanding decision-making authority (e.g., Ampe et al. 2021; Avelino 2017). The following paragraphs first examine the effects of each motivation on policy experimentation and then provide an overarching conclusion on how the two motivations together may shape it.

Reputation-maximising motivations

The analysis demonstrates that policymakers are likely to discard those experimental interventions whose blame-credit calculus is regarded as negative or 'too low', as it would mean the interventions pose risks to their reputation and/or re-election chances (**II**, Bauknecht et al. 2020; Raudla et al. 2023). Alternatively, the preference is likely to be given to experimentation of a more incremental nature – i.e., the one that has a smaller scope, shorter timeline and/or entails smaller institutional changes. In this case, the reputational risks of potential policy failure tend to be much lower, making such interventions politically safer to undertake (**II**, Hood 2010; Rozin & Royzman 2001). Both observations are importantly linked to the *negativity bias* that proves to characterise policymakers (Ampe et al. 2021; Rodriquez 2022). The term denotes a psychological tendency to attribute more weight to potential losses than potential gains, i.e., prioritise minimisation of the former over maximization of the latter (Kahneman & Tversky 1984). To underscore the influence of this bias on political decision-making, the motive of avoiding blame is intentionally put first in the calculus.

In the Estonian-Finnish study, the blame-credit calculus generally ascribed to public policy experimentation varies starkly between the countries. In Estonia, the failure of experiments is perceived as a blame-generating outcome, drawing policymakers' attention away from the potential benefits such interventions could deliver. As a result, views on experimental policymaking as "a waste of public resources" or "too politically risky undertakings" dominate the perception among Estonian policymakers (**II**). In Finland, a contrasting narrative prevails: the failure of policy experiments there is seen as an opportunity to learn and save the public budget from potential larger losses. Consequently, Finnish policymakers tend to consider experimentation as a *political* tool for garnering credit by demonstrating to the electorate that their concerns are being addressed in a smart and cautious way (**II**).

The study on 'job-guarantees', in turn, elaborates on how the political calculus may influence the upscaling of already launched experimental programs (**I**, **IV**). In the case of TZCLD, interviewees were concerned that the Labour Minister would argue against extending the scheme, the implementation of which relied on effective collaboration between *Left* and *Right* policy actors, as this could disprove one of his core electoral messages: namely, that the two political sides cannot work together. Consequently, legitimising the program as permanent at the national level could expose this contradiction to the broader public and thereby undermine the Minister's credibility.

By contrast, the case of Kinofelis illustrates that policymakers' aspiration to avoid electoral accusations may work in favour of politically unpalatable policy innovations,

reversing the 'sign' of an otherwise negative political calculus. Specifically, policymakers' fear of making unpopular decisions and losing voter support, amplified by the negativity bias, helped to retain the program adopted under previous leadership. Although the new Ministerial office was openly opposed to *Kinofelis*, it could not discard the program largely due to its high popularity among the Greek public. In other words, abolishing the program proved more politically costly than sustaining its (limited) enrolments (**I**).

Power-maximising motivations

Regarding the second type of strategic interests, the thesis indicates that policymakers tend to oppose experimentation when it threatens their status quo or provides no opportunities for expanding their current authority (**I**; Natili 2019*;* Hong & Lee 2018; Nair & Howlett 2016). Such opposition may stem from the need to reallocate available resources (authority and/or funds) in favour of experimental institutions and/or experimenters. In particular, two types of resource reallocation may be required: from conventional organisations to newly established experimental institutions and from central-level to local-level authorities. Both types may lead experimentation to clash with objections from the 'losing side' (**I**, **IV**). For instance, policymakers whose decision-making autonomy or areas of control risk being compromised by experimentation may try to weaken experimenters' positions by discarding, reversing, or interrupting the interventions or by preventing their scaling-up (**I**, **IV**; Nesti et al. 2025; Casula 2022). As the TZCLD stakeholders emphasised, scaling-up the scheme entails an increase in the autonomy of local actors at the expense of operational resources available to the Ministry of Labour (hereinafter: MOL) – such as disposable budget and control over the French territories involved in experimentation (in the TZCLD context, a "territory" denotes an administrative unit (of experimentation) of up to 10,000 residents). Consequently, the MOL's fear of losing part of its political advantage due to the cuts in funds and area of control made it reluctant to support the project's permanent adoption after the second experimental term (2021-2026) (**IV**).

Strategic motivations and policy experimentation: general impact

Summing up, the conducted analysis indicates that, ceteris paribus, experimental policy interventions are unlikely to be launched or upscaled unless the strategic incentives the experimental process entails *sufficiently* outweigh the costs it imposes on policymakers. The costs associated with experimentation may be linked to the institutional changes mentioned above, including the needs to arrange complex funding processes, establish novel entities and collaborations, reallocate operational resources, and/or revise the rules of governance (e.g., Raudla et al. 2025; Sager & Gofen 2022; Binz et al. 2016; Raven et al. 2019). Altogether*,* these demands are likely to disturb the convenience of policymakers' status quo, who may withdraw their support for experimentation in the absence of strong reputational *or* power-related incentives (**I**; Nair & Howlett 2016). The experiences of Austrian civil servants piloting labour-market innovations, including MAGMA (representatives of AMS Lower Austria, a local public employment service; hereinafter: PES) corroborate this point. The interviews have revealed that several experimental solutions preceding MAGMA proved effective as well, yet were similarly discontinued after the pilot phase, as the higher-level policymakers responsible for the final decisions (MOL, state-level PES) lacked incentives to allocate the limited funds for scaling them up.

Along with that, the thesis points out that policymakers' strategic motivations may shape the *primary* purpose of experimentation, which may diverge from the officially declared one ("to test/learn"). The analysis of the case-studies and existing scholarship

has shown that policy experiments may be launched to mitigate political objections, achieve greater social acceptance for novel solutions, ensure that these solutions align with the interests of certain policy actors, or signal proactiveness and responsible policy engagement to the electorate (**I**, **II**). In such contexts, the prospects of the experimental solutions tend to be predefined already at the stage of their inception – suggesting that in policy experimentation practice, scaling-up may be both a *pre-planned* and an *unplanned* deliverable (**I;** in line with Neuwinger 2022).

2.2 Ideological Motivations: Preferences and Commitments

Policymakers' ideological motivations – i.e., their natural aspirations to support those policy choices that resonate with the ideas, values, and beliefs they hold (e.g., Garcia-Hombrados et al. 2024; Wolak 2020) –constitute another driving force of public policy experimentation. This claim, however, contradicts the dominant understanding of experiments in early policy sciences literature, where they were typically framed in line with the technocratic doctrine: as "scientific and objective tools, independent from ideological meaning" (Parth & Nyby 2022, p. 774, referring to Rogers-Dillon 2005; Hallsworth 2011). Although the influence of such conceptualisations can still be observed in some recent studies (e.g., see Varazzani et al. 2023; Jones & Whitehead 2018; Ruggeri et al. 2020), where experiments are understood primarily as *research* rather than *political* undertakings, the thesis argues that policymakers' decisions regarding policy experimentation are subject to their partisan preferences and commitments and can hardly preserve ideological neutrality. Rather, it is a question of prioritisation: whether evidence-based considerations or other motivations prevail over policymakers' ideological predispositions in a given setting (**I**; Esaiasson & de Fine Licht 2024; Raudla et al. 2023). This subsection illustrates this argument by identifying three ways in which ideological motivations may influence public policy experiments at their different stages.

First, policymakers' ideological commitments may directly motivate experimental interventions or serve as key obstacles to their launch. As existing scholarship demonstrates, the role of ideology in political decision-making tends to vary across policy domains, being substantially higher in those that deal with contentious societal issues or involve highly polarised political opinions (Roosma 2022; Wolak 2020; Heilmann 2008; Strøm & Müller 1999). Experimentation in such domains is likely to be more explicitly ideology-driven and, consequently, more difficult to undertake due to greater political obstacles. In particular, the launch of experiments with ideologically contested solutions may be blocked if they appear incompatible with the political identity of the ruling party, authorised policymakers, or particular voter groups (**I**; Nesti et al. 2025; Raudla et al. 2023). Taxation, welfare, and employment-creation policies exemplify this dynamic, as the innovative solutions they encompass or the challenges these solutions promise to address are often subject to fundamental disagreements between political sides. Literature illustrates that these disagreements may emerge from different political grounds: including party vs. party, governing coalition vs. opposition, and municipal/local vs. central government tensions over ideas and agendas (Parth & Nyby 2022; Cairney 2022; Jessoula & Natili 2020). In such political setting, experimentation may act as a strategic attempt to oppose the prevailing policy course and/or reframe it with alternative solutions, implying that experiment initiators are ideologically committed to the solutions they 'test' (in line with Nesti et al. 2025; Jessoula & Natili 2020; Sacchi & Bastagli 2005). The job-guarantee case-study supports this proposition. The primary purpose of experimentation differed across the studied interventions – reflecting

different 'default states' of the initiators' resource autonomy (see the second argument). Yet, in all three cases, the initiators shared a common commitment to voluntary, individually tailored employment-creation, viewing this approach as an effective alternative or complementary measure to conventional public employment services (I). That is, all three policy pilots were, to a certain extent, launched *because* the initiators anticipated pilot success on the grounds of their ideological (normative or ideational) commitment to the novel policy solution.

At the same time, the undertaken doctoral research interestingly shows that the *lack* of ideological commitment to innovative policies may *also* motivate the launch of experimental interventions. In particular, policymakers may initiate experiments to demonstrate the ineffectiveness of unpalatable policy solutions, and thus, acquire formal grounds to discard them in favour of an already chosen political trajectory (Wispelaere et al. 2018; Ettelt et al. 2015b; Nesti et al. 2025).

These two contrasting scenarios, however, support the same conclusion: policymakers' stance toward the experimental solution, whether favourable or not, proves to be an important driver of their engagement with experimentation – rather than the intention to genuinely test and learn. In other words, the doctoral research demonstrates that policy experiments tend to be conducted against pre-existing normative expectations rather than substantive uncertainty about the merits of the 'tested' solutions.

Second, policymakers' pre-existing attitudes towards novel policy solutions may influence both the evaluation and interpretation of experimental outcomes. Policy sciences research typically depicts policymakers as opinionated actors who tend to prioritise personal hunches over expert knowledge when deciding the fate of policy measures (e.g., Bravo-Biosca 2020; Boswell 2009; Cairney 2012). The thesis' findings confirm this pattern for public policy experimentation (**I**, **III**, **IV**; Mattocks 2026, 2025; Raudla et al. 2023). Following the inference above, the thesis argues that policymakers' ideological commitments create normative expectations against which the success of experimental policies will be judged. Two contrasting examples from the 'job guarantee' study illustrate this statement.

Specifically, the studied employment-creation programs indicate that the ideological bias of experimental evaluation may work both in favour of and against the adoption of novel policies (**I**, **IV**). In the case of Kinofelis, the initiator's commitment to the pilot policy influenced the role and focus of the experimental evaluation. Since the initiator had sufficient resources to decide the program's future autonomously (Kinofelis is a top-down intervention initiated by the MOL), the evaluation was assigned to collect evidence on *how* rather than *whether* the scheme works and to strengthen its implementation design before nationwide adoption. Thus, the label "policy pilot" used to describe the program (e.g., ILO 2018) in practice referred to the process of prototyping an innovative solution that had already been selected for its subsequent scaling-up.

The experiences of MAGMA and TZCLD, in turn, illustrate the opposite case: one where the scaling-up of pilot programs depends on the will of external decision-makers (i.e., higher-level authorities such as MOLs and national-level PES) who are either sceptical (the case of TZCLD) or openly opposed (the case of MAGMA) to the experimental solutions. In particular, the lack of ideological support for the idea of public employment creation combined with low strategic incentives left experimental evidence produced by internal evaluations of both programs (which showed them to be effective) unable to prove the programs' success to the policymakers in power. In the case of

MAGMA, rigorous evidence on the pilot's success (the evaluation was based on a pre-registered RCT conducted by reputable scholars) failed to convince the conservative government. Policymakers criticised the "clearly social-democratic idea" using their own alternative evaluation, which shifted the focus from the socio-economic effects generated (as estimated in the pilot) to the financial efficiency of the program. Ultimately, the conservative government declared the program "costly" and made this assessment public via the local media (**I**). During the same timeline, however, the success of MAGMA was recognised at the international level. The program became instrumental in shaping the argument *for* the Job Guarantee made by the United Nations in 2023 and informing the call for piloting employment-creation schemes across the EU made by the European Social Fund Plus in 2024 (**IV**; Markowitsch & Scharle 2024). In the TZCLD case, the outcomes of internal evaluations – which consistently indicated the scheme's success – diverged from the conclusions made by the Scientific Committee, the official evaluator of the program established under the Experimental Law (loi d'expérimentation territoriale No. 2016-231). While the primary goal of the Committee was to ensure political neutrality of experimental evidence, the interviewees emphasised that the central government, which did not favour permanent adoption of the scheme, was empowered to influence it by choosing what the Committee should and should not include in the program's cost-benefit calculus. In particular, the approved evaluation framework did not account for positive externalities that the program generated, such as social reintegration, territorial cohesion, and locally created economic activity, thereby depicting it as less efficient than it might actually have been (**IV**).

With that, the thesis demonstrates that the success of policy experiments is at least *as much* 'measured' by political preference *as* by experimental evidence. In other words, it is hardly possible to separate the framing of experimental evidence from the ideological or ideational preconceptions of policymakers who *judge* it. The lack of unified guidelines upon *what* should be evaluated and *how* provides policymakers with 'rooms' for speculation over experimental evidence irrespective of its rigour – either at the stage of experimental evaluation (evidence production) or while interpreting its outcomes for the public (evidence dissemination) (**I**, **IV**; in line with Mattocks 2026; Ruggeri et al. 2020; Strassheim & Kettunen 2014). Consequently, if by the end of the pilot term an experimental policy clashes with a dominant political agenda at the national level or at the level of the experimenting institution (e.g., municipal or territorial ones), the issue of power-based evidence prioritisation may undermine its potential to convince authorised policymakers of its success and get a pass for upscaling in any form (**I**, **IV**).These findings question the capability of experiments to generate policy change in the settings of low political support – suggesting that they rather serve as auxiliary tools for designing politically palatable innovations than means to enable shifts in ongoing policy agendas (in line with Mattocks 2025; McFadgen & Huitema 2017, 2018; Ettelt et al. 2015b).

Third, the diverging party leanings of policy actors may significantly affect the adoption of experimental programs, both facilitating and hindering it. Empirical studies on policy innovation and innovation governance illustrate that innovative policy projects often require establishing new (in)formal ties and collaborations between various actors (Sager & Gofen 2022; Verkerk & Buuren 2017; Casula 2022; Smith et al. 2022). In this process, the heterogeneous political identities of potential stakeholders may present a crucial obstacle to governing, scaling-up, and/or diffusing innovations (Ettelt et al. 2022; Casula 2022; Leino & Åkerman 2021). The more diverse potential stakeholders are across the political spectrum, the lower the chances that experimentation managers will find

common ground to motivate all counterparts to collaborate (Perkiö 2020; Turnheim & Sovacool 2020). The findings of the 'job guarantee' case-study both corroborate and challenge this point through the TZCLD case.

On the one hand, the program elucidates that the political fragmentation of an experimenting country – i.e., the presence of diverse party affiliations across its governmental levels and administrative units – may challenge the expansion of experimentation. Specifically, in order to establish *Local Employment Committees* and *Companies for Employment Purposes* that would respectively supervise and undertake job creation at the territorial level, the central management of TZCLD had to engage both left- and right-wing policy actors in experimentation and ensure both sides could effectively steer it together (**IV**).To reconcile the diverging political interests of potential counterparts, the scheme's principal benefits were framed in line with their respective ideological priorities: the project's societal outcomes were underscored for the *Left*, while the empowerment of local actors through larger resource autonomy that the innovative approach entails was emphasised for the *Right* (**I**). The importance of narrative diversification while introducing experimental programs or their outcomes to different policy actors resonates with the insights from recent studies on legitimation of disruptive innovative policies, such as universal, minimum, and basic income guarantees (Roosma 2022; Smith et al. 2022; Perkiö 2020).

On the other hand, the diverging political preferences of local policy actors may provide strategic opportunities for policy experiments over a longer timeline. As demonstrated with TZCLD, developing collaborative capacities and a sense of shared commitment between political sides can crucially enhance program resilience to electoral dynamics and resulting changes in government agenda. In particular, the political diversity of TZCLD stakeholders helped the program neutralise its initially *Leftist* affiliation, build an effective advocacy coalition, and ultimately achieve unanimous parliamentary support for its extension for a second (ongoing) experimental term (**I**, **IV**).

2.3 Temporal Motivations: Alignment with Political Timeframes

Policy experimentation is a process with its own temporality which, in democratic governance systems, may overlap and clash with the timeframes followed by policymakers. Literature on electoral budget cycles provides ample evidence that such political timeframes as electoral cycles and government terms in office shape the way policymakers think of, design, and adapt policy interventions (e.g., Bueno 2023; Howlett & Goetz 2014; Goetz 2014; Linz 1998). Drawing on I, III, IV and the synthesis of insights from diverse studies, the thesis demonstrates that similar inferences hold true in the context of experimental policymaking. Specifically, the analysis indicates that policymakers tend to align their policy choices with the dynamics of regular political life. These dynamics encompass both the political timeframes as such – including electoral cycles, governments' terms in office, and leadership turnover within experimenting institutions – as well as the anticipated disruptions between them, i.e., policy- and political changes at the juncture of these cycles. Such alignment is found to permeate experimental policymaking across its different stages – affecting the chances of experiments being launched, their research design, and the extent to which policymakers (are eager to) learn from them.

First, the findings confirm the general tendency of policymakers to align policy measures with anticipated electoral dynamics to make them "electorally expedient", as observed in the literature on electoral budget cycles (e.g., Bueno 2023; Goetz 2014;

Katsimi & Sarantides 2012; Mink & De Haan 2006). This may include deciding whether and when to start an experiment, for how long to implement it, and whether to speed it up. In the context of public policy experimentation, the pursuit of electoral expediency is found to push policymakers into making *opposite choices* regarding the launch and design of experimental interventions (**III**; Bernecker et al. 2021).

On the one hand, the pressures of political timeframes may motivate policymakers to opt for those experimental interventions that do not span beyond the current electoral cycle (**III;** Bueno 2023; Corduneanu-Huci et al. 2021; Nair & Howlett 2016). As the Estonian-Finnish study illustrates, the short-term nature of political temporalities tends to hinder the implementation of large-scale policy experiments by discouraging policymakers from launching such projects at all (as the Estonian case shows) or by 'pressuring' them towards designing shorter experimental interventions and prioritising experiments that can be completed within the ongoing term of government (as the Finnish case elucidates). Such 'hastening design practices' (**III**; Checkland et al. 2021; Nair & Howlett 2016), in turn, may negatively affect both the credibility of experimental evidence and the perception of it. While the need to generate results quickly may undermine the quality of conclusions, political circles' *awareness* of this may make policymakers sceptical about the significance and validity of the effects captured. As a result of both or either, the electoral time-pressures are likely to discourage policymakers from giving experimental evidence sufficient weight in the decision-making process **(III**; Parth & Nyby 2022; Perkiö 2020).

On the other hand, the findings indicate that electoral expediency may motivate policymakers to extend an experimental timeframe beyond the current electoral cycle or their term in office (supported by the Finnish empirical data, **III**; Bernecker et al. 2021; Callander & Hummel 2014). This contrasting observation draws attention to the crucial role of context in which policy experiments are (conceived to be) undertaken. In particular, policymakers' expectations regarding their chances of being re-elected may influence what electoral expediency means for them in a given setting. As existing scholarship indicates, experimentation may be undertaken as an attempt to preserve a preferred policy measure in the next electoral cycle, or a strategy to minimise electoral losses by delaying unpalatable reforms of larger scale beyond the office term of its initiators (**III**; Callander & Hummel 2014; Corduneanu-Huci et al. 2021; Nair & Howlett 2016).

The 'job-guarantee' study, in turn, brings to the fore the issue of political dynamics – elucidating how political changes influence the scaling of already launched experimental programs. In particular, the findings illustrate that for both bottom-up and top-down experimental programs that span a junction of two political timeframes (such as electoral cycles, government terms, or leadership turnovers occurring at the level of experimenting agencies), the start of the new political timing proves to change, and, most likely, challenge, their chances to be sustained (**I**, **IV**; Guerineau et al. 2024). Accordingly, in all the cases studied (TZCLD, MAGMA, Kinofelis), the initiators of the programs took measures to adapt them to an anticipated political change depending on the expected 'direction' of the latter, i.e. in favour or against public employment-creation schemes.

In particular, for MAGMA (bottom-up policy pilot initiated by local PES), the anticipated victory of social democrats at the 2024 Austrian legislative elections would have provided a favourable agenda to resume the program. In this light, the MAGMA initiator who changed his occupation during the piloting – from the agency's managing

director to a politician, a regional leader of the Social Democratic Party – was prepared to push for the program's renewal and further extension in case social democrats (centre-left Social Democratic Party) would have won the majority votes in Lower Austria (this scenario, however, did not materialise, as the far-right Freedom Party of Austria won the elections). The related change in the agency leadership, however, exposed the risk that the efforts to resume the program might be blocked by the new managing director with right-leaning preferences (**I**).

Conversely, for TZCLD and Kinofelis, the start of the new political cycle was associated with the risk that the pilot programs would lose their political appeal (**I**, **IV**). To mitigate the risk, the initiators in both cases resorted to formally legitimising the undertaken schemes, yet the legitimation was employed at different stages of piloting and with different purposes. In the TZCLD case (bottom-up experimentation initiated by a group of NGOs), various-level elections (legislative, presidential, and European) were expected to influence the political appeal of the scheme for the central government – shaping its support for the program accordingly. To minimise the risk that the project would be discarded for political reasons, its two phases (2016-2021, 2021-2026) were enacted via two respective *Experimental Laws* (Law No. 2016-231 and Law No. 2020-1577 respectively) designed and approved by the Parliament. The Laws specified the terms and conditions of experimentation enabling its extension to new French territories on the one hand and protecting it from downscaling or premature termination on the other – in the event of changes in the central government's agenda.

For Kinofelis (top-down pilot initiated by the ML), a legal mandate was utilised to anchor the scaled-up program among regular PESs imposing additional costs to its reversal under successive Ministerial office(s). As a result, while the changes in government, from Left to Right, and in program leadership deprived Kinofelis of its political support, reducing its scope and quality of enrolments, the program was not terminated. Combined with high support from the Greek population, the efforts required to close the program ultimately discouraged policymakers from doing so.

Altogether, the conducted analysis indicates that the deep-rooted tensions between experimental and political timeframes impose constraints on the *capacity* of policy experiments to generate evidence-driven policy change. The short-termism of political horizons challenges the potential of experimentation findings to inform policymaking practice in two ways. On the one hand, electoral pressures may push policymakers to hasten experimental evaluation at the expense of its quality or to decide the fate of experimental policies prematurely – i.e., before their evaluation is completed (**III**, **I**; Sondal et al. 2024). On the other hand, expectations that experimental evidence will be disregarded at the onset of a new political cycle may become self-reinforcing by entrenching the perception of policy experiments as 'solutions in themselves' among policy actors (**I**). This perception is likely to influence the design choices of experimenters, leading them to prioritise short-term outcomes of the interventions over their longer-term effects. In such setting, little room remains for the ambition to experiment with novel solutions that defy conventional policy(making) approaches and practices.

2.4 Interplay of Political Motivations: Role of Context

As the conducted analysis demonstrates, the explored strategic, ideological, and temporal motivations are never 'siloed' analytical categories but interplaying decision-making drivers whose ultimate effects on experimentation depend on its specific setting. In other words, the context of policy experiments proves to shape the political calculus

underlying policymakers' decisions around them. The findings obtained from the 'job-guarantee' and Estonian-Finnish case-studies agree on this point with evidence from existing scholarship – which, however, is scattered and infers the point mostly indirectly (e.g., see Mattocks 2025; Wispelaere et al. 2023; Jessoula & Natili 2020; Cinar et al. 2019; Spicer et al. 2014). Altogether, the three groups of contextual factors are found to shape the interplay of political motivations, including: the institutional set-up of experimental programs (I, IV), policy domain to which they pertain (I, III, IV), and the legitimacy of policy experimentation, encompassing both the legal frameworks that enable policy actors to experiment and the general electoral endorsement of experimentation as a valid policymaking strategy (II, IV). These contextual factors prove to mediate and moderate the effects of political motivations – explaining why and how the latter become activated in the first case, and defining their significance, 'sign' (i.e., whether the motivations facilitate or hinder experimentation), and the extent to which policymakers can act on them – in the second.

First, the institutional set-up of experimental programs, i.e., the *rules* defining how such programs are operated and governed, is found to mediate and moderate the effects of the strategic and ideological motivations on their adoption prospects (**IV**, **I;** in line with McFadgen & Huitema 2017, 2018; Capano & Lippi 2017; Spicer et al. 2014). More specifically, as the 'job-guarantee' study elucidates, the *modes of launch* of experimental interventions and the *types of institutional networks* employed to operate them condition the clashes of political interests that may arise *between* the involved policymakers and *among* them **(I**, **IV).**

On the one hand, scaling-up policy pilots is found to be much more politically challenging when they are launched *bottom-up*, i.e., by grassroot actors (e.g., NGOs, politicians, or social movements) or local-level agencies (e.g., municipal/territorial public agencies and innovation labs). Unlike top-down interventions, such experimental programs are characterised by the split in decision-making autonomy (over available authority, funding, or both) *between their initiators* and *higher-level policymakers* authorised to define their prospects. As a result, bottom-up policy pilots 'that work' typically cannot be scaled up based on the ideological commitment of the experimenters – unless the same commitment is present on the side of higher-level authorities or the absence thereof is compensated with sufficient strategic stimuli available to them. In particular, the lack of both – ideological commitment to experimental policies and the strategic benefits they offer – left the MAGMA pilot unscaled despite the program being found successful according to experimental evaluation, and complicated accessing funds to expand experimentation to new territories in the case of TZCLD (respectively: the programs were initiated by local-level PES and by NGOs and politicians) (**I**, **IV**). With that, the thesis illustrates that the same ideological commitment has different leverage over experimentation outcomes for bottom-up and top-down experimentation. Namely, the inability to garner political endorsement from higher-level policymakers constitutes a recurring explanation for why policy experiments that proved successful according to internal/official evaluations remain unscaled (**I**, **IV**; Beisheim et al. 2025; Nesti et al. 2025; Sondal et al. 2024; Grin 2020; Checkland et al. 2021; Vreugdenhil & Rault 2010).

On the other hand, the study indicates that those experimental programs that build on new (experimental) rather than established institutional networks entail a much greater potential for political tensions that may undermine their implementation and upscaling (**I**, **IV**; Sager & Gofen 2022; Casula 2022; Smith et al. 2022). As the comparison of the three 'job guarantees' reveals, by relying on existing networks of entities and

actors, experimental programs are likely to avoid two types of political tensions: ideological and resource-related. The first may arise among policy actors of different party affiliations whose buy-in is needed to deliver the programs. The second, in turn, may emerge between the existing and newly established, experimental entities when the political resources (authority and/or funds) are to be (re)allocated in favour of the latter. Resolving the ideological and resource-related tensions may be especially challenging when the functions of experimental (newly established) and existing institutions substantially overlap, when experimentation is conducted in highly politically fragmented countries, or when the experimental solutions are highly ideologically contentious per se (**I**, **IV**). Yet at the same time, the programs that develop operational networks during and as a part of experimentation may end up being more politically resilient in the longer term compared to those that draw on existing entities and arrangements. As the 'job guarantee' study demonstrates, establishing new collaborations among diverse policy actors may ultimately help mitigate existing political disagreements and neutralise ideological biases regarding experimental programs, such as the programs' perceived 'party affiliations', thereby enhancing their upscaling prospects (**IV**).

Second, the domain to which novel policy solutions pertain is found to shape the level of political costs associated with policy experimentation and, consequently, moderate the relevance of ideological aspirations among policymakers' decision-making grounds. Specifically, the doctoral analysis indicates that the level of political costs associated with potential experimentation – measured in potential losses of electoral support, resource autonomy, or 'steering convenience' – varies across policy domains, being systematically higher in those where the issues of ideological disagreements and ethical sensitivity come to the fore (**I**; Wolak 2020; Carmines et al. 2012; Strøm & Müller 1999). The policy domains of labour-market, welfare, law, and taxation serve as vivid examples of politically costly 'application areas' where experimentation remains a relatively scarce experience (e.g., as compared to health, environmental, and administrative policy domains where the ideological and ethical misalignments are limited or easier to address). For instance, experiments with legislation and tax regimes risk violating citizens' basic rights to equal treatment and legal certainty, as well as challenging the core political choices such as 'more state vs. less state' and 'equality vs. liberty' (e.g., Eneqvist 2022; Raudla et al. 2023; Abramowicz 2019). In the labour-market- and welfare policies, in turn, the concerns of ideological nature – both normative and ideational – stand out. Experimental efforts in these areas face multiple clashes of interests, including those between different parties and political sides as well as between the proponents of mainstream and alternative conceptual approaches to poverty and unemployment. As a result, the related political debates in these domains complicate launching new policy pilots and scaling up those already undertaken (**I**, **IV**; Tcherneva 2024; Antonopoulos 2023; Parth & Nyby 2022; Perkiö 2020; Natili 2019).

Moreover, certain policy domains are inherently more associated with disruptive policy solutions, i.e., those that defy established institutional arrangements and conventional policymaking practices. Experimentation with disruptive solutions poses higher risks for the status quo and reputation of policymakers which, consistent with their risk-aversion and negativity bias (Ampe et al. 2021; Rodriquez 2022), further reinforces the uneven distribution of political costs across domains. Altogether, the foregoing considerations suggest that ideological commitment to novel policy solutions plays a decisive role in domains that tend to be more politically costly by default – serving

as a primary driver for top-down experiments, and an enabling prerequisite for bottom-up interventions, where gatekeepers' endorsement of experimentation allows initiators to access resources for its launch.

Third, the analysis illustrates that the legitimacy of policy experimentation both mediates and moderates the interplay of policymakers' strategic and temporal motivations – directly affecting their willingness to engage with it (**I**, **II**, **IV**; in line with McFadgen & Huitema 2017; Eneqvist, E. 2022; Spicer et al. 2014; Ranchordás 2013). The thesis distinguishes between the perceived and formal 'legitimacies' of experiments – corresponding, respectively, to the input and throughput types of democratic legitimacy in Scharpf's (1999) and Schmidt's (2013) terms. The former (perceived legitimacy) refers to the general electoral acceptance of experimentation as a valid policymaking practice (Scharpf 1999; Eneqvist 2022; Leino & Åkerman 2021). The latter (legitimation) spans existing legal frameworks that enable different policy actors (local- and national-level public institutions, NGOs, politicians) to experiment, as well as the 'rooms' in general legislation that allow such projects to be legalised on an ad-hoc basis, reflecting the procedural and institutional conditions that render experimental policymaking accountable and transparent (Schmidt 2013; Philipsen et al. 2021; Ranchordás 2013). Together, the perceived and formal 'legitimacies' create a backdrop against which policy experiments are (if so) conducted. Still, although interconnected and mutually reinforcing, each proves to shape the identified political influences differently.

On the one hand, public endorsement of policy experiments may enable policymakers to glean credit from experimenting with novel policies, especially when the risk of policy failure is high. As the analysis demonstrates, in the contexts where policy experiments are perceived as legitimate and encouraged, failed experimentation may be interpreted as an opportunity to receive praise – for learnings acquired and saving public budgets from greater losses – rather than a source of accusation for wasting these budgets (**II**; Leino & Åkerman 2021). The media proved to have an important role in framing and further engendering this perception. In particular, the analysis indicates that if policymakers anticipate negative media reactions to experimentation (e.g., due to the dominance of austerity or previous policy failures), they are likely to refrain from launching it (**II**; in line with Albadri 2024; Strömberg 2015; Maurer 2023). Low legitimacy of policy experiments, thus, risks becoming self-reinforcing, as the initial lack of (successful) experimental experiences may ultimately prevent policymakers from undertaking further, potentially successful, experimental efforts.

On the other hand, experimental programs without effective legal protection prove to be *critically* vulnerable to the short-termism of political interests (**I**, **III**, **IV**). The lack of formal guarantees that experiments will not be discarded for political reasons may discourage policymakers with limited decision-making autonomy from launching the programs (**IV**; Eneqvist 2022; Koon et al. 2020; Torrens et al. 2018). The issue is especially relevant for bottom-up experimental programs as well as those whose timeframe overlaps with the junction of two political cycles (**IV**). In particular, the 'job-guarantee' study indicates that experimental legitimation may be equally important for the prospects of both small-scale policy pilots initiated by civil servants (local-level public institutions) and ambitious grassroots interventions of much larger scale (**IV**).

Along with that, the legitimation cost *per se* proves to shape the interplay of political motivations – affecting whether they work in favour or against experimental programs. The thesis illustrates that the potential losses in political time associated with (de)legitimation processes involved in experimentation may lead to two contrasting

scenarios. Under the first, the investments of time required to create legal groundwork for policy experimentation may make policymakers reluctant to support it, thereby preventing the launch of promising pilot programs (**IV**). The experience of TZCLD exemplifies this scenario: facing weak strategic and ideological incentives, higher-level policymakers (the French Parliament) lacked motivation to invest their limited time-budgets in preparing a legal framework enabling grassroots actors, such as NGOs, to experiment at the territorial level (**IV**; consistent with **III**, Perkiö 2020). As a result, the first attempt to launch the experimentation (1994) was halted; the required legal permission was enshrined in the French Constitution later (as part of Act II of decentralization in 2003; Torre & Bourdin 2023). Under the second scenario, the political costs of reversing the formally adopted (and scaled-up) experimental programs – i.e., those legitimised as regular policy tools – may help sustain them against alternating government offices and changing political agendas (**I**, **IV**). As the case of Kinofelis demonstrates, the losses in political time associated with the program's 'delegitimation' (i.e., withdrawal of its legal and operational status) enabled it to survive under successive center-right Ministerial offices despite their explicit non-support (**I**, **IV**; echoing the point made in 3.3). As of 2026, the program continues its enrolments at the national level (though limited in scope and support services provided to the unemployed).

3 Concluding Discussion

Policy experimentation is increasingly promoted as an effective approach to 'making public policies' in the face of rapidly evolving societal challenges (Hughes et al. 2020; Raudla et al. 2023, 2025; Markowitsch & Scharle 2024; OECD 2020). Yet, its political dimension lacks sufficient and systematic attention in existing scholarship – which distorts expectations regarding the opportunities and pitfalls experimentation holds in real-world contexts. This thesis makes its contribution to filling in the gap. Drawing on two comparative studies and four associated publications, it examines how different types of political motivations shape experimental policymaking in the public sector – addressing the following research questions:

1. *What are the key political motivations that influence public policy experimentation?*
2. *Which aspects of experimentation do they influence?*
3. *How do they influence those aspects?*

This section discusses the findings of the thesis accordingly, concluding with their graphic summary, Figure 1.

What are the key political motivations that influence public policy experimentation?

The doctoral analysis (Section 3) emphasises the three groups of political motivations as crucial decision-making drivers that shape experimental policymaking: strategic, ideological, and temporal.

The first group – *strategic political motivations* – refers to policymakers' aspirations to maximise their *private* political advantage while experimenting with *public* policy solutions. The findings distinguish between *reputation*- and *power*-maximisation interests as two key strategic motivations guiding the choices of *for* or *against* experimentation. On the one hand, policymakers are shown to be interested in maintaining and enhancing their electoral appeal by receiving public praise for actions taken or by minimising risks of accusations for potential policy failures (**II**, **I**, **IV**). In the thesis, the effects of reputational interests on experimentation are analysed through the lens of *blame avoidance-credit claiming calculus,* developed in **II**. Accordingly, the analysis illustrated that policymakers seek to maximise the calculus while making decisions regarding experimentation – prioritising potential losses over potential gains due to risk-aversion and negativity bias that typically pertain to them. On the other hand, policymakers prove to be interested in maintaining and enhancing their status quo by preserving, consolidating, or expanding access to political resources such as available authority and operational budgets (**I**, **IV**).

The second group – *ideological political motivations* – spans party preferences, ideational commitments, and normative preconceptions of policymakers – all of which proved to shape their choices regarding policy experimentation. Furthermore, the findings indicate that in certain policy domains, policymakers' attitudes towards potential experimental solutions or the challenges these solutions promise to address tend to dominate decision-making grounds. In particular, the issues of divergence and dynamics of ideological motivations proved to be largely explanatory for the differences in the (up)scaling trajectories of the studied 'job guarantees' (**I**, **IV**).

The third group – *temporal political motivations* – denotes policymakers' orientation toward political time, i.e. their aspiration to align decisions with the dynamics of the

timeframes they operate in. These dynamics encompass both the political timeframes as such – electoral cycles, governments' terms in office, and leadership turnovers within the experimenting institutions – and the anticipated disruptions between them. The analysis has illustrated that in democratic settings, the tensions between the political and experimental timeframes are barely avoidable and prove to shape policymakers' reasoning regarding policy experimentation on a par with the other two political motivations. Specifically, such aspects of political short-termism as limited time-budgets, preference for electoral expediency over long-term outcomes, and (anticipated) political and policy changes are shown to influence the variety of choices involved in experimental policymaking in the public sector (**I**, **III**, **IV**).

Which aspects of policy experimentation do the political motivations influence?

The strategic, ideological, and temporal motivations proved to shape policy experimentation across its different stages, interfering with and challenging the rationale. Namely, the doctoral analysis points out the four 'affected areas' where the political effects are salient and significant. These include: *the chances of policy experiments being launched* (1), *their research design* (2), *evaluation* (3), *and adoption prospects and (up)scaling dynamics* (4).

First, the political motivations, and primarily the strategic ones, are found to influence public policy experiments *ex ante* – affecting whether a *specific intervention* will be initiated as well as whether *experimentation in general* is adopted across public institutions as an approach to 'making novel policies'.

Second, the duration, scope, and dynamics of policy experiments prove to be exposed to the effects of strategic and temporal motivations. The analysis indicates that these experimentation specifications are subject to political choice, rather than being driven solely by technocratic intentions to optimise policy learning from the intervention.

Third, the role, focus, and outcomes of experimental evaluation are shown to be largely influenced by ideological and strategic motivations. In the doctoral analysis, both groups proved to affect which aspects the evaluation spans and (consequently) whether experimental interventions and experimental solutions per se are recognised, interpreted, and promoted as *successful*.

Fourth, the political motivations prove to have pivotal influence on the adoption trajectories of experimental solutions – affecting their scaling during experimentation (i.e., piloting extension in space and time) and beyond (referring to both permanent adoption of the solutions and their 'preservation' at scale against the backdrop of (alternating) political cycles).

How do the political motivations influence these aspects?

Chances of launch

The findings demonstrate that when experimental interventions are not ideologically driven, policymakers tend to assess them via the lens of strategic palatability, weighing expected political gains over the risks and losses experimentation entails. Ceteris paribus, the lack of strategic incentives discourages policymakers from undertaking experiments. The lack of strategic incentives may arise from multiple grounds, including the experiments' negative blame-credit calculus or the risks they impose on policymakers' status quo. In the first case, the framing of experimentation – and particularly, the framing of experimentation failure – proved crucial for adopting it as an approach to policy innovation across public institutions. As the analysis illustrates, experimental

policymaking tends to be a much more pervasive practice when the failure of the trialled solution is viewed as a credit-claiming opportunity rather than a reason for electoral accusations. The country cases studied illustrate this contrast: in Finland, the failure of policy experiments was associated with saving public budgets from greater losses (positive calculus), whereas in Estonia it was framed as 'wasting taxpayers' money' (negative calculus) (**II**). In the second case and relatedly, policymakers are found unwilling to support those experimental interventions that jeopardise their current power or electoral appeal. Specifically, the findings elucidate that policy experiments face greater barriers to launch under either of the following two conditions. First, when their implementation entails a reallocation of available authority or funds from one policy actor to another (e.g., vertical resource reallocation from the central government to municipalities or horizontal resource reallocation from conventional to experimental operating institutions). Second, when the experiments risk eroding voter trust in policymakers (e.g., by disproving their prior policy choices, core opinions, or publicly made claims) (**I-IV**).

Furthermore, the analysis emphasises that policymakers' strategic incentives reflect the political dynamics they operate within – having a short-term nature in democratic systems. This short-termism, ceteris paribus, inclines policymakers to refrain from commissioning experiments that span beyond a single political cycle – such as their own term in office or the term of the incumbent government. Without such temporal alignment, strategic benefits are likely to diminish, since policymakers face uncertainty about the power positions they will hold by the end of the intervention. This, in turn, leads them to doubt both the reputational relevance of public reactions to such experimental programs and their own authority to influence those programs' fate (**III**, **II**).

Research design

The findings illustrate that the duration, scope, and dynamics of policy experiments are subject to political discretion rather than driven solely by technocratic intention to optimise learning from the interventions. In line with the point above, policymakers are found to prefer experimental interventions of *restrained* research design, i.e., those that have a shorter timeline and require a smaller budget to be implemented (II). On the one hand, such experiments are perceived as less risky and, consequently, more politically palatable to roll out, especially in contexts where experimentation failure is associated with ascribing public blame. On the other hand, policymakers often lack incentives to design experimental interventions beyond their term in office, as they may not be authorised to influence further adoption of the innovative policies and/or the electorate's favour earned for experimentation may lose its strategic relevance (**II**, **III**). Alternatively, when the research design of experimentation cannot be directly changed (e.g., in the case of already launched experimentation), the chase for electoral expediency may push policymakers into 'hastening practices'. For example, policymakers may hasten experimental evaluation to employ evidence on the program's success as a tool for enhancing their reputation before the elections (**II**). With that, the analysis indicates that ceteris paribus, in democratic systems, policymakers' temporal and strategic motivations work against experimentation with disruptive policy solutions – inclining them to favour incremental experiments as less risky and more controllable political means. This observation aligns well with the prevalence of experimental programs that help to improve conventional policymaking approaches rather than challenge them among existing case-studies.

Yet, the thesis further illustrates that the longer design can also be employed as a strategic tool. Driven by electoral expediency or ideological commitment, policymakers may intentionally extend the duration of experimentation beyond their term in office to avoid painful reforms in the first case, or in an attempt to preserve the preferred tool after political change in the second.

Experimental evaluation: role, focus, and outcomes

In terms of experimental evaluation, the findings demonstrate that the normative choices it entails are often shaped by political reasons. On the one hand, policymakers' strategic priorities and ideological preferences may take control over the role and focus of experimental assessments. As the findings reveal, the evaluations involved in policy experimentation in practice may serve different purposes – going far beyond testing novel solutions, as its rationale suggests (**I**). In particular, when policymakers favour an experimental solution and have sufficient resources to decide its fate autonomously, they may divert official evaluation from collecting evidence on policy effectiveness to serving political ends. The evaluations may be employed to justify the pre-chosen policy courses and established opinions (which, interestingly, does not necessarily imply that the novel solution will be promoted as effective and adopted beyond experimentation), receive electoral praise for undertaking a 'scientific' approach, or identify the drawbacks of the program's design to strengthen it before full-scale adoption. Relatedly, experimental evaluation may also have a symbolic role, being included in the interventions to meet electoral expectations or regulatory requirements. In all these contexts, labels like "policy pilot" may still apply to the 'experimental' interventions even though in essence they constitute prototyping processes or a phase-wise adoption of pre-selected policy tools (**I**, **IV**).

On the other hand, the findings indicate that policymakers' ideological preferences significantly influence the outcomes of experimental evaluations, i.e., whether the interventions and the solutions they 'test' will be recognised and publicly promoted as successful. In the context of experimental policymaking, the notion of "policy success" proves to be a highly contentious matter given its conditionality on the vantage point, criteria, and interpretations applied by those who assess it. Consequently, when the choices along the listed triad are limited only to the preferences of the 'judge' (i.e., the policymaker authorised to evaluate or communicate the outcomes of policy experiments) the issue of success provability risks being unresolvable. As the thesis elucidates, the lack of standardised guidelines on what and how to evaluate opens up 'rooms' for policymakers to interfere with the outcomes of experimentation – either by directly affecting the criteria of official evaluations, or while communicating experimentation results to the public (**I**, **IV**; Ruggeri et al. 2020; Checkland et al. 2021; Mattocks 2026). And although experimental evaluations can potentially be made politically neutral (e.g. by preregistering the methods and criteria they apply, as it typically works in the case of RCTs), it does not protect policy experiments from being re-evaluated by authorised policymakers on the basis of alternative evidence**.** Thus, irrespective of scientific rigour and evaluation approaches, policymakers may ultimately relabel the success of policy experimentation into a certain form of failure seeking to align its outcomes with their pre-existing attitudes towards the novel solutions. The objection developed may then be utilised to (re)shape public opinion on the intervention or serve as a formal argument against upscaling it. With that, the thesis points to the complexity of the relationship between "success" and "scaling" in

experimental policymaking – illustrating that the former is neither a *sufficient* nor *primary* condition for the latter (**I**, **IV**; Mattocks 2026).

Adoption and (up)scaling dynamics

The strategic, ideological, and temporal motivations are shown to shape whether and how experimental policies are extended under pilot status, and what will happen to them once experimentation ends: whether they will be legitimised among regular policy tools or terminated unscaled. Echoing the point above, scaling up experimental policy solutions proves to be a matter of political choice rather than a technical, evidence-driven replication of solutions that 'work'. As the thesis demonstrates, low strategic incentives that upscaling could entail for the gatekeepers (i.e., policymakers regulating access to resources required to adopt novel solutions beyond experimentation) as well as ideological incompatibility of the solutions with the party affiliations of the latter are likely to terminate experimentation unscaled irrespective of the outcomes of internal and/or independent evaluations (**I**, **IV**; Mattocks 2026; Checkland et al. 2021; Sondal et al. 2024; Vreugdenhil & Rault 2010; Capano & Lippi 2017). In this light, top-down policy experiments appear better positioned, as they benefit from the decision-making autonomy of their initiators and can therefore be scaled up based on their commitments to the policy rather than – or in addition to – experimental evidence. The 'shield' of ideological commitment, however, proves to be a short-term privilege, as top-down policy experiments remain vulnerable to political changes arising from electoral cycles and leadership turnovers occurring at the organisational level (i.e., when the initiators of experimental programs no longer lead them) (**I**, **IV**).

Relatedly, the thesis emphasises that the political palatability of experimental programs tends to be relative in time and changes along with political cycles. The short-term nature of political timeframes is shown to challenge the scaling-up of effective pilot policies (implying both their adoption and preservation among regular policy tools) – requiring their initiators to plan and develop *anticipatory* adaptations to a range of potential political changes. In particular, various types of elections – such as legislative, presidential, and European – as well as leadership turnovers in the experimenting institutions prove to 'change the game' for the scaling trajectories of the studied experimental programs. Depending on the starting conditions, the beginning of the new political cycle *substantially* affected, or was expected to affect, the chances of the programs being extended or preserved at scale (**I**, **IV**, **III**).

Finally, the thesis emphasises that policymakers' strategic, ideological, and temporal motivations interact with one another and the broader settings of policy experimentation – which largely defines their effects on the latter. In particular, the following groups of contextual factors prove to shape the interplay of political motivations: the institutional set-up of experimental programs, the policy domain they pertain to, and their formal and perceived legitimacy. These factors are found to shape whether the political motivations facilitate or hinder policy experiments, how important they are among policymakers' decision-making grounds, and which group dominates the others among these grounds. More specifically, the findings indicate that the modes of launch and the types of institutional networks underlying experiments shape the risks that the clashing political motivations of different policy actors will undermine the scaling-up of successful experimental programs. The combination of bottom-up launch and innovative institutions established to operate the experimental programs proved the most notable in this regard – potentially the most politically vulnerable in the short run and the most politically resilient over the longer term, in the face of political changes.

The policy domain of novel policy solutions, in turn, proves to influence both the willingness of policymakers to engage with experimentation and the relevance of ideological motivations among their decision-making grounds. The thesis demonstrates that experimenting within domains that feature high ideological or party polarisation, or fundamental ethical concerns tends to be the most politically unpalatable 'by default' and, consequently, requires a lever of political commitment to be present or developed in the first place. In particular, it was found that launching experiments in such policy areas as taxation, labour-market, and welfare policies is barely feasible without strong political support from authorised policymakers, popular social movements, or other opinion leaders.

Lastly, the legal frameworks enabling policy experimentation along with its electoral endorsement as an approach to 'making public policies' create a backdrop against which experiments are (if so) conducted. Specifically, the thesis illustrates that both formal and perceived dimensions of policy experimentation legitimacy shape the political calculus of its potential initiators. For higher-level policymakers, the calculus involves weighing opportunities for credit claiming against the disposable political time and its potential losses related to legitimising an experiment. For experimenters with limited or no decision-making autonomy, such as civil servants, innovation labs, and NGOs, formal legitimacy of experimentation provides essential protection against changes in the policy agenda.

All things considered, the doctoral thesis has illustrated that multiple aspects of experimentation with novel policy solutions – such as the chances of the experiments being launched, their scaling dynamics, research design, and evaluation specifications – are subject to political influence. Policymakers' strategic, ideological, and temporal motivations proved to be both salient and consequential decision-making drivers that shape how public policy experimentation operates in practice. In particular, the thesis has shown that the identified political motivations largely explain why the interventions of smaller scope and scale dominate current experimental practices; why the criteria of experimental evaluations are so chosen; and, importantly, why the links between success and scaling of experimental policies remain elusive. Finally, the thesis has elucidated that the effects of political motivations are never fixed but depend on a specific experimentation setting – including such factors as the institutional set-up of experimentation, its policy domain, and legitimacy. Figure 1 summarises these findings and invites the reader to reflect on the foregoing discussion.

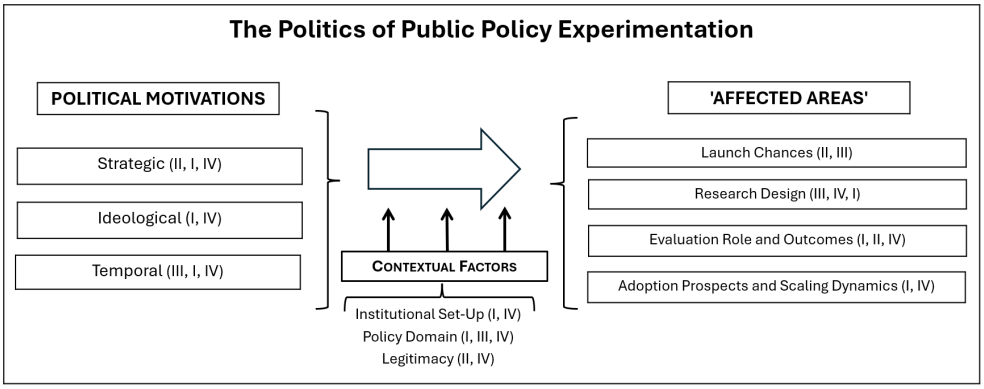


Figure 1. *Thesis Summary*
Source: Author

4 Propositions for Further Research

The conducted research has revealed a series of avenues for advancing knowledge of politics in experimental policymaking further. First, future studies could verify and deepen the doctoral findings in other comparative settings, policy cases, and domains. Figure 1, a graphic summary of the thesis, could serve as an analytical map for that: to guide the subsequent studies on the topic and help situate their contributions in a broader and more systematic, relational context. The figure depicts the relationships between the key elements of the framework developed – the three groups of political motivations; the 'areas', or aspects, of policy experimentation they affect; and the key contextual factors that mediate and moderate these effects – offering a system of potential research inquiries. In particular, future research could deepen understanding of the framework's separate compounds or the links between them. For instance, the studies could examine whether and how the effects of a specific group of political motivations vary across experimentation of different types (e.g., RCTs with pre-registered research designs vs. non-RCTS), explore how the three groups interact in different policy domains, or zoom in on the interplay of a specific group of political motivation and a particular contextual factor.

Second, further studies could expand the list of contextual factors that shape the role of politics in experimental policymaking. Specifically, more (comparative) research is needed to understand to what extent and how the size of a country, its political regime, level of public trust, and different authority-distribution settings, both vertical and horizontal, influence the politics of experimental policymaking (**I-IV**; Heilmann 2008; Zhu & Zhao 2021). Alternatively, the researchers could take a closer look at the interactions of political motivations *between* and *among* different policy actors, including: bottom-up initiators vs resource gatekeepers (i.e., higher-level decision-makers); experimentation stakeholders involved in its governance and implementation across different stages. Also, there remains very limited knowledge on the politics of international policy experiments. In particular, it would be fruitful to explore the potential clashes (and their resolutions) of political motivations between the engaged experiments' stakeholders representing different countries (and therefore, committed to different policy priorities and agendas).

Third, the thesis draws attention to the importance of conceptual framings in shaping research and its outcomes on the matter. In particular, one could explore how the narratives and analytical lenses utilised to interpret "policy experimentation" and the phenomenon of "politics" in existing scholarship inform its findings, resulting in the knowledge gap that this thesis engages with. As the thesis indicates, examining policy experimentation through the technocratic doctrine lens contributed to the general neglect of its political dimension; whereas framing "politics" as a set of separate factors compromised its systematic understanding in the relevant debates; finally, focusing on the challenges politics poses in (experimental) policymaking imbued it with a negative connotation, leading to the gap in knowledge concerning *opportunities* and *enabling mechanisms* it may offer for experimental policy interventions (Jones & Whitehead 2018; Huitema et al. 2018; Strassheim & Kettunen 2014; Neuwinger 2022). Therefore, future studies could contribute to developing more reflective conceptual tools to better capture the complexity and dynamics of the political reality of policy experimentation.

Importantly, the issue of conceptual framings concerns both the primary categories of the doctoral inquiry (such as "politics" and "policy experimentation") and the related

notions actively utilised in the analysis. In particular, the notion of *scaling(-up)* proves to shape the findings on the role of politics in evidence-based and experimental policymaking. In existing discussions, the term "scaling(-up)" is typically associated with either an *expansion* of experiments in *scope/scale,* or the *ultimate* outcome of experimentation (e.g., Reymond et al. 2020; Spicer et al. 2014; Vreugdenhil & Rault 2010). By contrast – and as a fourth specific proposition – the thesis suggests revisiting the term to facilitate a stronger analysis on the matter. Drawing on the thesis' findings and in line with Sondal et al. 2024, the thesis proposes defining "scaling-(up)" of experimental policy programs as *a process,* or *dynamics, of their change in both time and scope/scale*. Incorporating both temporal and scope dimensions in the analysis – and explicitly delineating them – would advance understanding of the programs' adoption trajectories and how politics interferes with them. This proposition rests on two empirical observations. On the one hand, the challenges that hamper scaling-up of policy experiments and their causes may differ along each dimension. This observation indicates the need to distinguish between the issues of expansion and prolongation of experimental programs. In particular, policy experiments that have proven effective may ultimately be discontinued for separate reasons. For instance, the limited budget of the initiating agency may hinder local adoption of the program (prolongation issue), while the inability to expand the experimentation to new sites may reflect its political unpalatability for the government or higher-level policymakers (**IV**). Moreover, preserving the dimensional dichotomy of (up)scaling would help recognise the difference between the concerns of experimentation 'scalability' (scope-dimension, focus on expansion challenges of a mostly institutional nature) and the risks posed to policy experiments by political change (time-dimension, focus on prolongation challenges of a mostly political nature). While each group may prevent further adoption of experimental programs, different explanatory mechanisms would underlie them. On the other hand, the 'scaling journeys' of experimental programs do not end with their adoption or termination. Taking a longer perspective – that spans beyond an electoral cycle – enables one to acknowledge, inter alia, the issues of downscaling due to changes in political support for the program, as well as stakeholders' attempts to resume terminated or discontinued experimental programs under a conducive political agenda. In this light, the thesis promotes "scaling(-up)" of policy experiments as a *process* dependent on the dynamics of political timeframes.

A fifth and final research proposition addresses the practical implications of this point. Namely, for policy practitioners who consider commencing experimental interventions, it is crucial to adopt a *long-term, anticipatory* approach to their governance – developing adaptations for political risks of two kinds: short-run (related to adoption of experimental programs at scale as permanent policy tools, in case of success) and long-run (linked to sustaining the programs against changing political timeframes). Future studies could thus contribute directly to the empirical plane of experimental policymaking – investigating what adaptations prove effective in mitigating both short- and long-run risks for policy experiments of different types and in different policy domains. In particular, one could expand research on the links between the institutional set-up and politics of experimental programs in order to better grasp the mechanisms behind the political influences and, consequently, discover which institutional adaptations could improve the prospects of successful programs (following up on **I, IV**; McFadgen & Huitema 2017, 2018).

References

Abramowicz, M. (2019). Tax experimentation. *Florida Law Review*, 71, 65–122.

Al-Ubaydli, O., List, J. A., LoRe, D., & Suskind, D. (2021). How can experiments play a greater role in public policy? Twelve proposals from an economic model of scaling. *Behavioural Public Policy*, 5(1), 2–49.

Albadri, L. (2024). The influence of the media on political decisions. *Academic International Journal of Social Sciences and Humanities*, 2(2), 42–50.

Ampe, K., Grin, J., & Loorbach, D. (2021). Power struggles in policy feedback processes: Incremental steps towards a circular economy within Dutch wastewater policy. *Policy Sciences*, 54, 579–607.

Antonopoulos, R. (2023). Has the time for a European job guarantee policy arrived? *Working Paper No. 1022*, Levy Economics Institute of Bard College, New York.

Avelino, F. (2017). Power in sustainability transitions: Analysing power and (dis)empowerment in transformative change towards sustainability. *Environmental Policy & Governance*, 27, 505–520.

Bauknecht, D., Broekel, T., Fichter, K., Jacob, K., Kivimaa, P., Rogge, K., & Wolff, S. (2020). Exploring the pathways: Regulatory experiments for sustainable development – An interdisciplinary approach. *Journal of Governance and Regulation*, 9(3), 49–71.

Baumgartner, F., & Jones, B. (1993). *Agendas and instability in American politics*. University of Chicago Press.

Béland, D., & Cox, R. (2016). Ideas as Coalition Magnets: Coalition Building, Policy Entrepreneurs, and Power Relations. *Journal of European Public Policy*, 23(3), 428–445.

Beisheim, M. et al. (2025). Politics Matters! Political Will as a Critical Condition for Implementing the Sustainable Development Goals. *Earth System Governance*, 24: 100244.

Berghöfer, A., Schroeder, J., Spahn, A., & Müller, R. (2020). The German Health Care Innovation Fund – An incentive for innovations to promote the integration of health care. *Journal of Health Organization and Management*, 34(8), 915–923.

Bernecker, A., Boyer, P., & Gathmann, C. (2021). The role of electoral incentives for policy innovation: Evidence from the US welfare reform. *American Economic Journal: Economic Policy*, 13(2), 26–57.

Besley, T. (2007). *Principled agents? The political economy of good government*. Oxford: Oxford University Press.

Binz, C., Harris-Lovett, S., Kiparsky, M., Sedlak, D. L., & Truffer, B. (2016). The thorny road to technology legitimation: Institutional work for potable water reuse in California. *Technological Forecasting and Social Change*, 103, 249–263.

Boswell, C. (2009) *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*. Cambridge: Cambridge University Press.

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

Britz, V., Ebrahimi, A., & Gersbach, H. (2022). Incentive pay for policymakers? *Journal of Public Economic Theory*, 24, 259–275.

Bueno, N. (2023). The timing of public policies: Political budget cycles and credit claiming. *American Journal of Political Science*, 67(4), 996–1011.

Cairney, P. (2012). *Understanding public policy: Theories and issues*. Palgrave Macmillan.

Cairney, P. (2022). The myth of “evidence-based policymaking” in a decentred state. *Public Policy and Administration*, 37(1), 46–66.

Callander, S. (2011). Searching for good policies. *American Political Science Review*, 105(4), 643–662.

Callander, S. (2008). Political motivations. *The Review of Economic Studies*, 75(3), 671–697.

Callander, S., & Hummel, P. (2014). Preemptive policy experimentation. *Econometrica*, 82(4), 1509–1528.

Capano, G., & Lippi, A. (2017). How policy instruments are chosen: Patterns of decision makers’ choices. *Policy Sciences*, 50(2), 269–293.

Carmines, E. G., Ensley, M. J., & Wagner, M. W. (2012). Who fits the left–right divide? Partisan polarization in the American electorate. *American Behavioral Scientist*, 56(12), 1631–1653.

Casula, M. (2022). How different multilevel and multi-actor arrangements impact policy implementation: Evidence from EU regional policy. *Territory, Politics, Governance*, 12(7), 1048–1072.

Checkland, K., Hammond, J., Coleman, A., MacInnes, J., Mikelyte, R., Croke, S., Billings, J. R., Bailey, S., & Allen, P. (2021). “Success” in policy piloting: Process, programs, and politics. *Public Administration*, 101(2), 463–480.

Cinar, E., Trott, P., & Simms, C. (2019). A systematic review of barriers to public sector innovation process. *Public Management Review*, 21(2), 264–290.

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.

Creswell, J. W., & Poth, C. N. (2023). Qualitative inquiry and research design: Choosing among five approaches (5th ed.). SAGE.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.

Duflo, E. (2020). Field Experiments and the Practice of Policy. *American Economic Review*, 110(6), 195–204.

Duflo, E., & Bertrand, M. (2016). Field Experiments on Discrimination. *Handbook of Economic Field Experiments*, 1, 309–393.

Dunlop, C. A., Radaelli, C. M., Wayenberg, E., & Zaki, B. L. (2024). Beyond powering and puzzling: The political dimensions of policy learning. *Journal of European Public Policy*, 31(7), 1979–1992.

Eneqvist, E. (2022). Experimental governance: Capacity and legitimacy in local governments. *PhD Thesis*, KTH Royal Institute of Technology.

Esaiasson, P., & de Fine Licht, J. (2024). Walking a mile in their shoes – How local politicians reason about trade-off decisions. *Local Government Studies*, 1–22.

Ettelt, S., Mays, N., & Allen, P. (2015a). 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.

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

Ettelt, S., Williams, L., & Mays, N. (2022). National policy piloting as steering at a distance: The perspective of local implementers. *Governance*, 35, 385–401.

Freeman, R., Griggs, S., Boaz, A. (2011). The practice of policy making. *Evidence & Policy: A Journal of Research, Debate and Practice*, 7(2), 127–136.

García-Hombrados, J., Jansen, M., Martínez, A., Özcan, B., Rey-Biel, P., & Roldán-Monés, A. (2024). Ideological alignment and evidence-based policy adoption. *IZA Discussion Paper No. 17007*, IZA Institute of Labor Economics.

Geels, F. W. (2014). Regime resistance against low-carbon transitions: Introducing politics and power into the multi-level perspective. *Theory, Culture & Society*, 31(5), 21–40.

Gersbach, H. (2004). Competition of politicians for incentive contracts and elections. *Public Choice*, 121(1–2), 157–177.

Giger, N., Nelson, M. (2011). The electoral consequences of welfare state retrenchment: Blame avoidance or credit claiming in the era of permanent austerity? *European Journal of Political Research*, 50, 1–23.

Goetz, K. H. (2014). Time and power in the European Commission. *International Review of Administrative Sciences*, 80(3), 577–596.

Grin, J. (2020). 'Doing' system innovations from within the heart of the regime. *Journal of Environmental Policy & Planning*, 22(5), 682–694.

Guerineau, M., Jacob, F., Kleszczowski, J. (2024). Codesign in action: Design principles to successfully manage transformative social innovation. *IEEE Transactions on Engineering Management*, 71, 12037–12052.

Hallsworth, M. (2011). Policy-making in the real world. *Political Insight*, 2(1), 10–12.

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

Hong, S., Lee, S. (2018). Adaptive governance, status quo bias, and political competition: Why the sharing economy is welcome in some cities but not in others. *Government Information Quarterly*, 35(2), 283–290.

Hood, C. (2010). *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton University Press.

Howlett, M., Goetz, K. H. (2014). Introduction: Time, temporality and timescapes in administration and policy. *International Review of Administrative Sciences*, 80(3), 477–492.

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.

ILO. (2018). Towards a stronger Kinofelis: Recommendations for the strengthening of future phases, *ILO Report*.

Jessoula, M., Natili, M. (2020). Explaining Italian "exceptionalism" and its end: Minimum income from neglect to hyperpoliticization. *Social Policy and Administration*, 54, 599–613.

Jones, R., Whitehead, M. (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.

Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist*, 39(4), 341–350.

Kangas, O. (2021). Chapter 14: The feasibility of universal basic income. In: *Experimenting with Unconditional Basic Income*. Cheltenham, UK: Edward Elgar Publishing.

Laczik, A., Dabbous, D., Patel, J., & Wilson, E. (2023). Good policy making: What can we learn from past policies? London: The Edge Foundation.

Leino, H., & Åkerman, M. (2021). The politics of making Finland an experimenting nation. *Critical Policy Studies*, 16(4), 441–459.

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.

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

Liarte, I., Criado, J. I., & Alcaide-Muñoz, L. (2024). Exploring barriers to innovation in public administration: An empirical study of the local layer of government. *International Journal of Public Administration in the Digital Age*, 11(1), 1–23.

Linz, J. J. (1998). Democracy's time constraints. *International Political Science Review*, 19(1), 19–37.

Maddison, S. (2012). Evidence and contestation in the indigenous policy domain: Voice, ideology and institutional inequality. *Australian Journal of Public Administration*, 71, 269–277.

Markowitsch, J., & Scharle, Á. (2024). Towards zero long-term unemployment in the EU: Job guarantees and other innovative approaches. Social Innovation Initiative, *Report for the European Commission*.

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

Mattocks, K. (2025). Can policy experiments achieve policy change? The politics of experimentation in Canadian cultural policy. *Policy Sci*, 58, 221–244.

Mattocks, K. (2026). What is successful policy experimentation?. *Policy & Politics*, 54(1), 91–112.

Maurer, P. (2023). Perceptions of media influence and performance among politicians in European democracies. *International Communication Gazette*, 85(5), 347–364.

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.

McFadgen, B., & Huitema, D. (2018). 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.

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

Natili, M. (2019). The politics of minimum income: Explaining path departure and policy reversal in the age of austerity. Palgrave Macmillan.

Nesti, G., Cittadini, M., & Bassoli, M. (2025). The politics of piloting: The case of minimum income schemes in European cities. *Review of Policy Research*, 42, 291–309.

Neuwinger, M. (2022). The revolution will not be randomized: Universal basic income, randomized controlled trials, and 'evidence-based' social policy. *Global Social Policy*, 22(1), 27–45.

OECD. (2020). Building capacity for evidence-informed policymaking: Lessons from country experiences. OECD *Public Governance Reviews*, Paris.

Parth, A. M., & Nyby, J. (2022). Welfare experiments as tools for evidence-based policy making? The political debate on Twitter about the basic income trial in Finland. *Policy Studies*, 43(4), 772–790.

Patton, M. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). SAGE Publications.

Perkiö, J. (2020). Legitimising a radical policy idea: Framing basic income as a boost to labour market activity. *Policy & Politics*, 48(2), 277–293.

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.

Ranchordas, S. (2013). The whys and woes of experimental legislation. *The Theory and Practice of Legislation*, 1(3), 415–440.

Raudla, R., Sarapuu, K., Douglas, J. W., Vallistu, J., Harbuzova, N. (2025). Budgeting for innovation: How funding strategies enable or constrain public sector experimentation. *Public Budgeting & Finance*, 45(4), 28–40.

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

Raudla, R., Karo, E., Valdmaa, K., Kattel, R. (2015). Implications of project-based funding of research on budgeting and financial management in public universities. *Higher Education*, 70(6), 957–971.

Raven, R., Van den Bosch, S., Weterings, A. (2019). Urban experimentation and institutional arrangements. *European Planning Studies*, 27(2), 258–281.

Reymond, Ph., Chandragiri, R., & Ulrich, L. (2020). Governance arrangements for the scaling up of small-scale wastewater treatment and reuse systems – Lessons from India. *Frontiers in Environmental Science*, 8, 72.

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-Dillon, R. (2005). *The welfare experiments: Politics and policy evaluation*. Stanford: Stanford University Press.

Roosma, F. (2022). A struggle for framing and interpretation: The impact of the 'basic income experiments' on social policy reform in the Netherlands. *European Journal of Social Security*, 24(3), 192–212.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320.

Ruggeri, K., Yarkoni, T., Lucas, J., Lewis-Peacock, J. A., Griffiths, T. L., Brooks, S. (2020). Standards for evidence in policy decision-making. *Nature Research: Social and Behavioural Sciences*, 39905.

Sager, F., & Gofen, A. (2022). The polity of implementation: Organizational and institutional arrangements in policy implementation. *Governance*, 35(2), 347–364.

Sarapuu, K., & Trei, M. (2025). Organizational dilemmas of EU funded policy development: The case of Estonian government task forces. *Public Money & Management*, 45(5), 400–403.

Scharpf, F. (1999). *Governing in Europe: Effective and democratic?* Oxford: Oxford University Press.

Schmidt, V. (2013). Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'. *Political Studies*, 61(1), 2–22.

Schönhage, N. L., Bækgard, M., & Geys, B. (2024). The politics of distributing blame and credit: Evidence from a survey experiment with Norwegian local politicians. *European Journal of Political Research*, 63, 599–620.

Sheffer, L. (2021). Political accountability, legislator gender, and the status quo bias. *Politics & Gender*, 17(3), 365–401.

Smith, Natalie R., Mazzucca, Stephanie, Hall, Marissa G., Hassmiller Lich, Kristen, Brownson, Ross C., Frerichs, Leah (2022). Opportunities to improve policy dissemination by tailoring communication materials to the research priorities of legislators. *Implementation Science Communications*, 3(1), 24.

Sondal, J., Hellquist, A., & Balfors, B. (2024). From pilot to practice: Navigating pre-requisites for up-scaling sustainable urban solutions. *Urban Transformations*, 6, 7.

Spicer, N., Bhattacharya, D., Dimka, R., Fanta, F., Mangham-Jefferies, L., Schellenberg, J., Tamire-Woldemariam, A., Walt, G., Wickremasinghe, D. (2014). 'Scaling-up is a craft not a science': Catalysing scale-up of health innovations in Ethiopia, India and Nigeria. *Social Science & Medicine*, 121, 30–38.

Strassheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy*, 10(2), 259–277.

Strøm, K., & Müller, W. C. (1999). Political parties and hard choices. In W. C. Müller & K. Strøm (Eds.), *Policy, office, or votes? How political parties in Western Europe make hard decisions*. Cambridge: Cambridge University Press.

Strömberg, D. (2015). Media and politics. *Annual Review of Economics*, 7, 173–205.

Tcherneva, P. (2019). The federal job guarantee: Prevention, not just a cure. *Challenge*, 62(4), 253–272.

Tcherneva, P. (2024). The job guarantee: Modern Money Theory's proposal for full employment and price stability. In Y. Nersisyan & L. R. Wray (Eds.), ch. 15, 196–212. Edward Elgar Publishing.

Tcherneva, P., & Lalucq, A. (2022). A job guarantee for Europe: Economic policy innovation for stability, security and sustainability. *FEPS Policy Brief*.

Torre, A., & Bourdin, S. (2023). The French Territorial Reform of the Regions: Objectives, Risks and Challenges for some Forgotten Territories. *International Journal of Public Administration*, 46(11), 761–772.

Torrens, J., Johnstone, P., & Schot, J. (2018). Unpacking the formation of favourable environments for urban experimentation: The case of the Bristol energy scene. *Sustainability*, 10(3), 879.

Turnheim, B., & Sovacool, B. (2020). Forever stuck in old ways? Pluralising incumbencies in sustainability transitions. *Environmental Innovation and Societal Transitions*, 35, 180–184.

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.

Van Parijs, P., & Vanderborght, Y. (2017). *Basic income: A radical proposal for a free society and a sane economy*. Cambridge, MA: Harvard University Press.

van Thiel, S. (2014). Research methods in public administration and public management: An introduction (1st ed.). Routledge.

Varazzani, C., Tuomaila, H., Emmerling, T., Brusoni, S., Fontanesi, L. (2023). Seven routes to experimentation in policymaking: A guide to applied behavioural science

methods. *OECD Working Papers on Public Governance*, No. 64, OECD Publishing, Paris.

Verkerk, J., & Van Buuren, A. (2017). Developing collaborative capacity in pilot projects: Lessons from three Dutch flood risk management experiments. *Journal of Cleaner Production*, 169.

Vreugdenhil, H., & Rault, P. (2010). Pilot projects for evidence-based policy making: Three pilot projects in the Rhine Basin. *German Policy Studies*, 6(2), 1–42.

Wenzelburger, G. (2014). Blame avoidance, electoral punishment and the perceptions of risk. *Journal of European Social Policy*, 24(1), 80–91.

Wispelaere, J., Halmetoja, A., & Pulkka, V.-V. (2023). A primer on the Finnish basic income experiment: From design and implementation to evaluation and impact. In M. Torry (Ed.), *The Palgrave International Handbook of Basic Income*, 413–435. Springer.

Wispelaere, J., Halmetoja, A., & Pulkka, V.-V. (2018). The rise (and fall) of the basic income experiment in Finland. *CESifo Forum*, 19, 15–19.

Wolak, J. (2020). *Compromise in an age of party polarization*. Oxford: Oxford University Press.

Xi, T., Yao, Y., & Zhang, M. (2018). Capability and opportunism: Evidence from city officials in China. *Journal of Comparative Economics*, 46(4), 1046–1061.

Yin, R. K. (2018). Case study research and applications.

Zhu, X., & Zhao, H. (2021). Experimentalist governance with interactive central-local relations: Making new pension policies in China. *Policy Studies Journal*, 49(1), 13–36.

Acknowledgements

A PhD is one of those life destinations where the journey itself is an ultimate purpose. My doctoral journey was intense and much elucidating, filled with excitement, experience, and many lessons to learn from – rewarding me with far more than a degree – to a degree much higher than a doctorate.

I am grateful to all fellow humans who made this journey what it was, and grateful to myself – for passing it with decency and dedication.

Particularly, I would like to express my respect and gratitude to both of my supervisors, Professor Ringa Raudla and Professor Külli Sarapuu. First – for entrusting me with the right to develop an independent pillar-study for my doctorate – the comparative research on 'job guarantees' – considering all the politico-economic controversy and complexity of the case. Second – for investing their time and careful attention in this thesis – willing to help me produce the best scholarly output from our joint research under PRG1125 "Experimental approaches and institutional innovations in the domains of fiscal and financial policy". Third and importantly, Ringa and Külli have also supported my endeavours in advancing knowledge about and beyond the matters of this thesis, disseminating doctoral findings, and engaging with the relevant networks and communities that promote research on the political economy of (experimental) economic policymaking – including multiple summer schools, workshops, talks and conference presentations. All these considerations are greatly appreciated, as each has provided its essential input to my professional development – equipping me well for conducting international comparative research of substantial value.

I also would like to express my appreciation to the Ragnar Nurkse Department of Innovation and Governance, TalTech, which has hosted my doctoral studies. The Department endorses interdisciplinarity and methodological pluralism in research and teaching, which is a crucial aspect to emphasise. On the one hand, this has simply made my doctoral research possible – empowering me to fulfil the ambition of exploring the 'backstage' of economic policymaking and innovation governance, paying special attention to the interplay of institutional set-up and politics of these processes. On the other hand, it has allowed me to learn from the diversity of mindsets, frameworks, and perspectives of other academic fellows – including opportunities to review the works of doctoral and postdoctoral researchers at our doctoral seminars, supervise and act as an opponent for MA theses of different foci, and complete educational modules of interest. All of this has enriched my academic growth further, sharpening the capacity to engage critically and productively across disciplinary boundaries of both economic and policy sciences.

This doctoral thesis was supported by the Estonian Research Council grant PRG1125 "Experimental approaches and institutional innovations in the domains of fiscal and financial policy".

Abstract

The Politics of Experimental Policymaking: How Strategic, Ideological, and Temporal Motivations Shape Policy Experimentation

Novelty and Problem Definition

Policy experimentation – referring to well-structured evidence-producing interventions of limited time and scale intended to test, evaluate, and further iterate novel policy solutions – is gaining traction across the globe: more and more public institutions adopt it as a means for devising effective responses to modern complex and rapidly evolving societal challenges. The political dimension of such interventions, however, remains overlooked, understudied, and lacks systematic attention in existing academic and practitioner-oriented discussions. Consequently, a remarkable divergence is observed between what policy experiments are expected to be and deliver and what policymaking practice shows they *are* in real-world contexts. In particular, a growing body of evidence suggests that policy experiments often unfold not in line with their rationale, that the interventions evaluated as 'successful' remain unscaled and/or end up terminated, and that the lessons learnt from experimentation have little, if any, bearing on policymakers' decisions – proving policy experiments unable to facilitate meaningful policy change.

This doctoral thesis constitutes a scholarly effort to bridge the gap between the rationale and practice of public policy experimentation by developing a more systematic, internally coherent, and empirically well-informed understanding of the politics involved in and influencing the latter. In order to delineate and explore the variety of political effects that public policy experiments are exposed to, the thesis pioneers the lens of political motivations – a novel analytical framework that brings together policymakers' intrinsic aspirations to maximise their private political advantage (such as available authority, disposable budgets, or electoral appeal) from the decisions made, and to align these decisions with specific party interests, ideational commitments, or operational timeframes. Accordingly, three research questions are addressed: *what are the key political motivations that influence public policy experimentation?*; *which aspects of experimentation do they influence?; and how do they influence those aspects?*

Methodology

The thesis draws on two comparative qualitative studies conducted under Grant PRG1125, Estonian Research Council. Each study employs a small-N case-study design ($N_1 = 3$, $N_2 = 2$) employing in-depth semi-structured interviews and desk research as key complementary methods for collecting data. The first study – the comparative research on European 'job guarantees' – compares three ideationally similar yet institutionally, operationally, and contextually different employment-creation schemes piloted in France (Zero Long-Term Unemployment Territories), Austria (MAGMA Job Guarantee), and Greece (Kinofelis Ergasia). This study pays special attention to the interplay of politics and the institutional set-up of the schemes, intending to reveal how it accounts for the differences in their scaling dynamics and adoption prospects. The ultimate inquiry of the study was to discover how to 'design better' – i.e., to pilot the 'job guarantees' that are (operationally) scalable, (politically) scaled-up, and sustainable in the long run – against various political and policy changes occurring at different levels (internationally, nationally, and at the level of experimenting agencies). Empirically, the study builds on semi-structured interviews (18) and follow-up discussions with the programs'

stakeholders (initiators, managers, evaluators, direct implementers). The second study – the Estonian-Finnish comparative case – takes a close look into policymakers' general experiences with experimentation in two neighbouring countries. The study examines *why* the experimental trajectories in the two countries with similar governance institutions vary and to what extent and how political motivations explain this variation. Empirically, the study builds on semi-structured interviews with policy practitioners from key public institutions and innovation labs in Estonia (32) and Finland (34).

The findings of the two studies are presented through four doctoral publications: three peer-reviewed articles and a single-authored book chapter.

Main Findings

The doctoral analysis identifies and puts forth three groups of political motivations as crucial forces that shape public policy experimentation: strategic, ideological, and temporal. The groups constitute interplaying yet distinct analytical categories that altogether operationalise the political dimension of public policy experimentation. Strategic motivations originate in policymakers' aspirations to maximise private political advantage – encompassing both reputation-maximising aspirations (studied through the lens of blame avoidance and credit claiming) and power-maximising aspirations (preserving or expanding available authority and operational budgets). Ideological motivations capture policymakers' inclinations to act in line with party affiliations, ideational preferences, and normative convictions. Temporal motivations refer to policymakers' orientation toward political time – their aspiration to align policy decisions with the dynamics of the timeframes they operate within, encompassing political cycles per se (electoral cycles, office terms, leadership turnovers within experimenting agencies) and anticipated policy and political changes at their junctures. Altogether, the strategic, ideological, and temporal motivations prove to affect: whether experiments are conducted (chances for launch), how they are conducted (research design and evaluation purpose), and what outcomes the latter deliver (recognition of policy success or failure, adoption prospects, and scaling dynamics).

Importantly, the effects of political motivations are shown to largely depend on the *broader contextual setting* of policy experimentation, which is shown to mediate and moderate them. Specifically, the institutional set-up of experimental interventions (modes of launch, types of institutional networks), their policy domain, and the perceived and formal 'legitimacies' of policy experimentation – corresponding, respectively, to the input and throughput dimensions of democratic legitimacy – prove to explain why and how the motivations become activated, how significant they are among decision-making grounds, whether they facilitate or hinder experimentation, and to what extent policymakers are able to act on them.

Contribution

The thesis demonstrates that various policy actors engaged with experimentation at its different stages – including the initiators of such programs, the stakeholders contributing to their launch, design, operation, and evaluation, and, importantly, the 'gatekeepers' controlling access to resources for their scaling and, therefore, deciding on their prospects (in the case of bottom-up interventions) – are not politically neutral but strategically, ideologically, and temporally motivated. These findings render expectations regarding the essence and deliverables of experimental policymaking more realistic – calling on practitioners conceiving experimental interventions in the public sector to adopt a *long-term anticipatory approach* to their governance. In particular, the importance of anticipatory adaptations against *two types* of political risks is revealed:

short-term risks – referring to adaptations navigating the *diversity* of political motivations involved in complex experimentation networks; and long-term risks – indicating the need to account for the *dynamics of change* in those motivations against the backdrop of electoral cycles, government office terms, and leadership turnovers in the experimenting agencies.

On the theoretical plane, the thesis offers a two-fold contribution to existing scholarship. First, it aids in collecting, systematising, and advancing existing knowledge on the politics of experimental policymaking. Second, it provides a novel analytical framework which could serve as a map to guide further studies on the matter – helping to both structure their research inquiries and to situate them within a broader, more systematic context.

Lühikokkuvõte

Katsetava poliitikakujundamise poliitilised mõjutegurid: strateegiliste, ideoloogiliste ja ajaliste motivatsioonide roll

Uudsus ja probleemi püstitus

Katsetav poliitikakujundamine (mis tähendab ajaliselt ja ulatuselt piiratud ning tugevalt struktureeritud tõenduspõhiseid sekkumisi, mille eesmärk on uudseid poliitikalahendusi katsetada, hinnata ja edasi arendada) kogub üle maailma populaarsust: üha enamad avaliku sektori institutsioonid võtavad seda kasutusele, et leida tänapäeva keerulistele ja kiiresti muutuvatele ühiskonnaprobleemidele tõhusaid lahendusi. Senises akadeemilistes ja valdkonnas tegutsejatele suunatud kirjanduses on aga selliste sekkumiste poliitilist mõõdet vähe uuritud ja sellest puudub süstemaatiline ülevaade. Seega on täheldatav märgatav erinevus ootuste (milleks poliitikaeksperimente peetakse ja mis tulemusi neilt oodatakse) ning tegeliku poliitikakujundamise praktika vahel (mis tulemusi poliitikaeksperimendid tegelikult toovad). Üha rohkem on tõendeid sellest, et poliitikaeksperimendid ei kulge sageli vastavalt oma algsele eesmärgile, „edukaks" hinnatud sekkumiste ulatust ei laiendata ja/või need lõpetatakse ning katsetamisest saadud teadmistel on poliitikakujundajate otsustele vähe või ei mingit mõju, mis näitab, et poliitikaeksperimendid ei suuda alati kaasa tuua tähendusrikkaid poliitilisi muutusi.

Selle doktoritöö eesmärk on vähendada lahknevust katsetava poliitikakujundamise teooria ja praktika vahel, luues süstemaatilisema, sisemiselt sidusama ja empiiriliselt põhjendatud arusaama selles osalevatest ja seda mõjutavatest poliitilistest mõjuteguritest. Selleks, et piiritleda ja uurida mitmesuguseid poliitilisi mõjutegureid, millest katsetav poliitikakujundamine mõjutatud on, rakendatakse töös uuenduslikku poliitiliste motivatsioonide analüüsiraamistikku. See ühendab poliitikakujundajate soovi maksimeerida tehtud otsustest tulenevat isiklikku poliitilist kasu (nagu saavutatav võimupositsioon, kasutatav eelarve ja valimistel edukas olemise potentsiaal) ning viia otsused kooskõlla konkreetsete erakondlike huvide, ideoloogiliste veendumuste ja tegevuste ajaraamidega. Töös on püstitatud kolm uurimisküsimust:

1. Millised on peamised poliitilised motivatsioonid, mis mõjutavad katsetavat poliitikakujundamist?
2. Milliseid katsetava poliitikakujundamise aspekte need mõjutavad?
3. Kuidas nad neid aspekte mõjutavad?

Metoodika

Doktoritöö tugineb kahele võrdlevale kvalitatiivsele uuringule, mis viidi läbi Eesti Teadusagentuuri uurimisgrandi PRG1125 raames. Mõlemas uuringus kasutatakse väikese juhtumiarvuga juhtumiuuringu disaini $N_1 = 3$, $N_2 = 2$), rakendades andmete kogumise põhiliste, üksteist täiendavate meetoditena poolstruktureeritud süvaintervjuusid ja dokumendianalüüsi. Esimene uuring – võrdlev teadustöö Euroopa „töötagatiste" kohta – võrdleb kolme ideoloogiliselt sarnast, kuid institutsionaalselt, tegevuslikult ja kontekstuaalselt erinevat tööhõive loomise kava, mida katsetati Prantsusmaal (*Zero Long-Term Unemployment Territories*), Austrias (*MAGMA Job Guarantee*) ja Kreekas (*Kinofelis Ergasia*). Uuring pöörab erilist tähelepanu poliitika ja kavade institutsionaalse ülesehituse vastastikusele mõjule, eesmärgiga selgitada välja, kuidas see põhjustab erinevusi nende laiendamise dünaamikas ja rakendamisväljavaadetes. Uuringu peamine eesmärk oli avastada, kuidas „paremini disainida" ehk kuidas katsetada töötagatisi, mis oleksid (tegevuslikult) skaleeritavad,

(poliitiliselt) laiendatavad ja pikaajaliselt jätkusuutlikud erinevate poliitiliste ja regulatiivsete muutuste kontekstis eri tasanditel (rahvusvaheliselt, riiklikult ja katsetavate asutuste tasandil). Empiiriliselt tugineb uuring poolstruktureeritud intervjuudele (18) ja jätkuaruteludele programmide erinevate osapooltega (algatajad, juhid, hindajad, otsesed elluviijad). Teine uuring – Eesti-Soome võrdlev juhtumiuuring – vaatleb lähemalt poliitikakujundajate üldiseid kogemusi katsetamisega kahes naaberriigis. Uuringus analüüsitakse, *miks* ja millisel määral erinevad kahe sarnaste valitsemisinstitutsioonidega riigi katsetamisteekonnad ning kuidas selgitavad seda erinevust poliitilised motivatsioonid. Empiiriliselt tugineb uuring poolstruktureeritud intervjuudele Eesti (32) ja Soome (34) peamiste avalike institutsioonide ja innovatsioonilaborite praktikutega.

Kahe uuringu tulemused esitatakse doktoritöö nelja publikatsiooni kaudu, milleks on kolm eelretsenseeritud artiklit ja üks ainuautoriga raamatupeatükk.

Peamised järeldused

Doktoritöö raames tehtud analüüs tuvastas ja määratles katsetavat poliitikakujundamist mõjutavate oluliste teguritena kolm poliitiliste motivatsioonide gruppi: strateegilised, ideoloogilised ja ajalised. Need grupid on omavahel seotud, kuid eraldiseisvad analüüsikategooriad, mis üheskoos moodustavad katsetava poliitikakujundamise poliitilise mõõtme. Strateegilised motivatsioonid tulenevad poliitikakujundajate soovist maksimeerida isiklikku poliitilist kasu, hõlmates nii mainekasu maksimeerimise soove (mida vaadeldakse süü vältimise ja tunnustuse endale võtmise kaudu) kui ka võimukasu maksimeerimise soove (võimupositsiooni säilitamine või laiendamine ja kasutata olevate eelarvete säilitamine või suurendamine). Ideoloogilised motivatsioonid tähendavad poliitikakujundajate kalduvust tegutseda kooskõlas oma erakondliku kuuluvuse, ideeliste eelistuste ja normatiivsete veendumustega. Ajalised motivatsioonid viitavad sellele, kuidas poliitikakujundajad võtavad arvesse poliitilist ajaskaalat, ehk nende soovile viia poliitilised otsused kooskõlla ajaraamidega, milles nad tegutsevad, mis hõlmab nii poliitilisi tsükleid endid (valimistsüklid, ametiajad, juhtide vahetumine katsetavates asutustes) kui ka eeldatavaid poliitika ja poliitilisi muutusi nende puutepunktides. Kokkuvõtvalt on näha, et strateegilised, ideoloogilised ja ajalised motivatsioonid mõjutavad järgnevat: kas eksperimente viiakse läbi (käivitamise tõenäosus), kuidas neid läbi viiakse (uurimisdisain ja hindamise eesmärk) ning milliseid tulemusi nad toovad (poliitikameetme edu või ebaõnnestumise tunnistamine, rakendamisvõimalused ja laiendamise dünaamika).

Tuleb arvesse võtta, et on näha, et poliitiliste motivatsioonide mõjud sõltuvad suuresti katsetava poliitikakujundamise *laiemast kontekstist*, mis mõjusid vahendab ja modereerib. Konkreetselt võib välja tuua, et eksperimentaalsete sekkumiste institutsionaalne ülesehitus (käivitamisviisid, institutsionaalsete võrgustike tüübid), nende valdkond ning katsetava poliitikakujundamise tajutav ja ametlik „legitiimsus" (mis vastavad demokraatliku legitiimsuse sisend- ja läbilaskemõõtmele) seletavad motivatsioonide aktiveerumise põhjuseid ja viise, nende olulisust otsuste langetamisel, katsetamise soodustamist või takistamist ning seda, mil määral poliitikakujundajad saavad nende ajel tegutseda.

Panus

Doktoritöö näitab, et erinevad poliitikakujundamise osapooled, kes osalevad katsetamise eri etappides – sealhulgas selliste programmide algatajad, nende käivitamisse, disaini, tegevusse ja hindamisse panustavad osapooled ning, mis on eriti oluline välja tuua, eksperimentide ulatuse laiendamiseks vajalikele ressurssidele

ligipääsu kontrollivad isikud, kes otsustavad seega nende tulevikuväljavaadete üle (alt-üles sekkumiste puhul) – ei ole poliitiliselt neutraalsed, vaid strateegiliselt, ideoloogiliselt ja ajaliselt motiveeritud. Need tulemused muudavad ootused katsetava poliitikakujundamise olemuse ja tulemuste suhtes realistlikumaks, kutsudes avalikus sektoris eksperimentaalseid sekkumisi kavandavaid otsustajaid üles lähenema neile *pikaajalise ettevaatava lähenemisega*. Eriti tuleb esile tuua ettevaatava lähenemise tähtsust kahte tüüpi poliitiliste riskide suhtes: lühiajalised riskid (kohandused, mis on vajalikud keerukates katsetamisvõrgustikes esinevate *mitmesuguste* poliitiliste motivatsioonide seas navigeerimiseks) ja pikaajalised riskid (vajadus arvestada motivatsioonide *muutuste dünaamikaga* valimistsüklite, valitsuse ametiaegade ja katsetavate asutuste juhtide vahetumise tõttu).

Teooria osas panustab see doktoritöö teaduskirjandusse kahel viisil. Esiteks aitab see koguda, süstematiseerida ja edasi arendada olemasolevaid teadmisi katsetava poliitikakujundamise poliitiliste mõjutegurite kohta. Teiseks loob see uue analüüsiraamistiku, mida saab kasutada edasise uurimistöö suundade leidmisel, kuna see aitab struktureerida uurimisküsimusi ja paigutada neid laiemasse, süstemaatilisemasse konteksti.

Appendix 1

Publication I

Harbuzova, N. (2026). Why successful policy pilots are (not) scaled up: The case of European employment-creation schemes. *Halduskultuur: The Estonian Journal of Administrative Culture and Digital Governance*, 25(1), 24–49. (1.1)

Why Successful Policy Pilots Are (Not) Scaled Up: The Case of European Employment-Creation Schemes

NASTASSIA HARBUZOVA, TALLINN UNIVERSITY OF TECHNOLOGY, RAGNAR NURKSE DEPARTMENT OF INNOVATION AND GOVERNANCE, ESTONIA

Abstract

While existing research demonstrates that successful policy pilots often remain unscaled, the reasons for this observation lack clarity and systematisation. This study addresses the gap. Drawing on the synthesis of various literatures, it first develops four propositions explaining why successful pilots may not scale up in practice, including: success may not be a scaling prerequisite (1), success may not be proven to decision-makers (2), success may lack further funds and incentives (3), and success may not be scalable (4). To "probe" and strengthen the propositions, the study then employs them as lenses to analyse a comparative case-study of three European employment-creation (pilot) schemes: Zero Long-Term Unemployment Territories (France), MAGMA Job Guarantee (Austria), and Kinofelis Ergasia (Greece). Empirical analysis draws on interviews with program stakeholders facilitated with follow-up discussions and desk research. The findings illustrate that: the success of policy pilots is neither a sufficient nor a primary condition for their upscaling (a), and whether the solutions "that work" find further adoption can be largely determined by the interplay of policymakers' ideological preconceptions and the institutional design of pilot interventions (b).

Keywords: experimental policymaking, policy pilots, policy success, scaling-up, employment-creation schemes

1. Introduction

The growing turbulence of the modern world puts policymakers in a quandary where regular policy solutions no longer work in the face of rapidly unfolding challenges, while their relevant alternatives are yet to be developed. In that context, adopting policy pilots as an approach to designing effective policy responses has gained momentum[1] (Hughes, Yordi and Besco 2020; OECD 2020; Varazzani et al. 2023). Policy pilots exemplify one of the core types of policy experimentation (along with design experiments, experimental legislation, and regulatory sandboxes) and can be defined as time- and scope-limited interventions intended to test

1 While the opportunities and pitfalls of piloting differ across policy domains due to a number of ethical, legal, political and technical reasons (e.g., Cinar et al. 2019; Liarte et al. 2024; Raudla et al. 2023), a growing use of experimental practices has been observed both in the areas where experimentation have been adopted as a common approach for designing innovations (e.g., healthcare, environmental, urban polices), as well as in those where it previously got only limited application – such as welfare- and employment- policies (Markowitsch and Scharle 2024; Nesti, Cittadini and Bassoli 2025; Varazzani et al. 2023).

novel policy solutions and use evidence obtained to decide their future (Bailey, Hodgson and Checkland 2019; Raudla et al. 2024). In this study, both randomised controlled trials (RCTs) and non-randomised interventions are considered policy pilots as long as they fit that definition.

According to the rational understanding of policy experimentation – which, as Huitema et al. (2018) argue, still dominates the literature – policy-piloting is a process with an attractively linear logic. It implies that if experimental evidence proves a pilot policy successful, the latter will be upscaled – i.e., extended in time, scope, or both (e.g., Stoker 2010). Yet, what the rationale expects pilots to be often runs counter to what the reality shows they are. As a growing body of empirical scholarship shows, many successful policy pilots remain unscaled due to the multiple considerations of political, institutional, administrative, relational, and technical nature (Corduneanu-Huci, Dorsch and Maarek 2021; Mauw, Smith and Torrens 2023; Nair and Howlett 2016). These considerations, however, are scattered across different studies and policy domains, preventing a holistic, empirically informed understanding of both the pragmatic grounds underlying the upscaling decisions and the role that the success of policy pilots plays among these grounds. This study addresses the gap by developing a systematic perspective on the relations between *success* and *scaling-up* in the practice of policy-piloting.

The aim of this study is two-fold: to identify and systematise the reasons that explain a limited record of upscaling effective policy innovations (1), and to strengthen this knowledge further with evidence from the relevant comparative case-study (2). With that, the research question of the study is: why are successful policy pilots scaled up and *are not* scaled up?

To address the question, the study proceeds in two steps. First, it conducts a systematic review of existing literatures and synthesizes four discrepancies between the rationale ("scaling-up in case/because of success") and the practice of policy-piloting, as follows: *success may not be a scaling prerequisite (1)*; *success may not be proven to decision-makers (2)*; *success may lack further funds and incentives (3)*; and *success may not be scalable (4)*. To develop these propositions, the author has examined the existing scholarship, asking the following sub-questions: (a) is the policy pilot success *always* a factor for its upscaling?; (b) if not – *when* is it a factor?; and (c) what are the *other factors*? How, when and why do the other factors affect the prospects of effective pilot solutions? The literature reviewed covers studies on experimental- and evidence-based policymaking, public innovation, and the politics of policymaking. The studies span multiple innovation areas and policy domains, including environmental, health, welfare, and fiscal policies, and administrative, urban and social innovation.

Second, the study examines how each of the four discrepancies manifests itself in a specific policy case – the case of *employment-creation schemes*. The discrepancies serve as analytical lenses to compare the adoption experiences of three (pilot) programs: *Zero Long-Term Unemployment Territories* in France (hereinafter: TZCLD, from *Territoires Zéro Chômeur de Longue Durée*), *MAGMA Job Guarantee* in Austria (hereinafter: MAGMA), and *Kinofelis Ergasia* in Greece (hereinafter: Kinofelis). The program's pilot(ed) is a conceptually similar tool that is commonly labelled as a *job guarantee*, an active labour-market policy aimed at eradicating (long-term) unemployment by creating *voluntary*, *individually tailored* and *locally beneficial* job opportunities that do not compete with the regular economy and provide at least a living wage (e.g., Tcherneva 2019). The comparative case-study analysis enables us to verify and

strengthen the discrepancies further, which is crucial given their "cumulative" formation. By comparing the theoretical propositions with empirical insights, the study explores the role of *context* in policy-piloting, identifying its potentially key elements for the upscaling trajectories of the employment-creation schemes analysed, both policy (domain)- and program-wise.[2] As sources of empirical data, the study draws on semi-structured interviews with 18 program stakeholders (2023-2024; primary data source), complemented with desk research and follow-up discussions (with program leadership and evaluators; 2024-2025) intended to verify and enhance intermediate findings.

This study is structured as follows. Section 2 outlines the existing knowledge on the "success-scaling" relations, specifies both terms and introduces the discrepancies developed. Section 3 provides an overview of the programs analysed, and Section 4 outlines the methodology. Empirical analysis is provided in Section 5, and its key findings and inferences are discussed in Section 6. Section 7 concludes the study, indicates its limitations and suggests avenues for further research.

2. Literature Overview: Success and Scaling-up in Policy-Piloting

2.1. Research gap and definitions

Although policy pilots are recognised as primary examples of evidence-based policymaking (e.g., Ko and Shin 2017), the complexity of the real world hardly leaves them a chance to evolve according to the linear scenario where an experimental solution proven successful becomes regular. While a growing body of empirical research supports this claim, the evidence lacks a systematic understanding. Two circumstances account for that.

First, the questions posed in existing studies tend to treat success and scaling-up as two separate outcomes of policy experimentation. Specifically, the literature that (in)directly suggests that not all successful experimental policies are able or intended to be scaled up can be conditionally divided into three overlapping yet self-contained thematic groups. The first group focuses on the issues of policy evaluation, indicating that so far, there is no commonly recognised notion of "success" in experimental policymaking (Checkland et al. 2021; Martin and Sanderson 1999; Stoker and Evans 2016). The second group investigates the challenges in policy experimentation at its different stages – inferring that multiple considerations may prevent policy experiments from evolving according to the rationale (Cairney and Oliver 2017; Reymond, Chandragiri and Ulrich 2020; Vreugdenhil and Rault 2010). The third group of studies explores how policy experiments influence policymaking and what this influence depends on, providing an overview of the factors that affect different outcomes of policy experimentation, including policy scale-up, policy learning, and knowledge diffusion (Baird et al. 2014; McFadgen and Huitema 2017, 2018; Vreugdenhil et al. 2009).

Second, the lack of consensus on the definitions of *success* and *scaling-up* in experimental policymaking prevents their systematisation. In the scholarly discussion, the success of

2 I.e., distinguishing between the specifics of the policy case/field (a) and the specifics of the (pilot) program (b); in the article: employment-creation schemes (a) and TZCLD, MAGMA, and Kinofelis (b).

experimental policies has various interpretations. As Checkland et al. (2021, 466) note, "a multifaceted set of frameworks for judging policy success emerges, defining success from multiple perspectives, across time and through varied lenses." While in earlier studies, the success of an experimental solution is associated specifically with its positive evaluation (Haynes, Goldacre and Torgerson 2012; Jowell 2003), more recent research tends to include additional dimensions in the definition. For instance, Ettelt, Mays and Allen (2015a) distinguish between the material and political outcomes of policy experimentation, emphasising the ability of the latter "to catalyse implementation" as an essential ground for judging its success. Whereas Checkland et al. (2021) go further and suggest policy success to be a complex concept that incorporates the dimensions of process, outcomes, and effects of experimentation, in both program and political terms. Alternatively, there are studies that associate the success of policy experiments with the very fact of their scale-up (Binz et al. 2016; Reymond, Chandragiri and Ulrich 2020). Similarly, the scope of the term "scale-up" varies across studies. Scaling up of experimental projects tends to be associated with an increase in their scale along one or several dimensions, such as: the number of enrolled participants, geographical area, institutional spread, time horizon, budget, range of activities and interaction with the related programs (Gillespie 2004; Reymond et al. 2018; Reymond, Chandragiri and Ulrich 2020; Sondal 2024). Furthermore, some studies consider scaling-up a form of policy diffusion along with knowledge dissemination (Vreugdenhil et al. 2009), or as a special case of policy "amplification", along with its stabilisation and speeding-up (Lam et al. 2020).

Reflecting on both circumstances, this study zooms in on the relations between the success and scaling-up of policy pilots, specifying the terms in line with the aim set: to identify why *the rationale* of these relations ("scaling-up in case/because of success") may not work in practice. Specifically, *the success of a policy pilot* is defined as an evidence-based assessment of high policy performance made according to the criteria set by the evaluators. The definition accounts for two points. First, it considers the success to be an outcome of *solely and specifically* the pilot evaluation. Second, it acknowledges success to be based on judgments that reflect the perspective of its evaluators on what and how *should be* assessed. *Policy pilot scale-up* is defined as both the process and outcome of policy extension in time and/or scope, including policy expansion and prolongation, regular adoption and maintenance at scale, yet excluding knowledge diffusion to avoid confusing the term with policy learning.

Drawing on these definitions, the rationale for the "success-scaling" relations looks as follows: policy pilots evaluated as successful are expected to be scaled up, i.e., the fact of their success is expected to be a necessary and sufficient condition for the latter. The next subsection contrasts the rationale with evidence from the scholarship to explore why it may not work in practice.

2.2. Discrepancies between rationale and practice

2.2.1. Success may not be a scaling prerequisite

Existing studies question the very conditionality of pilot policy success for its upscaling, i.e., the extent to which the outcomes of pilot adoption stem from the outcomes of its evaluation.

Literature demonstrates that in practice, policy experiments are launched with multiple motivations and for various purposes (e.g., Ettelt, Mays and Allen 2015b; Torrens and von Wirth 2021), thus challenging the rationale with two alternative scenarios.

Under the first scenario, *scaling up appears to be a predefined decision.* Specifically, whether the experimental solution will be scaled up can be decided before the experimental evidence is evaluated (Nair and Howlett 2016; Sondal, Hellquist and Balfors 2024; Vreugdenhil et al. 2009), without evaluating/collecting it (Bailey, Hodgson and Checkland 2019) or when the evidence has a formal or symbolic character and is used to support rather than question the chosen policy objectives (Ettelt, Mays and Allen 2015a,b; van der Heijden 2014). Three general motivations account for that.

First, policy-piloting may be a tool for gaining political advantage or earning credit from the public (Raudla et al. 2025b). Policymakers may use policy pilots to demonstrate the effectiveness of decisions already made (Zurbriggen and Lago 2019), signal to the electorate their commitment to the "rigour of evidence" (Leino and Åkerman 2021; Sanderson 2002) or provide a "shield of argument" if their policy choices face criticism (Corduneanu-Huci, Dorsch and Maarek 2021). Additionally, policymakers may try to scale up the advocated solution before its evaluation is completed if they doubt their authority to do so in due time – e.g., if the experiment goes beyond an electoral cycle (Bueno 2023; Raudla et al. 2024). Second, a phase-wise adoption of the policy may be *labelled* as a "pilot project" when the intervention is initiated to catalyse the implementation of an already decided solution (Behaghel, Macours and Subervie 2019). Adopting a novel policy in steps may be a way to ensure it would work as intended by the initiators (Ettelt, Mays and Allen 2015a; Nair and Howlett 2016) or to stimulate a shift in public perception and achieve higher social acceptance for the tool (Ansell and Bartenberger 2016). Third, policy-piloting in practice may entail the process of policy prototyping when the emphasis is on evaluating *how* rather than *whether* the solution works (Behaghel, Macours and Subervie 2019; Nair and Howlett 2016). In that case, the pilot is launched to fine-tune the policy before rolling it out at scale (Checkland et al. 2021) or to inform its implementation design (Hughes, Yordi and Besco 2020).

Under the second scenario, *scaling up is not a primary goal of experimentation.* In certain contexts, policy pilots may act as "solutions in themselves," which means that their upscaling in case of success is deemed neither necessary nor attainable. For instance, policy-piloting may be used as a tool for conflict avoidance (Raudla et al. 2025b). Politicians may launch a pilot to delay politically unpalatable reforms beyond their term in office to dodge "electorate retribution" (Bailey et al. 2017). Similarly, piloting may serve as a way to replace unpalatable larger-scale policy reforms with a less painful option (Partners for Health Reform Plus 2004). Alternatively, policy pilots may act as "trailblazers" when the primary purpose of the experimentation is to learn about the tool. In that case, policy learning and policy change appear to be rather disconnected experimental outcomes: the tested solution is in an early phase of development, and its adoption in a given environment is not considered feasible (De Wispelaere, Halmetoja and Pulkka 2023; Ettelt, Mays and Allen 2015b).

2.2.2. Success may not be proven to decision-makers

The rationale of policy-piloting tends to underestimate the fact that policy success is a *value judgement* dependent on what is evaluated, how, and by whom (Checkland et al. 2021; Leong and Howlett 2022; Ruggeri 2025). As empirical studies demonstrate, different policy actors may not share a common view on whether the results of the pilot qualify it as successful (Ettelt, Mays and Allen 2015a; McConnell, Grealy and Lea 2020; Parth and Nyby 2022) – drawing attention to the issue of "provability" of success in experimental policymaking.

First, policymakers may question pilot policy success if they doubt the quality of its evaluation (Head 2010; McFadgen and Huitema 2017, 2018). While the "credible evidence" is argued to be crucial for catalysing scaling-up of experimental policies (Spicer et al. 2014), it is unclear *what kind of* evidence qualifies as such. Pilot policies can be evaluated in various ways, yet these *ways* are unlikely to be seen as equally valid (Abramowicz 2019; McFadgen and Huitema 2018; Patton 2003). The policy experimentation literature indicates that policymakers tend to prioritise one type of experimental evidence/design over another. The priority may be given to RCTs (Bravo-Biosca 2020; Webber and Prouse 2018) or non-randomised interventions (Ettelt, Mays and Allen 2015a; Ko and Shin 2017; Raudla et al. 2023) – depending on what is valued most in a given environment: scientific rigour or hands-on knowledge (Ruggeri 2025). Besides, the reputation of evaluators and their affiliation with the pilot may influence the perceptions of evidence credibility (Jacobson and Azzam 2018; Roseland, Lawrenz and Thao 2015). As the studies on evidence-based policymaking show, policy evaluations conducted by external experts are generally seen as more credible than the internal ones (Jacobson and Azzam 2018; Koon et al. 2020; Vetulani-Cęgiel and Cianciara 2025).

Second, policymakers may have different understandings of what constitutes a "successful" pilot policy and thus judge it on different grounds (Head 2010; Marston and Watts 2003; Ruggeri et al. 2020). The differences may stem from both normative choices and political reasons – referring to the issue of the "right" measurements of policy performance in the first case, and the embeddedness of policymakers' attitudes towards pilot solutions in their evaluations in the second (Koon et al. 2020; Lewis, McGann and Blomkamp, 2020; Ruggeri et al. 2020; Strassheim and Kettunen 2014). As the evidence-based and experimental policymaking literatures show, policymakers tend to prioritise their "hunches" over scientific conclusions when deciding the fate of policy measures (Bravo-Biosca 2020; Capano and Lippi 2017; Ettelt, Mays and Allen 2015a; Raudla et al. 2023). Accordingly, the success of pilot policies is likely to be judged against certain expectations that policymakers already have about them (Ettelt, Mays and Allen 2015a,b; Maddison 2012; Parth and Nyby 2022). This may seriously complicate proving success to decision-makers who are initially sceptical or not in favour of the pilot tool (Ettelt, Mays and Allen 2015a,b; McFadgen and Huitema 2018; Perkiö 2020). Diverse political interests and party programs compound the case further by giving rise to potentially conflicting discourses on what lessons the experimentation has *actually* delivered (Parth and Nyby 2022; Perkiö 2020).

Both considerations contribute to the fact that in practice, multiple evaluations of policy pilots may co-exist and contradict each other (Cairney 2022; Checkland et al. 2021; Pearce and Raman 2014). In particular, experimental evidence may clash with alternative sources and types of knowledge (Cairney 2022; Wesselink, Colebatch and Pearce 2014), which raises

the issue of evidence prioritisation: the outcomes of *which/whose* evaluation *should be* utilised for judging policy success? Literature on evidence-based policymaking identifies the relations of power as a key factor of knowledge hierarchy, arguing that evidence that matches the interests of the most "empowered" policy actor is likely to be decisive in political decision-making (Marston and Watts 2003; Mattocks 2021). This circumstance may produce the scenario where a pilot proven successful for one actor/agency is promoted as a failure and devalued by another (Leong and Howlett 2022; McConnell 2010; McConnell, Grealy and Lea 2020).

2.2.3. Success may lack further funds and incentives

Policy pilots can be initiated by various policy actors and at different governance levels, which may largely affect what funding strategies and sources will be applied to support experimentation and what level of financial autonomy the experimenters will enjoy (Cinar, Trott and Simms 2019; Raudla et al. 2025a). As empirical studies illustrate, the lack of financial autonomy among pilot initiators may become a central barrier to adopting successful policy solutions after experimentation (Bello et al. 2018; Raudla et al. 2025a; Sondal, Hellquist and Balfors 2024). The issue may arise within different funding regimes, due to the limited budgets of experimenting agencies, temporary nature of support when experiments are financed from the targeted (international) funds (e.g., European Social Fund Plus), or the resource dependency "by default" when the funding comes from the budgets of higher-level policymakers or private sponsorship (Ko and Shin 2017; Markowitsch and Scharle 2024; Raudla et al. 2025a). In all these cases, the need for a third-party "buy-in" to scale up the pilots makes their prospects dependent on external political will. Recent research supports this inference with multiple examples of successful policy innovations being unscaled as they failed to find and/or sustain such "political will" (Beisheim et al. 2025; Grin 2020; Nesti, Cittadini and Bassoli 2025; Sondal, Hellquist and Balfors 2024). Two circumstances explain much of the issue.

First, the unwillingness of policymakers to take risks and the availability of safer policy alternatives may leave successful pilot policies temporary. Literature on public policy and innovation describe (higher-level) policymakers as risk-averse actors who generally favour the status quo over policy change, as the latter may jeopardise their political advantage in terms of power or reputation among voters (Bello et al. 2018; Hong and Lee 2018; Liarte, Criado and Alcaide-Muñoz 2024). Associating "less risky" with "more politically palatable," policymakers tend to comply with less-than-optimal yet safer policy alternatives (Callander 2011; Hong and Lee 2018) and may thus be reluctant to replace regular policy measures with experimental solutions (Mauw, Smith and Torrens 2023; Turnheim and Sovacool 2020) or opt for adopting incremental innovations over more radical ones (Grin 2020).

Second, policy pilots may clash with the dynamics of political timeframes. As the respective studies indicate, policy pilots may lose their "appeal" during experimentation – due to the changes in government when the pilots overlap several electoral cycles, changes in voter demands, or various shocks that may divert policymakers from their original plans (Bueno 2023; Leino and Åkerman 2021; Mauw, Smith and Torrens 2023; Raudla et al. 2025b). Altogether, the conflicting timeframes of the piloting and its political environment may ultimately "undo" the adoption of successful policy innovations (Bueno 2023; Raudla et al. 2024).

2.2.4. Success may not be scalable

Finally, the rationale of policy-piloting disregards the fact that successful pilots may have limited *scalability* and thus fail to replicate the same results at scale. According to existing discussions, the following considerations may constrain the scalability of interventions: the "smallness" and research design of policy pilots, the diversity of contexts that the target scale encompasses, and the challenges involved in operating pilot solutions at scale.

To start with, the essential characteristics of policy pilot design, such as their limited scope and timeframe, may compromise the external validity of findings they deliver (Bédécarrats, Guérin and Roubaud 2019; Beets et al. 2020; Werner and Riedl 2019). On the one hand, policy pilots risk not capturing those effects that the tested solutions would generate at a larger scale or over a longer timeline (Bravo-Biosca 2020; Raudla et al. 2023). On the other hand, pilots may fail to capture multiple contextual variations that an experimenting country incorporates – including both the uneven distribution of the issues tackled (e.g., when the pilots target "narrow" social/demographic groups or the issues are intrinsic to certain geographical landscapes (Eboreime et al. 2021; Verkerk and Van Buuren 2017)), and the political and legal fragmentations. These circumstances may introduce a higher generalisability bias in larger piloting countries and RCT-designs that seek "optimal conditions" to ensure internal validity (Beets et al. 2020; Muralidharan and Niehaus 2017).

While the uneven distribution of the tackled issues directly challenges the external validity of pilot evidence, political and legal fragmentations constrain the scalability of the latter, affecting the demand for the pilot solutions or the opportunities for their adoption across the country. The dominance of different parties across local governments or diverging political agendas of local and central authorities may cause political palatability of the solutions to vary and thus block their adoption in areas of low political support (Cairney 2022; Jessoula and Natili 2020; Zhu and Zhao 2021). Similarly, the lack of legal consistency between the experimenting and potential receptive jurisdictions (including the lack or inability to provide legal support for policy experimentation within certain administrative units) may prevent the replication of successful pilot policies (Binz et al. 2016; Eneqvist 2022; Koon et al. 2020; Torrens, Johnstone and Schot 2018).

Finally, the capacities required to effectively operate pilot policies at scale may not be available (Bulthuis et al. 2020; Reymond et al. 2018; Reymond, Chandragiri and Ulrich 2020; Sondal, Hellquist and Balfors 2024; Vreugdenhil and Rault 2010). On the one hand, not all skills and infrastructures underlying the experimental process may be feasible to replicate in the target environment (Howlett and Leong 2022; Lewis, McGann and Blomkamp, 2020; Simmons et al. 2007; Verkerk and Van Buuren 2017). As studies indicate, the success of pilot programs may be inseparable from certain intrinsic and barely scalable elements of the experimental process (Eneqvist 2022; Howlett and Leong 2022). The organisational culture and informal routines of experimenting agencies, individual expertise, leadership charisma and certain technical facilities are some examples (Liarte, Criado and Alcaide-Muñoz 2024; Reymond, Chandragiri and Ulrich 2020; Torrens, Johnstone and Schot 2018). On the other hand, scaling-up may entail an unacceptable increase in costs or complexity of operating pilot policies (Reymond, Chandragiri and Ulrich 2020; Sondal, Hellquist and Balfors 2024). The growing network of policy/institutional actors involved with the policies may require specific

complementary capacities to be developed – to ensure effective coordination between them, establish new collaborations and manage resource provisioning (Cinar, Trott and Simms 2019; Eboreime et al. 2021; Vreugdenhil and Rault 2010). The inability to develop such capacities may prove successful pilots to be "unscalable" as well (Howlett and Leong 2022; Sondal 2024).

Altogether, the four discrepancies introduced describe the process of policy-piloting at its different stages: initiation (1), evaluation (2), and adoption of experimental solutions (3, 4) – providing a plausible systematic overview of how the "success-scaling" relations look in practice. Table 1 summarises the propositions.

Table 1: Why successful policy pilots may not scale up: summary of theoretical propositions

Discrepancies between rationale and practice	Reasons identified
Success may not be a prerequisite	Scaling-up may be predefined → piloting for: - voter validation (symbolic role of evidence) - phase-wise adoption - prototyping: to *optimise* rather than *test* novel policies Scaling-up may not be a purpose, or pilots as "solutions in themselves": - political tools for conflict avoidance -"trailblazers" (piloting for learning)
Success may not be proven to decision-makers	Pilot success is a value judgement → decision-makers may: - not consider pilot evidence as credible (*normative* issues of research design, evaluators, and measurement choices) - be negatively predisposed to pilot policies (*political* issues of biased evidence, evidence (re/de)valuation and "power-based" prioritisation) ↓ Clash of multiple evaluations, relativity of success/failure
Success may lack funds and incentives	Scaling-up requires (access to) resources → policymakers may lack: a) financial autonomy: insufficient disposable budgets; temporary funding; financing from higher-level policymakers/private sponsorship (↓) b) political incentives to buy-in due to: - status quo, risk-aversion and more palatable policy alternatives - loss of "appeal" due to electoral cycles, changes in voter demands, shocks
Success may not be scalable	Pilots may not yield the same results at scale due to: - limited external validity: "smallness" and research design of interventions - political and legal fragmentation: varying agendas, laws and regulations - constraints of operational design: inability to replicate pilot capacities or develop complementary ones (needed to steer upscaled solutions)

3. Case Selection

To examine how the theoretical propositions manifest themselves in a specific policy case, the author conducted a comparative case-study in the domain of employment(creation) policies. In the literature on experimental- and evidence-based policymaking, the insights from that domain remain very scarce, which is likely due to limited experiences of experimenting with such policy solutions, until recently (Markowitsch and Scharle 2024). This circumstance makes the analysis of employment-creation schemes highly beneficial theory-wise, as it enables the study both (a) to discover the *domain-specific* context of piloting the schemes and (b) to reflect on its importance for the discrepancies to occur. This is particularly relevant given the rise of political interest in the "job guarantees" observed in Europe and beyond over the past few years (Markowitsch and Scharle 2024; Tcherneva and Lalucq 2022).

The case study includes three programs that share the following characteristics. The programs were launched as *policy pilots* (according to official media/reports), are conceptually similar (labelled as "job guarantees"), relevant and comparable in time (ongoing/recent programs). All the programs have been promoted as successful in official media/research/reports (e.g., ILO 2018; Lehner and Kasy 2024; Tcherneva and Lalucq 2022; program websites),[3] yet their scaling records vary. With such characteristics, the programs strengthen the study further, enabling the examination of why scaling-up a similar pilot solution may follow different scenarios – and thus reflect on the role of the *program-specific* context for the prospects of policy pilots. The programs analysed are introduced next.

3.1. Zero Long-Term Unemployment Territories (Territoires Zéro Chômeur de Longue Durée), France

TZCLD is an ongoing pilot program extended for a second experimental term (2021-2026). It combats long-term unemployment at a territorial level by creating jobs that match the skills of the unemployed with the unmet needs of their community while facilitating social cohesion. The aim of the program is to make effective the right to employment for all French residents by demonstrating that no one is unemployable, and there is no shortage of work and public funds to do so.

TZCLD was initiated in 2016 by four French NGOs (grassroots) as a 5-year pilot program with the scope of 10 experimental territories (referring to local administrative units of up to 10,000 residents); facilitated by top-level politicians and funded (mainly) from the state (central government budget). In 2021, the program was extended for a second term and subsequently expanded to at least 50 new territories. By 2025, TZCLD will have been piloted at 75 territories employing 3,629 residents (as of 02.2025). The program has inspired piloting similar schemes across Europe, including the initiatives in Belgium and Italy (Markowitsch and Scharle 2024).

The operation of the scheme draws on a novel institutional network created as a part of the program. The Companies for Employment Purposes established based on existing social enterprises are the main implementing bodies, while the program governance is provided by the Local Committees for Employment, chaired by local authorities at the territorial level, and

3 Which does not necessarily entail that the program is successful according to definition adopted in the study (2.1). Whether it is the case is analysed in Section 5.

the TZCLD Association (management/coordination), Experimental Fund (finding allocation to the territories) and the TZCLD Observatory (research/evaluation coordination) at the general level. The scheme is iterative and evolves consensually, drawing on an open, diverse and dynamic network of stakeholders.

3.2. MAGMA Job Guarantee (Modellprojekt Arbeitsplatzgarantie Marienthal), Austria

MAGMA represents "the world's first universal job guarantee experiment" (Lehner and Kasy 2024), a preregistered RCT initiated and funded by the local public employment service (PES), and designed in collaboration with Oxford University researchers (its main pro bono evaluators). The program was aimed at eliminating long-term unemployment and improving the general well-being of the area by offering jobs that match the needs of the unemployed and their community. The purpose of the pilot was to provide scientific evidence on policy performance.

The program was conducted in the town-village Marienthal from 2020 to March 2024, creating 150 unconditional, voluntary and non-competitive jobs for long-term unemployed residents. Employment was created in several forms, including jobs in newly established social enterprises, subsidised jobs in the regular economy, and projects developed by the unemployed. The program was not scaled up and discontinued after the pilot term.

The initiating PES and its regular partner, *it.works*, a non-profit provider of labour market services, implemented MAGMA. To ensure credible evidence, no politicians were involved in the official evaluation and/or operation of the scheme.

3.3. Kinofelis Ergasia, or 'Of Common Benefit' (ΚΟΙΝΩΦΕΛΗΣ ΕΡΓΑΣΙΑΙ), Greece

Kinofelis is an ongoing public employment program led by the Greek Ministry of Labour (MOL) since 2011. The launch of the pilot was motivated by a sharp rise in unemployment observed in the country due to the global and domestic economic crises. Since then, the program has aimed to reintegrate unemployed residents into the labour market, reduce poverty and social exclusion by creating temporary jobs beneficial for both the unemployed and their municipalities. Limited disposable funds (European Social Fund Plus, MOL budget) and high demand from the population made the program selection-based (the chances for being enrolled varied by age, status, and unemployment term) (ILO 2018).

The scheme was co-designed with the Greek Trade Union Confederation in 2011 and the ILO (next pilot phases, 2016-2017). In 2016, its improved version was introduced and subsequently adopted as a regular policy measure nationwide. Overall, the scaling-up included three steps: from 17 (2011-2013) to 51 (2013-2015) and to all 326 municipalities (2016) with 25,000-45,000 participants annually. In 2019, the program was downscaled and has since been conducted in a reduced form.

The implementation of Kinofelis draws on the collaboration of the MOL (defines enrolment

terms, reimburses employment costs) with the municipalities (creates and steers employment projects).

4. Methods

To collect data, the study combines interviews with program stakeholders and complementary desk research. Altogether, 18 semi-structured interviews were conducted (5 for TZCLD, 7 for MAGMA, 6 for Kinofelis; 2023-2024). To select the interviewees, a criterion-based sampling (different roles in the pilot) was mixed with the snowball technique. The sample includes initiators, managers, evaluators and direct implementers of the programs (Appendix 1).

The interview questionnaire was iteratively adjusted based on the interviewees' experience and previously collected data. The questions asked dealt with piloting challenges, details of the evaluation process and results, scaling-up factors and expected/fulfilled scenarios for program adoption, and personal lessons learned by the interviewees during experimentation. The goal of each interview/discussion was to collect specific data, verify intermediate inferences or compare the perspectives of different stakeholders. The interviews of an hour on average, were conducted online, recorded and transcribed. To preserve anonymity, the interviewees are identified by codes in the analysis (e.g., "Fr1/Au1/Gr1"; Appendix 1).

Data analysis followed an iterative and interpretive approach consistent with a small-N case-study design (Creswell and Poth 2023). It was conducted in parallel with data collection so that one informed the other. Specifically, each of the four discrepancies (Subsection 2.2) was alternately "probed" on each of the three programs (Section 3) to analyse *to what extent* the discrepancy resonates with the program and *why,* i.e., whether the factors explaining its occurrence in the case are similar to explanations offered by the literature. In the next section, each case follows the logic described.

In addition, the analysis has benefited from feedback received during follow-up discussions with program stakeholders (program leadership and/or evaluators; 2024-2025). The discussions were held in connection with the author's presentations of intermediate (2024) and general (2025) results of the study on piloting "job guarantees" and enabled the verification and enhancement of the inferences made.

To help the reader follow the analysis (Section 5), Appendix 2 provides the charts that visualise program timelines and key events that might have affected them.

5. Analysis

5.1. TZCLD

The program proved to be the only case where the results of experimental evaluation directly affected its chances for scaling up. The primary motive for launching TZCLD was to demonstrate that the scheme works and ultimately achieve its legal adoption nationwide (Fr1-2, Fr5). While

the initiators themselves expected the scheme to be successful (Fr3-4), scaling up required proving that to the state (central government, MOL) (Fr1). As an interviewee elaborated, "because government is funding a significant part of the project, it is the government's decision: to move it forward, stop it, or reduce it" (Fr2).

The pilot featured two simultaneous evaluations. On the one hand, the scheme was evaluated "from within," by stakeholders. The initiators adopted a broad approach to evidence collection, gathering multiple types of learnings (scientific, hands-on, grassroots-feedback) on the program's effects, outcomes and costs (Fr1, Fr4). Although the evidence indicated the success of the program (Fr1-5) as well as encouraged its scalability, elucidating how the scheme should be iterated to operate across diverse French territories (Fr1), the central government did not consider it credible due to the risk of "dependent" evaluation (Fr1-3). As an interviewee commented, "we [stakeholders] might try to find what we are convinced of" (Fr3). On the other hand, and relatedly, the evaluation was conducted by the Scientific Committee assigned to the project by law (Fr1). This evaluation, in contrast, reflected the priorities of the state, focusing on collecting scientific evidence on the project's cost-benefit ratio (Fr2-3). At the time of the analysis, the outcomes of this evaluation were uncertain, yet the stakeholders were concerned that it might show the program to be a failure due to the high costs (Fr2-4).

Thus, the program faced the challenge of potentially conflicting evidence when the pilot success "according to stakeholders" might be devalued by "external" decision-makers. Furthermore, the interviewees questioned the neutrality of the Scientific Committee (Fr3-4). As one of them noted, the government had room for "tweaking" its conclusions by choosing what the Committee should and should not include in the program cost-benefit ratio: "sometimes we had the feeling that one of the tasks of the Scientific Committee was to kill us – to show how expensive the project could be – while we don't see how externalities are examined" (Fr4). The interviewees linked the lack of state support to the risks that the program posed to the status quo of the MOL. The adoption of the scheme at scale would imply an increase in the autonomy of local-level authorities at the expense of the MOL, including the cuts in disposable funds and reduced control over the territories (Fr4).

While the lack of political support from decision-makers threatened the program's upscaling beyond its legally protected pilot-term (Fr1), the political diversity of French territories complicated expanding the pilot. As interviewees clarified, expanding the experimentation to new territories required reconciling the conflicting interests of different political parties in power in different territories (Fr4, follow-up discussion). To make the pilot ideologically palatable for both "sides," the importance of diversifying political narratives to promote the scheme and adopting a consensual decision-making model for pilot governance was emphasised. Ultimately, such adaptations turned the challenge into a strategic advantage for the project. By demonstrating that "they [left and right] can make the project together" (Fr4), TZCLD neutralised its initially leftish affiliation and grew resilient to changes in the political agenda due to legislative, presidential and European elections (follow-up discussions).

Beyond political concerns, the issues of scalability made the interviewees question the nationwide rollout of the program. The fact that the governance (and in part, implementation) of TZCLD relied on the novel institutions created within the project caused its effective expansion to be particularly difficult. First, scaling up the scheme required defining how to

integrate the novel entities into the system of existing institutions and harmonise their mutually informed work at scale (Fr2). Second, different operationalisations of the scheme across French territories cast doubt on whether the program can be kept transparent nationwide. An interviewee explained, "the challenge will be to ensure that all territories still comply with the key fundamentals of the project: not compete with the normal economy, hiring people without selection. To monitor this at a larger scale will be much more difficult" (Fr2). Finally, the uncertainty of functional overlaps between the program and the recently reformed PES-system compounded the case. The stakeholders had to envision "how it is going to match" (Fr1) and "prove that the experimentation can work with the reform" (Fr4). As a result, a *truncated scaling-up of the* program became a likely scenario (Fr4).

5.2. MAGMA

For MAGMA, scaling up in case of success was neither a prescribed nor an anticipated scenario. The program was launched as a temporary project with an intention to produce rigorous scientific evidence on policy performance that could be used at some point in the future, under a conducive political agenda (Au1-7). With that, MAGMA resembled the case of a "trailblazer" where policy learning and policy change are spaced out in time (2.2.1). Yet in this context, the initiator's limited decision-making autonomy (the authority and funds of local PES sufficed to pilot but not to upscale labour market innovations), coupled with the program's low political palatability, explained the case. As the agency representative noted, using pilots as "solutions in themselves" was a common practice for the agency: "we had different pilots for many years and, to be honest, it always comes back to money. There were really successful projects. But we just didn't have the money to scale them up, so they ended up temporary" (Au4).

The conservative government in office (Au1, Au3-4, Au7) and low unemployment rate in the region (Au5-6) reinforced the expectation that the program would remain unscaled: "the job guarantee program is clearly a social democratic idea. And now the government party [conservatives] is not interested in financing such projects" (Au3). Consequently, it was not a surprise for the stakeholders that evidence of policy success, despite its rigour, would fail to convince the government (Au1, Au3-4, Au7). As an interviewee explained, "even with this scientific proof, they [government] can always say: "you just measure what you want to measure." And I don't think they will be very much influenced by it" (Au4). Ultimately, the program was criticised by conservatives on alternative grounds (program financial efficiency) (Au1, Au3-4). MAGMA was declared "costly" (Au4), which could be partly misleading: "they [conservatives] declared to the Vice-Chancellor that the researchers wasted a lot of money, although there was no money involved" (Au1).

Along with the conducive agenda, the interviewees emphasised the importance of political facilitation to resume the program in the future (Au1-2, Au4, Au6-7; follow-up discussion). The advocacy leverage for initially politically neutral pilot was developed "from within" when the (former) head of initiating PES became a leading regional politician (head of the Social Democratic Party). As an interviewee commented, "at some point, you always have to involve politics to push it forward. And now we are at that point" (Au7) [the interview was conducted before the 2024 Austrian legislative election]. The change in leadership, however, created the

risk that the efforts to resume the program might not be supported by the new PES manager, who "is not from social democrats" (Au7).

Importantly, when discussing the program resumption, the interviewees referred to piloting renewal rather than a permanent adoption of the scheme. The initial scope of the pilot (150 residents), affordable within the disposable budget, was cited as a potential constraint on the program's scalability (Au3-6). As an interviewee noted, a better contextual awareness should be developed first: "I'm sure you would first have to adapt it in a city because it's just not the same. To get more learning there and then see if this is the way it should work" (Au4).

5.3. Kinofelis

For Kinofelis, scaling up of the scheme was driven by political commitment rather than experimental evidence (Gr1, Gr4, Gr6). The program's initiator (MOL) had sufficient decision-making autonomy to proceed with adopting it nationwide before the pilot evaluation was completed: "the program was for eight months, but four months later we rolled it out for the rest of the country," commented an interviewee (Gr1).

The ministerial office expected the program to be successful based on findings from prior research (Gr1, Gr3), and there was no need to corroborate these expectations further. Consequently, the evaluation of the program itself was aimed at *optimising* rather than testing the scheme. As several interviewees noted, despite being labelled a "pilot" (e.g., ILO 2018), the experimentation was more akin to a phase-wise adoption of the policy (Gr1) or policy prototyping (Gr2, Gr4), with the evaluation focused on *how* rather than *whether* the scheme works. One interviewee explained that "the aim of evaluation was to check if the project is being implemented right and whether any adjustment needs to be done" (Gr2). Accordingly, the pre-established nature of the program scaling-up left no room for valid criticism of its evaluation or effectiveness during the pilot phases.

The "shield" of political commitment, however, proved to be a short-term privilege. After the leadership turnover (2018) followed by the change in government from the radical left to the centre-right (2019), Kinofelis lost its political appeal (Gr1, Gr4-6). The new ministerial office did not support employment creation, neither ideologically (Gr1) nor did it regard it as an expedient expense item (Gr4). Consequently, the program was *scaled down* in scope, frequency, and quality of enrolments. As interviewees clarified, "they [new ministerial office] decided they would keep it, but very small: like, 30,000 people every 1.5 year" (Gr1), while "the training, the capacity building, the updates to the platforms that we did – they didn't happen in future implementations" (Gr4).

Nevertheless, the proactive leadership during the "pilot-phases" enabled ensuring that the program is (a) scalable (by targeted development of municipal capacities), (b) properly legitimised among regular PES and (c) creates a sustainably high demand from the unemployed (Gr1-2, Gr4-6). As a result, Kinofelis became "so popular among the local population" (Gr1) and "so well-entrenched in the system" (Gr4) that the scenario of its termination proved to be too politically expensive to implement (Gr1, Gr4-5). As an interviewee elaborated, "the new government explicitly stated that they want to abolish Kinofelis, the agenda and priorities have

changed... but they still keep it running, because abandoning it would entail such political costs that it would never be" (Gr4).

6. Discussion

The analysis has illustrated that the relations between the success and scaling-up in policy-piloting tend to be complex, dynamic, deeply contextualised and highly political – reflecting the rationale ("scaling-up in case/because of success") only to a limited extent. To structure and further discuss this empirical observation, the section applies the lenses of four discrepancies developed. Table 2 summarises the discussion.

The first discrepancy – *success may not be a prerequisite for upscaling* – finds support in the cases of Kinofelis and MAGMA. The programs corroborate both major scenarios: a pre-planned adoption and an expected termination of policy pilots. The Kinofelis case demonstrates that scaling up of pilot programs can be predefined when the initiating agency has sufficient decision-making autonomy (authority, funds) for that. Although labelled a "pilot" (ILO 2018), the program's national roll-out was decided already at the stage of its inception. The "pilot phases" and related evaluation were used to *optimise* the novel scheme rather than *test* it. The case of MAGMA, in turn, indicates that policy pilots can be launched as "solutions in themselves" when the initiators anticipate them to remain temporary irrespective of evaluation results. Specifically, the limited budget of local PES and an unfavourable political agenda (right-wing government, low unemployment rate in the region) created the initial expectations that MAGMA would likely be terminated after the pilot.

Along with that, the analysis has shown that policy pilots can pursue multiple purposes at once (as aligns with Ettelt, Mays and Allen 2015b). While the intention "to test" can be among these purposes (as in the case of MAGMA and arguably – TZCLD), pilots can also serve as *policy measures* aimed at resolving (a part of) the issue during experimentation (all three cases). Along with that, pilots can serve as *political tools* utilised to facilitate scaling-up of novel programs or "anchor" them at scale by creating political incentives/costs for the program adoption/termination. In particular, piloting can be employed to build an international support coalition for the novel program (TZCLD), acquire credible evidence to advocate for the program adoption under a conducive political agenda (MAGMA), or use the pilot term to ensure the program is effectively enshrined among regular public services and meets the needs of the target population (to create *sustainable* demand for it).

The second discrepancy – *success may not be proven to decision-makers* – is vividly corroborated by the examples of TZCLD and MAGMA, the programs where the success of the scheme *is to be* proven to higher-level policymakers to proceed with its further adoption. The programs indicate that the normative and political issues of pilot evaluations are deeply intertwined. The lack of common guidelines on *which* aspects of pilot policies should be evaluated can, and is likely to, be used in political interests. In both cases, policymakers had room to influence the outcomes of pilot evaluation irrespective of its rigour – either directly affecting the criteria of the official evaluation (TZCLD) or while criticising the pilot with alternative evidence (MAGMA). Consequently, the analysis suggests that the issue of *power-based* evidence prioritisation may become unresolvable for scaling-up ideologically/

ideationally contested pilots with the limited decision-making autonomy (as aligns with Mattocks 2021; Nesti, Cittadini and Bassoli 2025). Successful according to internal evaluations, neither TZCLD nor MAGMA managed to convince higher-level policymakers, who prioritised alternative evidence that resonated with their scepticism of the programs (respectively: official evaluation regulated by the state (TZCLD) and program financial assessments promoted by conservatives (MAGMA)).

Regarding the third discrepancy – *success may lack further funds and incentives* – the case study *altogether* demonstrates that the financial autonomy of policy pilots is a crucial starting point to understand why their scaling scenarios vary. While in the Kinofelis case, the ideological commitment of its leader *per se* provided sufficient grounds to proceed with scaling up the program nationwide, the prospects of TZCLD and MAGMA (led by no less committed leadership who, however, had no [TZCLD] or limited [MAGMA] financial autonomy) were threatened by the lack of political support from the higher-level policymakers regulating access to the corresponding funds. For TZCLD, the lack of support from the central government (mainly the MOL) complicated expanding the pilot to the new experimental territories and made its stakeholders doubt the chances of the pilot "to survive" after the second experimental term (Fr4). The risk that the program posed to the status quo of the MOL (due to the reallocation of resources from central- to local-level government, which it involved) proved to be a primary reason for that. For MAGMA, the issue of (low) ideological palatability came to the fore. Specifically, the lack of support for "clearly social democratic idea" (Au3) from the MOL and the state-level PES, who followed the agenda of the conservative government, prevented the program from acquiring funds for its extension and/or expansion.

At the same time, the analysis emphasises that political/ideological palatability of policy pilots is dynamic and changes along with the government/electoral cycles (all programs) and leadership turnovers (Kinofelis, MAGMA). For MAGMA, the potential victory of social democrats at the legislative elections was expected to provide a favourable agenda to resume the program, while the change in agency leadership was perceived as a risk for this scenario. For TZCLD, the various-level elections (legislative/presidential/European) were anticipated to influence the "appeal" of the scheme for the state, shaping its support for the program accordingly. For Kinofelis, the change in government from left to right and leadership turnover caused a substantial downscaling of the program. Yet, the case has also shown that the policymakers' reluctance to make electorally inexpedient decisions may allow for preserving politically/ideologically unpalatable (pilot) programs – suggesting that the tendency of policymakers to favour the status quo over policy change (2.2.3) may *also*[4] work "for" the pilot upscaling.

The fourth discrepancy – *success may not be scalable* – is supported by MAGMA and TZCLD. For MAGMA, the small scope of the pilot and, in part, its dominant approach to evaluation (preregistered RCT) questioned the feasibility of the program's national roll-out at the next step. While discussing the desired scenario for the pilot under a conducive political agenda,[5] the interviewees referred to promoting the experimentation to other Austrian territories, since "a better awareness of local contexts should be developed first" (Au4). Besides, the case also indicates that the research design of policy pilots may stem from budgetary constraints

4 In contrast to literature insights in 2.2.3.
5 Where social democrats would have won the most votes in the 2024 Austrian legislative election.

rather than (solely) normative preferences of initiating agencies, implying that the limited (access to) funding may constrain pilot scalability as well.

TZCLD corroborates the discrepancy in several respects. First, the high political polarisation of policymakers at the local level (across and within French territories) challenged the scalability of the program. The expansion of the pilot with initially "leftish" connotation risked not happening, had its leaders failed to appeal, reconcile and ensure effective collaboration of policymakers on both sides of the political spectrum. The case, however, also indicates that the challenge of political fragmentation/polarisation is resolvable and, moreover, could provide strategic opportunities for the prospects of the pilot programs (such as the resilience to political cycles, "demonstration effect" facilitating pilot expansion, stronger advocacy leverage to request resources from the state). Second, the program shows that the institutional set-up of policy pilots may constrain their scalability. Unlike MAGMA and Kinofelis, the implementation and governance of the French employment-creation scheme relied on novel institutions that first had to be set up in each experimental territory. As a result, the stakeholders were worried it might not be feasible to implement the scheme in the same form nationwide due to the issues of integration (novel vs existing institutions), coordination and transparency.

Table 2: Why successful policy pilots are not scaled up: empirical summary

Discrepancies between rationale and practice	Reasons identified
Success may not be a prerequisite	Scaling-up may be predefined: piloting for prototyping /adopting phase-wise (Kinofelis) Scaling-up may not be *the only* purpose /pilots as "solutions in themselves": - policy measures: piloting for solving issues during experimentation (all) - political tools/facilitators: piloting to create incentives/termination cost (all) - "trailblazers" (MAGMA)
Success may not be proven to decision-makers	Relativity of success and failure in policy-piloting (TZCLD, MAGMA): - normative and political issues are intertwined: the lack of evaluation guidelines can be used in political interests - ideological/ideational preconceptions are embedded in evaluations - issue of multiple evidence: conflicting evidence can coexist - issue of evidence prioritisation: pilot evidence can be devalued with alternative, more palatable
Success may lack funds and incentives	Limited financial autonomy of pilot initiators against low political support (TZCLD, MAGMA): - limited disposable budget and ideological incompatibility with the ongoing government agenda (MAGMA) - dependence on state funding and threat posed to the status quo (power, disposable resources) of decision-makers (TZCLD) Loss of appeal due to government/election cycles and leadership change (all)

Success may not be scalable	Pilots may not yield the same results at scale: - limited external validity: scope, evaluation choices, budgetary constraints (MAGMA) - political fragmentation/polarisation: varying "leanings" of local policymakers/ local-level dominance of different parties (TZCLD) - constraints of operational design: reliance on experimental institutions and related challenges (integration, coordination, transparency at scale) (TZCLD)

7. Conclusions

This study has systematically explored why successful policy pilots *are* and *are not* scaled up in policymaking realities. Drawing on the synthesis of existing scholarship, the study developed four theoretical propositions for the matter and applied them as lenses for subsequent case-study analysis. The analysis corroborated all the propositions, illustrating that the rationale of "scaling-up in case/because of success" in practice is subject to numerous limitations. On the one hand, pilot evidence that the policy "works" may not be necessary, convincing or politically appealing to proceed with its adoption beyond experimentation. On the other hand, successful as a pilot, the policy may fail to replicate the same results at scale.

While all four propositions found support in the comparative case-study, the reasons behind them (Table 2) only partially overlapped with the ones suggested by the literature (Table 1), were prioritised in a certain way and/or varied among the programs studied – indicating the importance of the piloting context for the prospects of novel policies. By comparing theoretical propositions with empirical findings (Section 6), the study has identified the key *policy*- and *program*-specific contextual factors which, while interacting with one another, prove to shape scaling-up of three employment-creation schemes analysed.

As a key policy-specific factor identified, the study emphasises the ideological/ideational preferences of policymakers. In all three cases, policymakers' preconceptions about employment-creation schemes motivated the launch of the pilots (referring to the initiators' commitment to the schemes), affected their evaluation results (referring to conflicting evidence (interpretation) and its embedded bias) and ultimately, largely influenced the programs' scaling-up – facilitating or hindering their extension, nationwide adoption and maintenance at scale. By contrast, the importance of ideology for the outcomes of policy experiments has received only limited attention in the existing discussions.[6] This is likely linked to a higher political complexity and thus the lack of experimenting with those policy solutions or in those policy domains/areas which involve high polarisation of political opinions over the issues tackled or approaches utilised for that – as the case-study exemplifies and aligns with Béchard et al. (2023).

As a key program-specific factor, the study indicates the importance of the institutional design of policy pilots. Specifically, such pilot characteristics as their modes of launch (top-down vs bottom-up) and related decision-making/financial autonomy of initiators, types of *institutions employed to operate them* (conventional vs experimental entities, established

6 With few identified exceptions examining the role of party leanings and/or narrative diversification in promoting (experimentation with) minimum/basic income schemes (Jessoula and Natili.2020; Parth and Nyby 2022; Van Parijs and Vanderborght.2017).

vs new collaborations), and *learning strategies adopted* (including the choices of research design, evaluation focus and evaluators) proved to shape their scaling-up scenarios – as resonates with the findings of McFadgen and Huitema (2017, 2018).

Altogether, the differences identified between (a) the programs analysed and (b) the theoretical propositions and empirical findings in no way claim to be attributed *solely* to the context of policy pilots, but may also stem from the limitations of empirical data (such as its interpretative nature, inability to "probe" all the reasons that could underlie the discrepancies, lack of access to certain interviewees), limits of literature review or knowledge gaps in existing scholarship. Besides, the overall generalisability of the findings offered is likely to be limited given the inductive nature of the study and the research design applied (small-N comparative case-study, N = 3).

Thus, future studies would play an important role in strengthening the findings of this study further – in line with the continuously evolving knowledge formation in qualitative research. The framework of four discrepancies could be employed in future comparative investigations to continue holistic exploration of the factors, mechanisms and patterns of scaling-up (or lack thereof) in experimental policymaking. Besides, more research is needed to explore how the choices of institutional design affect the prospects of policy pilots, and how politics mediates the way pilot policies are institutionalised.

In future research on scaling-up of experimental policies, the study proposes to conceptualise "scaling-up" in broader terms, i.e., as a *lasting two-dimensional process*.[7] Taking a longer perspective would enable to explore the factors that may: (a) have different effects depending on the timeframe considered (e.g., risk-aversion of policymakers, political diversity of pilot stakeholders), and (b) cause reduction or termination of pilot policies after their adoption at scale – embracing the issues of "downscaling" as another essential part of the pilot adoption journey – as aligns with Sondal, Hellquist and Balfors (2024). Incorporating both time-and scope-dimensions in the analysis would allow distinguishing between the expansion and prolongation challenges, which, although largely overlapped, could be different in their substance, effects or reasons.

Acknowledgments

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

References

Abramowicz, M. 2019. "Tax Experimentation." *Florida Law Review* 71(1), 65-122.

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

Bailey, S. et al. 2017. "The Policy Work of Piloting: Mobilising and Managing Conflict and Ambiguity in the English NHS." *Social Science & Medicine* 179, 210-217.

7 Contrary to existing scholarship that mainly associates the term with the *piloting expansion* (e.g., adopting similar pilots to other municipalities) or the *ultimate outcome* of experimentation (such as permanent adoption of the novel tool) – see Reymond et al. 2020; Spicer et al. 2014; Vreugdenhil and Rault 2010 as examples for both.

Bailey, S., D. Hodgson and K. Checkland. 2019. "Pilots as Projects: Policymaking in a State of Exception." In D. Hodgson et al. (eds). *The Projectification of the Public Sector.* UK: Routledge, 130-148.

Baird, J. et al. 2014. "Learning Effects of Interactive Decision-Making Processes for Climate Change Adaptation." *Global Environmental Change* 27, 51-63.

Bédécarrats, F., I. Guérin and F. Roubaud. 2019. "All that Glitters is not Gold. The Political Economy of Randomized Evaluations in Development." *Development and Change, International Institute of Social Studies* 50(3), 735-762.

Beets, M. et al. 2020. "Identification and Evaluation of Risk of Generalizability Biases in Pilot Versus Efficacy/Effectiveness Trials: A Systematic Review and Meta-analysis." *International Journal of Behavioral Nutrition and Physical Activity* 17, 19.

Behaghel, L., K. Macours and J. Subervie. 2019. "How can Randomised Controlled Trials Help Improve the Design of the Common Agricultural Policy?" *European Review of Agricultural Economics* 46(3), 473-493.

Beisheim, M. et al. 2025. "Politics Matters! Political will as a Critical Condition for Implementing the Sustainable Development Goals." *Earth System Governance* 24: 100244.

Bello, B. et al. 2018. "Does Austerity Drive Public Service Innovation? Evidence from Shared Senior Management Teams in Local Government." *Public Money & Management* 38(2), 131-138.

Béchard, B. et al. 2023. "Political Complexity and the Pervading Role of Ideology in Policy-Making." *Journal of Dynamic Decision Making*, 9(1).

Binz, Ch. et al. 2016. "The Thorny Road to Technology Legitimation. Institutional Work for Potable Water Reuse in California." *Technological Forecasting and Social Change* 103, 249-263.

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

Bueno, N. S. 2023. "The Timing of Public Policies: Political Budget Cycles and Credit Claiming." *American Journal of Political Science* 67(4), 996-1011.

Bulthuis, S. et al. 2020. "Factors Influencing the Scale-up of Public Health Interventions in Low- and Middle-income Countries: A Qualitative Systematic Literature Review." *Health Policy Plan* 35(2), 219-234.

Cairney, P. 2022. "The Myth of 'Evidence-Based Policymaking' in a Decentred State." *Public Policy and Administration* 37(1), 46-66.

Cairney, P. and K. Oliver. 2017. "Evidence-based Policymaking is not like Evidence-Based Medicine, so how far should you go to Bridge the Divide Between Evidence and Policy?" *Health Research Policy and Systems* 15, 35.

Callander, S. 2011. "Searching for Good Policies." *American Political Science Review* 105(4), 643-662.

Capano, G. and A. Lippi. 2017. "How Policy Instruments are Chosen: Patterns of Decision Makers' Choices." *Policy Sciences* 50(2), 269-293.

Checkland, K. et al. 2021. "'Success' in Policy Piloting: Process, Programs, and Politics." *Public Administration* 101(2), 463-480.

Cinar, E., P. Trott and C. Simms. 2019. "A Systematic Review of Barriers to Public Sector Innovation Process." *Public Management Review* 21(2), 264-290.

Corduneanu-Huci, C., M. Dorsch and P. Maarek. 2021. "The Politics of Experimentation: Political Competition and Randomized Controlled Trials." *Journal of Comparative Economics* 49(1), 1-21.

Creswell, J.W. and C. N. Poth 2023. "Qualitative Inquiry and Research Design: Choosing Among Five Approaches." Thousand Oaks, CA: SAGE. 552.

De Wispelaere, J., A. Halmetoja and V.-V. Pulkka. 2023. "A Primer on the Finnish Basic Income Experiment: From Design and Implementation to Evaluation and Impact." In M. Torry (ed.). *The Palgrave International Handbook of Basic Income.* Cham: Palgrave Macmillan, 413-435.

Eboreime E. A. et al. 2021. "Appraising and Addressing Design and Implementation Failure in Global Health: A Pragmatic Framework." *Global Public Health* 16(7), 1122-1130.

Eneqvist, E. 2022. "Experimental Governance: Capacity and Legitimacy in Local Governments." PhD Thesis, KTH Royal Institute of Technology.

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.

Gillespie, S. 2004. "Scaling-up Community-Driven Development: A Synthesis of Experience." FCND Discussion Paper 181.

Grin, J. 2020. "'Doing' System Innovations from within the Heart of the Regime." *Journal of Environmental Policy & Planning* 22(5), 682-694.

Haynes, L., B. Goldacre and D. Torgerson. 2012. "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials." *SSRN Electronic Journal*, ssrn.2131581.

Head, B. 2010. "Reconsidering Evidence-Based Policy: Key Issues and Challenges." *Policy and Society* 29(2), 77-94.

Hong, S. and S. Lee. 2018. "Adaptive Governance, Status Quo Bias, and Political Competition: Why the Sharing Economy is Welcome in Some Cities but not in others." *Government Information Quarterly* 35(2), 283-290.

Howlett, M. and C. Leong. 2022. "The 'Inherent Vices' of Policy Design: Uncertainty, Maliciousness, and Noncompliance." *Risk Analysis* 42, 920-930.

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

Huitema, D. et al. 2018. "Policy Experiments: Core Concepts, Political Dynamics and Impacts." *Policy Sciences* 51(2), 143-159.

ILO. 2018. "Towards a Stronger Kinofelis: Recommendations for the Strengthening of Future Phases." ILO Report.

Jacobson, M. and T. Azzam. 2018. "The Effects of Stakeholder Involvement on Perceptions of an Evaluation's Credibility." *Evaluation and Program Planning* 68, 64-73.

Jessoula, M. and M. Natili. 2020. "Explaining Italian 'Exceptionalism' and its End: Minimum Income from Neglect to Hyperpoliticization." *Social Policy and Administration* 54, 599-613.

Jowell, R. 2003. "*Trying it Out: the Role of 'Pilots' in Policy-Making: Report of a Review of Government Pilots*." London, UK: Cabinet Office, Strategy Unit.

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

Koon, A. D. et al. 2020. "A Scoping Review of the Uses and Institutionalisation of Knowledge for Health Policy in Low- and Middle-income Countries." *Health Research Policy and System* 18, 7.

Lam, D. P. M. et al. 2020. "Scaling the Impact of Sustainability Initiatives: A Typology of Amplification Processes." *Urban Transformations* 2, 3.

Lehner, L. and M. Kasy. 2024. "Employing the Unemployed of Marienthal: Evaluation of a Guaranteed Job Program," IZA Discussion Paper 16088.

Leino, H. and M. Åkerman. 2021. "The Politics of Making Finland an Experimenting Nation." *Critical Policy Studies* 16(4), 441-459.

Leong, C. and M. Howlett. 2022. "Policy Learning, Policy Failure, and the Mitigation of Policy Risks: Re-Thinking the Lessons of Policy Success and Failure." *Administration & Society* 54(7), 1379-1401.

Lewis, J., 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.

Liarte, I., J. I. Criado and L. Alcaide-Muñoz. 2024. "Exploring Barriers to Innovation in Public Administration: An Empirical Study of the Local Layer of Government." *IJPADA* 11(1), 1-23.

Maddison, S. 2012. "Evidence and Contestation in the Indigenous Policy Domain: Voice, Ideology and Institutional Inequality." *Australian Journal of Public Administration* 71, 269-277.

Markowitsch J. and Á. Scharle. 2024. "Towards Zero Long-term Unemployment in the EU: Job Guarantees and other Innovative Approaches." Brussels: European Commission. Report.

Marston, G. and R. Watts. 2003. "Tampering with the Evidence: A Critical Appraisal of Evidence-Based Policymaking." *Australian Review of Public Affairs* 3(3), 143-163.

Martin, S. and I. Sanderson. 1999. "Evaluating Public Policy Experiments: Measuring Outcomes, Monitoring Processes or Managing Pilots?" *Evaluation* 5(3), 245-258.

Mattocks, K. 2021. "Policy Experimentation and Policy Learning in Canadian Cultural Policy." *Policy Sciences* 54(4), 891-909.

Mauw, T., S. Smith and J. Torrens. 2023. "Sustainability Transitions in Los Angeles' Water System: The Ambivalent Role of Incumbents in Urban Experimentation." *Journal of Environmental Policy & Planning* 25(4), 368-385.

McConnell, A. 2010. "Policy Success, Policy Failure and Grey Areas in-Between." *Journal of Public Policy* 30(3), 345-362.

McConnell, A., L. Grealy and T. Lea. 2020. "Policy Success for Whom? A Framework for Analysis." *Policy Sciences* 53(4), 589-608.

McFadgen, B. and D. Huitema. 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, 161-187.

Muralidharan, K. and P. Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31(4), 103-124.

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

Nesti, G., M. Cittadini and M. Bassoli. 2025. "The Politics of Piloting. The Case of Minimum Income Schemes in European Cities." *Review of Policy Research* 42, 291-309.

OECD. 2020. "Building Capacity for Evidence-Informed Policymaking: Lessons from Country Experiences." *OECD Public Governance Reviews*. Paris: OECD Publishing.

Parth, A. M. and J. Nyby. 2022. "Welfare Experiments as Tools for Evidence-Based Policy Making? The Political Debate on Twitter about the Basic Income Trial in Finland." *Policy Studies* 43(4), 772-790.

Partners for Health Reform Plus. 2004. "*The Role of Pilot Programs*: *Approaches to Health Systems Strengthening*." Bethesda, MD: PHRPlus, Abt Associates.

Patton, M. 2003. "Utilization-Focused Evaluation." In T. Kellaghan and D. Stufflebeam (eds). *International Handbook of Educational Evaluation*. Dordrecht: Springer, 223-242.

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

Perkiö, J. 2020. "Legitimising a radical policy idea: Framing basic income as a boost to labour market activity". *Policy & Politics*, 48(2), 277-293.

Raudla, R. et al. 2023. "To Experiment or not to Experiment in Tax Policy?" *Halduskultuur*, 23(1), 27-48.

Raudla, R. et al. 2024. "It is About Time! Exploring the Clashing Timeframes of Politics and Public Policy Experiments." *Perspectives on Public Management and Governance* 7 (4), 137-148.

Raudla, R. et al. 2025a. "Budgeting for Innovation: How Funding Strategies Enable or Constrain Public Sector Experimentation." *Public Budgeting & Finance*, 45(4), 28-40.

Raudla, R. et al. 2025b. "The Politics of Experimental Policymaking: The Influence of Blame Avoidance and Credit Claiming." *Policy Sciences*, 58, 245-266.

Reymond, Ph. et al. 2018. "Scaling-up Small-Scale Wastewater Treatment Systems in Low- and Middle-Income Countries: An Analysis of Challenges and Ways Forward Through the Case of Egypt." *Utilities Policy* 52, 13-21.

Reymond, Ph., R. Chandragiri and L. Ulrich. 2020. "Governance Arrangements for the Scaling Up of Small-Scale Wastewater Treatment and Reuse Systems – Lessons from India." *Frontiers in Environmental Science* 8, 72.

Roseland D, F. Lawrenz and M. Thao. 2015. "The Relationship Between Involvement in and Use of Evaluation in Multi-Site Evaluations," *Evaluation and Program Planning* 48, 75-82.

Ruggeri, K. 2025. "Assessing Evidence Based on Scale can be a Useful Predictor of Policy Outcomes." *Policy Sciences* 58, 179-188.

Ruggeri, K. et al. 2020. "Standards for Evidence in Policy Decision-Making." Nature Research: Social and Behavioural Sciences, 39905.

Sanderson, I. 2002. "Evaluation, Policy Learning and Evidence-Based Policymaking." *Public Administration* 80(1), 1-22.

Simmons, R., P. Fajans and L. Ghiron. 2007. "*Scaling-up Health Service Delivery: From Pilot Innovations to Policies and Programmes*." Geneva: WHO.

Sondal, J., A. Hellquist and B. Balfors. 2024. "From Pilot to Practice: Navigating Pre-Requisites for Up-Scaling Sustainable Urban Solutions." *Urban Transformations* 6, 7.

Spicer, N. et al. 2014. "'Scaling-up is a Craft not a Science': Catalysing Scale-up of Health Innovations in Ethiopia, India and Nigeria." *Social Science & Medicine* 121, 30-38.

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 M. Evans. 2016. Evidence-Based Policy Making and Social Science. In G. Stoker and M. Evans (eds). *Evidence-Based Policy Making in the Social Sciences*. Bristol: Policy Press, 15-28.

Strassheim, H. and P. Kettunen. 2014. "When Does Evidence-Based Policy Turn into Policy-Based Evidence? Configurations, Contexts and Mechanisms." *Evidence & Policy* 10(2), 259-277.

Tcherneva, P. 2019. "The Federal Job Guarantee: Prevention, Not Just a Cure." *Challenge* 62(4), 253-272.

Tcherneva, P. and A. Lalucq. 2022. "A Job Guarantee for Europe: Economic Policy Innovation for Stability, Security and Sustainability." Brussels: Foundation for European Progressive Studies (FEPS). Policy Brief.

Torrens, J. and T. von Wirth. 2021. "Experimentation or Projectification of Urban Change? A Critical Appraisal and Three Steps Forward." *Urban Transformations* 3(1), 8.

Torrens, J., P. Johnstone and J. Schot. 2018. "Unpacking the Formation of Favourable Environments for Urban Experimentation: The Case of the Bristol Energy Scene." *Sustainability* 10(3), 879.

Turnheim, B. and B. Sovacool. 2020. "Forever Stuck in Old Ways? Pluralising Incumbencies in Sustainability Transitions." *Environmental Innovation and Societal Transitions* 35, 180-184.

van der Heijden, J. 2014. "Experimentation in Policy Design: Insights from the Building Sector." *Policy Sciences* 47, 249-266.

Van Parijs, P. and Y. Vanderborght. 2017. "*Basic Income. A Radical Proposal for a Free Society and a Sane Economy*". Cambridge, MA: Harvard University Press.

Varazzani, C. et al. 2023. "Seven Routes to Experimentation in Policymaking: A Guide to Applied Behavioural Science Methods", *OECD Working Papers*, 64.

Verkerk, J. and A. Van Buuren. 2017. "Developing Collaborative Capacity in Pilot Projects Lessons from Three Dutch Flood Risk Management Experiments." *Journal of Cleaner Production*, 169, 225-233.

Vetulani-Cęgiel, A. and A. K. Cianciara. 2025. "Informational Lobbying in a Polish-German Comparative Perspective: Uncovering Perceptions of Political Relevance of Expert Knowledge and Lobbying Tactics." *Journal of Contemporary European Studies*, 33(4), 1349-1371.

Vreugdenhil, H. and P. Rault. 2010. "Pilot Projects for Evidence-Based Policy Making: Three Pilot Projects in the Rhine Basin." *German Policy Studies* 6(2), 1-42.

Vreugdenhil, H. et al. 2009. "The Next Step in Policy Transitions: Diffusion of Pilot Projects." *IRSPM XII Conference Paper*, 1-25.

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. Ule (eds). *Handbook of Research Methods and Applications in Experimental Economics*. Cheltenham: Edward Elgar, 295-311.

Wesselink, A., H. Colebatch and W. Pearce. 2014. "Evidence and Policy: Discourses, Meanings and Practices." *Policy Sciences* 47, 339-344.

Zhu, X. and H. Zhao. 2021. "Experimentalist Governance with Interactive Central-Local Relations: Making New Pension Policies in China." *Policy Studies Journal* 49(1), 13-36.

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

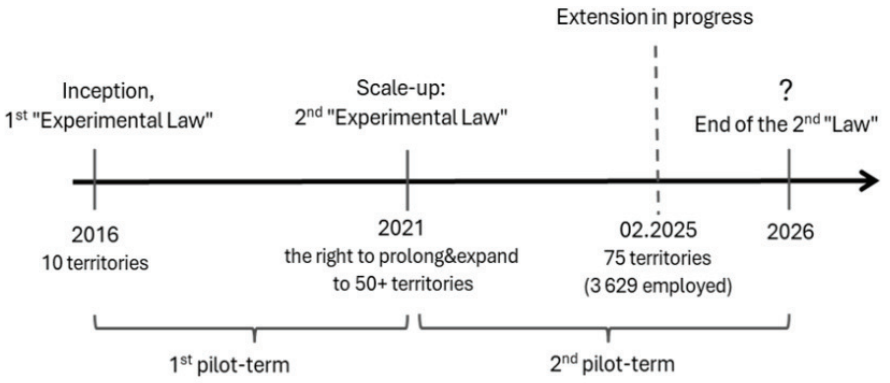
Appendices

Appendix 1. Anonymised list of interviewees

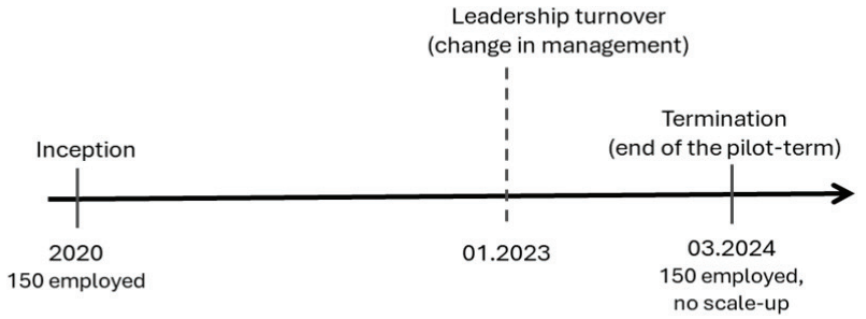
Reference	Role in the program	Interview date
	TZCLD, France	
Fr1	National Research & Evaluation Monitoring, Lead	02.06.2023
Fr2	National Management, Vice-President	28.06.2023
Fr3	Funding Management, Board of Directors	01.09.2023
Fr4	National Management & Implementation, Lead	05.10.2023
Fr5	Local Management & Implementation, Lead	19.10.2023
	MAGMA, Austria	
Au1	Research/Evaluation	30.05.2023
Au2	Research/Evaluation	07.06. 2023
Au3	Research/Evaluation	28.06.2023
Au4	Implementation, Lead	29.06.2023
Au5	Implementation, Lead	27.07. 2023
Au6	Implementation, Executive	27.07.2023
Au7	Initiator Representative	12.10.2023
	Kinofelis, Greece	
Gr1	Initiator	26.10.2023
Gr2	Design/Evaluation, Remote Executive	13.11.2023
Gr3	Research	15.11.2023
Gr4	Design, Implementation, Evaluation: On-site Coordination, Lead	16.11.2023
Gr5	Design, Implementation, Evaluation: Remote Coordination, Lead	08.12.2023
Gr6	Design and Technical Advice	20.01.2024

Appendix 2. Program timelines

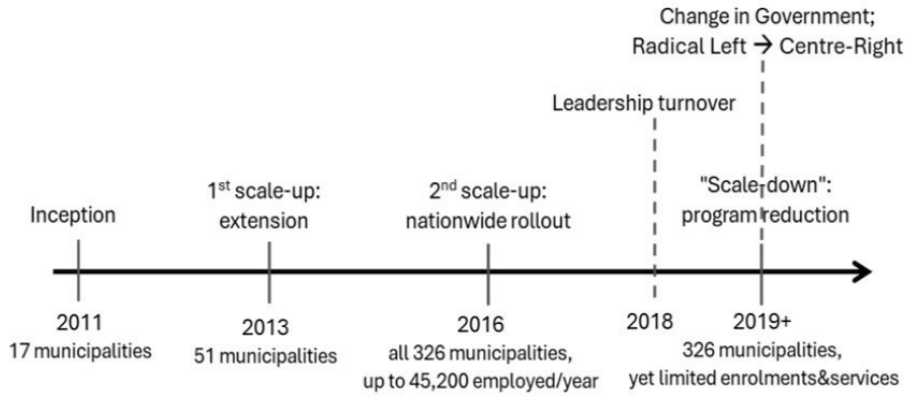
TZCLD, France



MAGMA, Austria



Kinofelis, Greece



Appendix 2

Publication II

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. (1.1)

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.

Appendix 3

Publication III

Raudla, R., Sarapuu, K., Vallistu, J., **Harbuzova, N.** (2024). It is about time! Exploring the clashing timeframes of politics and public policy experiments. *Perspectives on Public Management and Governance,* 7(4), 137–148. (1.1)

Perspectives on Public Management and Governance, 2024, **XX**, 1–12
https://doi.org/10.1093/ppmgov/gvae008
Advance access publication 16 October 2024
Article

OXFORD

It is about time! Exploring the clashing timeframes of politics and public policy experiments

Ringa Raudla*, Külli Sarapuu, Johanna Vallistu, Nastassia Harbuzova

Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Akadeemia 3, Tallinn, 12618, Estonia
*Corresponding author: Ragnar Nurkse Department of Innovation and Governance, Tallinn University of Technology, Akadeemia 3, Tallinn, 12618, Estonia.
Email: ringa.raudla@taltech.ee

Abstract

Although existing studies on experimental policymaking have acknowledged the importance of the political setting in which policy experiments take place, we lack systematic knowledge on how various political dimensions affect experimental policymaking. In this article, we address a specific gap in the existing understanding of the politics of experimentation: how political timeframes influence experimental policymaking. Drawing on theoretical discussions on experimental policymaking, public policy, electoral politics, and mediatization of politics, we outline expectations about how electoral and problem cycles may influence the timing, design, and learning from policy experiments. We argue electoral timeframes are likely to discourage politicians from undertaking large-scale policy experiments and if politicians decide to launch experiments, they prefer shorter designs. The electoral cycle may lead politicians to draw too hasty conclusions or ignore the experiment's results altogether. We expect problem cycles to shorten politicians' time horizons further as there is pressure to solve problems quickly. We probe the plausibility of our theoretical expectations using interview data from two different country contexts: Estonia and Finland.

Key words: policy experiments; experimental policymaking; politics of experiments; timeframes.

INTRODUCTION

The increasing complexity of economic, environmental, and societal problems has pressured governments to seek novel ways to adapt existing public policies, as well as design new policy solutions. Experimentation has been proposed as a key strategy for dealing with complexity and producing policies that better address acute challenges (e.g., Ansell and Bartenberger 2016; Bravo-Biosca 2020; McGann, Blomkamp, and Lewis 2018; Pearce and Raman 2014; Rangoni and Zeitlin 2021). Experimental approaches promise to provide a better understanding of policy problems that are characterized by high levels of uncertainty (Bravo-Biosca 2020). Numerous studies point to the increasing use of experimentation in contemporary policymaking (e.g., Ansell and Bartenberger 2016; Lee and Ma 2020; Pearce and Raman 2014).

Various definitions are used for delineating policy experiments. We adopt a definition of experiments as *time-limited tests of new policy solutions that provide information for further policy decisions* (Bravo-Biosca 2020, p. 195; McFadgen and Huitema 2018; Nair and Howlett 2016, p. 69). This definition is intentionally broad to accommodate different experimental approaches and satisfies the two core criteria put forward by McFadgen and Huitema (2018): to call something an experiment, there should be an intervention theory with explicit assumptions, which are tested, and the tested solution should be novel. Importantly, experiments have a clear timeframe and an explicit *ex ante* strategy for assessing the effects of the intervention (Ibid.; Bravo-Biosca 2020, p. 195). Our focus is on *field experiments* initiated and interpreted by governmental policy actors and not on laboratory experiments conducted by researchers. We concentrate on strategic large-scale policy experiments, which span several years and involve political decision making.

Despite high expectations, existing studies have often found that large-scale policy experiments face various impediments. Nair and Howlett (2016, p. 72) argue the obstacles have a lot to do with "the politics of policy experiments"—various political considerations featuring throughout the experimental process and posing "continual challenges to effective experimentation." From the "political learning" perspective, postulated by Corduneanu-Huci, Dorsch, and Maarek (2021, p. 4), electoral competition should encourage politicians to undertake more experiments since the knowledge gained from such experiments may offer politicians electoral advantages by signaling voters their "intentions to implement policies based on rigorous evidence, address criticism, and establish objectivity in politically contentious issues." The reported experience with policy experiments, however, indicates a much more complex picture: choices of policy action and interpretation of evidence have been strongly influenced by what is "politically preferred, conducive or acceptable" (Nair and Howlett 2016, p. 72). Despite extensive scholarly discussions on policy experimentation, however, there is surprisingly little systematic analysis and theoretical conceptualization of the *political dimensions* of public policy experiments.

A core characteristic of all experiments is that they have "timeframes set from the start to assess results and make decisions" (Bravo-Biosca 2020, p. 197). However, some studies have recognized that there is a potential "temporal disjoint between establishing evidence and demands for political action" (Pearce and Raman 2014, p. 393). As noted by Stoker (2010, p. 53), "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." The timeframes inherent to large-scale experiments (spanning several years) can be expected to interact with other temporalities in the politico-administrative system. The interaction of these temporalities may shape the experimental design, process, and results in unanticipated ways.

Hence, in this article, we zoom in on a specific challenge that the political setting poses for experimental policymaking: the potential clashes between political and experimental *timeframes*. Although the challenges related to timeframes have been tangentially mentioned by several studies discussing public policy experimentation, so far, they have not been investigated in a *systematic* way—either theoretically or empirically. In our article, we seek to address this important gap in existing research. The key research question of the article is: How do the timeframes inherent to politics influence experimental policymaking? In particular, we are interested in the *challenges* the political timeframes pose for large-scale policy experiments.

Our analytical strategy for tackling these questions is as follows. In the theoretical part of the article, we outline the key clashes between experimental and political temporalities that can be encountered in experimental policymaking. We do so by synthesizing ideas from the theoretical discussions on experimental policymaking, public policy, electoral politics, and mediatization of politics.[1] In the empirical part, we probe the plausibility of the theoretical expectations and examine how these clashes are expressed in the perceptions of public officials about experimental policymaking in two different country contexts: Estonia and Finland. As sources of data, we use 66 interviews conducted with public officials in Estonia and Finland in 2022–2023.

The novel theoretical contribution of our article is to put forth a framework that discusses the impacts of the various timeframes inherent to politics (electoral cycles and problem cycles) on experimental policymaking in a systematic and holistic way. Our empirical contribution is to test the plausibility of the theoretical expectations and to provide an empirical understanding of how the political timeframes influence experimental policymaking in different empirical contexts.

THEORETICAL DISCUSSION

The Promise and Nature of Public Policy Experiments

The growing pressures on governments to be "smarter," more "innovative," and "agile" have spurred a call for an increasingly experimental stance or even an "experimental turn" in public policy making (e.g., Ansell and Bartenberger 2016; Ettelt, Mays and Allen 2015; Lee and Ma 2020; McGann, Blomkamp and Lewis 2018; Rangoni and Zeitlin 2021; Raudla et al. 2023). The benefits of using experimentalist approaches are argued to be manifold. Testing the usefulness of novel ideas and solutions in a real-world setting before applying them on a larger scale is expected to increase the quality of policies by providing knowledge that is otherwise not available (Checkland et al. 2023; Ettelt, Mays, and Allen 2015; Farrelly 2008; Millo and Lezaun 2006). Furthermore, experimentation is viewed as a way to adjust to change in an increasingly uncertain, complex, and fast-changing social, economic, and environmental context, and to improve the future-readiness of the planned policies (Ansell and Bartenberger 2016; Bravo-Biosca 2020; Lee and Ma 2020; Millo and Lezaun 2006; Nair and Howlett 2016). Large-scale policy experiments have been undertaken in various policy fields, including education, employment, welfare, health, environmental sustainability, and fiscal policies. For example, policymakers have experimented with universal basic income, housing allowances, income maintenance, welfare-to-work, negative income tax, and water management (Ettelt, Mays, and Allen 2015; Kangas et al. 2021; Nair and Howlett 2016; Oakley 1998).

A wide range of experimental approaches are available for governments to fulfill these aims. By and large, they can be divided into three ideal types—randomized controlled trials (RCTs), non-randomized policy pilots, and design experiments (Raudla et al. 2023). RCTs offer experimenters the opportunity to draw valid causal conclusions about the effects of a project, program, or policy by randomly dividing the target population into two or more groups and treating them differently (Bravo-Biosca 2020; Pearce and Raman 2014). RCTs are often presented as the "gold standard" of experiments (e.g., Pearce and Raman 2014). Non-randomized policy pilots test a new policy approach on a small subset of population or jurisdictions and allow the introduction of a policy in a phase-wise manner (Farrelly 2008; Nair and Howlett 2016). In design experiments, a policy solution is iteratively refined on the basis of feedback gathered from the affected individuals, until acceptable results emerge (Ansell and Bartenberger 2016; Stoker and John 2009). In this article, we focus on large-scale experiments, which span several years and necessitate the involvement of politicians (e.g., for legislative mandate, funding, and public legitimization).

The Political Setting of Experiments

Huitema et al. (2018, p. 148) argue that in the policy sciences, experiments have been predominantly conceptualized as *a research method*. According to the "research logic" of public policy experiments, the experimentation would proceed in the following steps: identifying the policy problem, exploring potential ideas and solutions, selecting idea(s) for testing, choosing the appropriate experimental design, testing whether and how the solutions work, evaluating the evidence, deciding on the further policy steps (discontinue, scale-up, or new experiment), and disseminating the knowledge gained (e.g., Bravo-Biosca 2020; Cobb et al. 2003). As such, public policy experiments hold a promise of a rational approach to policy making (Checkland et al. 2023), "removed from the messy world of politics" (Rogers-Dillon 2004, p. 24).

[1]The mediatization of politics has been defined as "a long-term process through which the importance of the media and their spill-over effects on political processes, institutions, organizations and actors have increased" (Strömbäck and Esser 2014, p. 6).

However, since policymaking is an inherently political process, the reality of policy experimentation is much more complex than the "rational" ideal implies (e.g., Checkland et al. 2023; Rogers-Dillon 2004). Experiments can be exposed to political impacts throughout the entire process, and politics may undercut the aforementioned linear sequence of steps (e.g., Bédécarrats, Guérin, and Roubaud 2019; McGann, Blomkamp, and Lewis 2018). Indeed, the existing literature acknowledges that public policy experiments are not neutral evidence-creating research activities, and their interpretation is "an inescapably political process" (Huitema et al. 2018, p. 148). The political setting can influence "who gets to formulate the ideas, who is involved in producing the evidence on their efficacy, which kinds of information should be collected, and which rules of evidence are used" (Ibid., p. 145). Thus, political motivations are likely to affect the choice and set-up of experiments and the interpretation of results. In the following, we zoom in on a key challenge that the political setting poses for large-scale public policy experiments: the clashes between political and experimental timeframes.

How Political and Experimental Timeframes Clash

According to Schedler and Santiso (1998, p. 5), "[t]ime in its manifold manifestations represents a pervasive factor of political life." Studies on political time have distinguished between various temporal features (such as electoral terms, time budgets, and time horizons) and argued that all these aspects can profoundly affect policymaking (e.g., Goetz 2014; Goetz and Meyer-Sahling 2009; Howlett and Goetz 2014; Linz 1998; Schedler and Santiso 1998). Although in the existing literature on experimental policymaking, the potential clashes between political and experimental timeframes are mentioned, it has been done in a cursory fashion. In none of the existing studies have the tensions between experimental and political timeframes been examined in a systematic way—theoretically or empirically. Thus, to take the first step towards a more systematic understanding of these tensions, we developed the theoretical discussion in the following way. We juxtaposed the theoretical insights from various literatures on public policy, electoral politics, and mediatization of politics that discuss the nature of political timeframes with existing theoretical discussions on policy experimentation and derived the implications for experimental policymaking. We argue the temporalities of politics and experimentation can clash with each other in various ways. More specifically, the key sources of these clashes are disparities between the electoral and experimental timeframes and differences between the timeframes of political problem cycles and experiments.

Electoral Cycles Versus Policy Experiments

The modus operandi of electoral politics is that in a democratic setting, politicians strive to win competitive elections, and their decisions are influenced by their perceptions of which policies appeal to the electorate and maximize votes from their constituents (e.g., Bernecker, Boyer, and Gathmann 2021; Sørensen et al. 2020). In a democratic setting, "the most fundamental temporal unit is the electoral term" (Goetz and Meyer-Sahling 2009, p. 185). The electoral term provides politicians with the time budgets they have for getting things done and marks the time horizons they adopt—which in turn can influence the timing, starting, and ending of policies (Goetz 2014; Goetz and Meyer-Sahling 2009; Howlett and Goetz 2014; Linz 1998; Schedler and Santiso 1998). The impact of these temporal features on policy experiments can be viewed through various lenses. On the one hand, there are mechanisms through which the timeframes of electoral cycles can *facilitate* the use of policy experiments.

First, politicians may use experiments to extend their policy influence *beyond one electoral cycle*: to cast a longer shadow than their current term in office would otherwise allow and to tie the hands of subsequent governments in their policy choices (Callander and Hummer 2014; Corduneanu-Huci, Dorsch, and Maarek 2021). When politicians know that they may not be in office after the next elections, they may initiate policy experiments during their time in office in the hope that the evidence of positive results provided by trial would compel the future government to go ahead with their preferred policy direction. Thus, experiments initiated in one electoral cycle can serve as a mechanism for transmitting information to the next electoral cycle (to the electorate and to the next incumbents), for securing buy-in and ensuring the survival of the preferred policy (Callander and Hummel 2014; Corduneanu-Huci, Dorsch, and Maarek 2021). Used in that way, experiments can in fact lengthen the political time horizons, which may otherwise be confined to just one electoral cycle.

Second, politicians may launch policy experiments in the run-up to elections to avoid "larger pain" and to minimize electoral losses. As Nair and Howlett (2016) suggest, if addressing a particular problem entails painful and unpopular decisions, politicians may use an experiment as a "substitute" for large-scale policy action. Pilots can serve "as an excuse for policymakers to delay large-scale policy reforms beyond their term in office" (Nair and Howlett 2016, p. 71). In that way, the elected officials can demonstrate that they are "dealing" with a problem but at the same time avoid retribution from the electorate for politically unpalatable actions.

On the other hand, there are several ways how the electoral cycle can be expected to pose challenges for experimental policymaking because of the mismatches between experimental and electoral temporalities, especially when it comes to larger-scale policy experiments. The clashes between electoral and experimental timeframes can influence *whether* the experiments are initiated in the first place, how they are *designed*, and the extent to which policymakers *learn* from them.

First, given that electoral terms limit their time budgets and shorten their time horizons (Goetz and Meyer-Sahling 2009), politicians may be disinclined to launch large-scale experiments (spanning several years) in the first place (Bernecker, Boyer, and Gathmann 2021; Majumdar and Mukand 2004). If an experiment is initiated before elections but the results only materialize after, the political calculus may question the electoral benefit of such an undertaking (Bueno 2023; Majumdar and Mukand 2004; Mink and De Haan 2006). Prior to elections, politicians would be motivated to pinpoint concrete policy achievements and decisions that have already been adopted, highlight their visible benefits to the voters, and claim the credit (Bueno 2023; Harrington 1993; Rogoff 1990). If, instead, politicians have to point to an *ongoing* policy experiment with vague and uncertain future benefits (Corduneanu-Huci, Dorsch, and Maarek 2021), they may consider it to be an overly weak signal of their achievements in office, potentially undermining their chances in the electoral competition. Thus, the perceived short-term political benefits derived from faster policy decisions (by fitting them into the time budget remaining in the electoral

term) may outweigh the longer-term benefits of the information an experiment can offer. Such opportunistic uses of the time budgets shaped by the electoral cycle have been observed in other domains of political action and discussed extensively in the literature on electoral budget cycles (e.g., Bueno 2023; Katsimi and Sarantides 2012; Mink and De Haan 2006).

Second, even if politicians do decide to launch policy experiments, the considerations of the electoral term, which limits their time budgets (Goetz and Meyer-Sahling 2009), may motivate politicians to push the experiments to be done faster than warranted by scientific design (Corduneanu-Huci, Dorsch and Maarek 2021; Stoker and Evans 2016; Strydom et al. 2010). Thus, the elected officials may demand shorter experiments to ensure that the results can be shown within their time budget: before the next elections (Nair and Howlett 2016; Stoker 2010). Although large-scale strategic RCTs lasting several years hold the promise of delivering the most rigorous causal conclusions about the effects of a tested policy (e.g., Cotterill and Richardson 2010), politicians may object to longer timeframes needed for the intervention to take effect and demand shorter timeframes for the experiment. Due to the limited duration, however, the experiments may not be able to capture the longer-term impacts of the tested policy solution (Millo and Lezaun 2006).

Third, even if politicians agree to launch experiments that cross the electoral cycle (i.e., are designed to end after the subsequent elections), the proximity of upcoming elections may motivate politicians to draw conclusions from the ongoing experiment too hastily (Checkland et al. 2023; Martin and Sanderson 1999; Rogers-Dillon 2004). Stoker (2010, p. 53) argues even "if the right moment is seized to do experimental work, there can be issues about the impatience of policymakers in waiting for results." If politicians support the policy solution the experiment is testing but fear that they would not be in office after the next elections, they may wish to cement certain policy decisions during their time in office. As a result, they may be inclined to scale up the tested policy immaturely, before the experiment has an opportunity to provide solid evidence. This, in turn, would undermine learning from the experiment and important policy decisions may be made before the results of the experiment come in (Nair and Howlett 2016; Stoker 2010).

Fourth, electoral cycles may induce politicians to ignore the results of the experiments (Corduneanu-Huci, Dorsch, and Maarek 2021; Majumdar and Mukand 2004). By the time the results of an experiment are available, a new government may be in office and, consequently, political priorities may have changed (Bravo-Biosca 2020; p. 205; Nair and Howlett 2016; p. 70; Oakley 1998; Pearce and Raman 2014, p. 394). Thus, experiments may become regarded "as time capsules from a previous era" (Huitema et al. 2018, p. 148) and the new government in office after elections may be inclined to ignore the experimental evidence.

Political Problem Cycles Versus Policy Experiments

Alongside the considerations of the electoral cycle, politicians' decision making is strongly influenced by emergent problems in the public sphere and the collective attention to these problems (e.g., Jones 2017). The essence of what we label as a "problem cycle" is the following: when there is a societally perceived problem and politicians are expected to "fix" it—sooner, rather than later. The timeframes of the "problem cycles" are likely to affect politicians' inclination to launch and learn from experiments.

If delays in solving societal issues are perceived as problematic, the time horizons of policymaking can become even shorter than the timeframes posed by electoral cycles. If politicians perceive a strong time pressure in addressing a societal problem, gathering evidence via experimenting may seem as a "luxury," as pointed out by the literature on evidence-based policymaking (e.g., Cairney 2016; Stoker and Evans 2016; Strydom et al. 2010). Some problems may be perceived to be so acute that they do not allow time for conducting experiments (Clarke and Craft 2019; Pearce and Raman 2014). Instead, driven by ideology, politicians may prefer to trust their gut feeling about the potential impacts of different policies—and hence would not feel the need to demand more rigorous evidence via experiments (Beesley, Hawkins, and Moffitt 2022; Bravo-Biosca 2020; Cairney 2016). Experimenting strikes the kind of "chord of skepticism and indecision" that politicians seek to avoid (Peters 1998, p. 126), especially in a situation of collective attention to a problem that has been defined as urgent (Jones 2017). Even if an experiment is launched to test different solutions to an imminent and salient problem, then similarly to the dynamics in electoral cycles, politicians may want to hasten the evaluation of the data provided by the experiment to get answers more quickly (Nair and Howlett 2016; Trein and Vagionakis 2022).

The clashes between the experimental and political problem-solving timeframes are likely to be amplified by two factors: (1) media attention to societal problems and (2) crises. Media attention is likely to increase policymakers' perception of urgency in solving salient societal problems (Stoker and Evans 2016; Torfing and Ansell 2017). The increasing role of media in shaping political decisions has been captured by the term "mediatization of politics." Since the media plays a strong role in influencing public opinion, politicians are strongly motivated to take it into consideration in their decisions (Strömbäck and Esser 2014). The growing mediatization of politics means that the media attention amplifies the societal gaze under which political actors move and it forces political actors to take decisions more quickly (Blumler 2014; Mazzoleni and Schulz 1999; Strömbäck 2008; Strömbäck and Esser 2014; Torfing and Ansell 2017). Due to the perceived power and influence of the (news) media, political actors may subject the pace of their decisions to the media-cycle. As Blumler (2014, p. 35) puts it, the rhythm of news production embraces "the new" and "all involved in its making are drenched in the fluid immediacy of events and their coverage." Consequently, "politicians and their advisors often seem impelled to keep up with and respond to the news on its terms and in its time" (p. 35). As a result of the scandal-focused media, while at the same time suffering from information-overload and shortage of knowledge (e.g. ,Jones 2017), politicians may be tempted to dive into solutions too quickly (Torfing and Ansell 2017, p. 38), which leaves no time for discussing alternative options and testing them before scaling up.

The shortening of political time horizons resulting from mediatization would be further amplified if the policy problem is defined as a "crisis": where the core values, institutions, or functions of the society are perceived to be under threat (Boin and t'Hart 2022). A situation of a (perceived) crisis is likely to give rise to a sense of "policy urgency" among

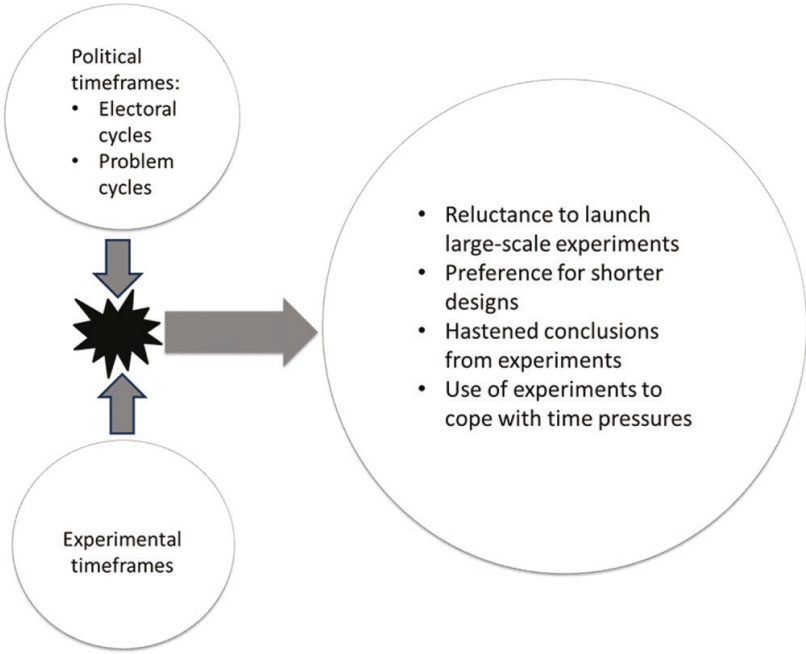


Figure 1. Impacts of clashes between political and experimental timeframes

politicians (Hulme and Hulme 2012; Lesch and Millar 2022). In such a situation, politicians are likely to forgo experimental policymaking, since such evidence-collection is likely to take more time than politicians have in their time budget (Hulme and Hulme 2012).

In sum, we expect the clashes between political and experimental timeframes to lead to the following impacts (see also fig. 1). First, both electoral and problem cycles are likely to hinder the launch of large-scale policy experiments. Second, even if large-scale policy experiments do get launched, electoral cycles can lead politicians to prefer shorter designs and draw conclusions before the completion of the experiments. Finally, politicians may use policy experiments to cope with pressures from electoral and problem cycles. Next, we test these conjectures empirically.

METHODS

To probe the plausibility of the theoretical expectations outlined the previous section, we draw on interviews carried out in Estonia and Finland from Fall 2022 to Summer 2023. The semi-structured interviews (32 in Estonia and 34 in Finland) were conducted according to the same interview protocol and engaged central government public officials from the Government Office/Prime Minister's Office, fiscal and financial policy institutions, and other public institutions. When selecting interviewees, we were guided by the following considerations. First, we contacted officials who are or have been in charge of coordinating experimental policymaking in the Prime Minister's Office in Finland and the Government Office in Estonia. Second, we wanted to interview officials from different hierarchical positions. Our set of interviewees included: 6 top-level officials from Estonia and 4 from Finland; 13 middle-level officials in Estonia and 12 in Finland; 13 expert-level officials in Estonia and 18 in Finland. Third, we wanted to ensure that our interviews cover the perspectives of officials working in ministries and agencies. Our set of interviewees thus included 32 ministry-level officials (16 in each) and 16 agency-level officials (7 in Finland and 9 in Estonia).[2] Finally, we used snowballing to identify further interviewees who had had experiences with policy experiments (see Appendix I for more details).

Despite being geographically close, the two countries represent different cases in terms of their historical-social background and experience with policy experimentation. Finland is regarded as a leading country of policy experimentation, embedding them in "anticipatory governance," while Estonia is an aspiring novice, having only taken its first steps towards a more systematic public policy experimentation. The interviews explored the perceptions of the Estonian and Finnish public officials regarding the benefits and pitfalls of policy experiments as well as their own experience with experimenting (see Supplementary Appendix II for the interview protocol). The average length of an interview was one hour.

All the interviews were recorded and transcribed. In coding the interviews with MAXQDA, we used a combination of deductive and inductive approaches. The coding exercise started with the initial coding scheme based on the preliminary analytical framework drawing on the international academic

[2]Given the overall focus of the broader research project within which the interviews were conducted and to ensure in-depth comparability between the cases, the interviewees included 11 Ministry of Finance officials in Finland and 15 in Estonia; 7 officials from the Tax and Customs Board in Estonia and 4 officials from the Finnish Tax Administration.

debate on policy experimentation. In the process, the initial coding scheme was complemented with the additional codes emerging from the analysis of the interview transcripts. All the statements of the interviewees concerning the role and perceptions of politicians and political institutions towards policy experimentation as well as all the statements regarding temporal aspects of experimenting were coded with respective codes (see Supplementary Appendix II for more details).

The research design allowed us to gain rich information about the perceptions and experiences of public officials regarding the different temporalities involved in policymaking and experimentation. To maintain the anonymity of the interviewees, they are identified by a short code in the analysis (e.g., Est1, Fin1).

EMPIRICAL ANALYSIS

A Short Overview of the Case Contexts

Estonia and Finland—neighboring countries located at the opposing shores of the Baltic Sea—are both small parliamentary democracies with parliaments *Eduskunta* (FIN) and *Riigikogu* (EST) elected by proportional representation. The countries have strong multi-party systems that, in combination with the proportional representation, lead to coalition governments where common interests and aims must be sought by political parties.

The development of the experimental culture in Finland dates back several decades as a central theme in innovation policy and the activities of the Finnish innovation fund SITRA but was strongly re-emphasized during the government period of Prime Minister (PM) Juha Sipilä in 2015–2019. During that period, policy experimentation was institutionalized as a high political priority through the program "Experimental Finland" (2016–2019) (Leino and Akerman 2022). A framework of public sector experimentation was developed and funding for experiments was allocated. The program designed by the Experimental Finland Team in the PM Office foresaw several types of experiments including strategic experiments (policy trials) and pilot pools/partnerships (regionally relevant or sector-specific experiments) (Leino and Akerman 2022, p. 45). Alongside hundreds of experiments and policy pilots conducted across the country both at the central government and municipal level, six strategic experiments were initiated, including the world-famous universal basic income (UBI) experiment (Kangas et al. 2021). The subsequent government of PM Sanna Marin launched various large-scale policy experiments as well, including an RCT testing the extension of preschool education to 5 year olds, an RCT testing the effects of a recruitment subsidy for sole entrepreneurs, and piloting the delegation of employment services to municipalities, among others.

In Estonia, the emergence of policy experimentation as an element of public sector innovation policy has been much more recent and characterized by a bottom-up approach. Experimentation has been promoted by public sector innovation programs financed from the European Union structural funds and so-called "design sprints" led by the Public Sector Innovation Team at the Government Office. In June 2023, a methodological guide for experimenting in the public sector was published. However, the policy experiments carried out have been mostly small-scale design experiments (e.g., on improving the explanatory note of the state budget), policy pilots (e.g., on performance budgeting), or RCTs (e.g., concerning recycling of packaging) not demanding political attention or intervention.

Electoral Cycle Versus Experimental Cycles

Our interviews indicate that in both Finland and Estonia, the electoral cycles have influenced experimental policymaking in important ways. The clashes between electoral and experimental timeframes have influenced decisions on *whether* and *when* to initiate experiments, the temporal features of their *design*, and *learning* from experiments.

Whether, When, and How Long to Experiment?

Even though the governments in Finland have managed to launch large-scale experiments spanning several years, many interviewees did consider the electoral cycle to be a major obstacle to policy experiments. It was noted that since governments are in office only for the term of four years it may *discourage* politicians from launching longer-lasting experiments since they feel they do not have enough time in their time budget to wait for the results. As an interviewee explained, "politicians know that they have only four years in office and if they want to get votes in the next elections, they have to signal to the voters that they have done something beneficial. That creates a problem with these experiments because it would be ideal to have them running for longer time-periods. But the politicians don't want to, because the effects would be coming after the elections, and they may not be in the next government" (Interview Fin20).

Estonian public organizations have undertaken several smaller-scale experiments that have remained under the political radar—for those, the political timeframes have not presented significant challenges. The interviewees acknowledged that for large-scale policy experiments (the Finnish UBI experiment was often referred to as an example), politicians would need to be involved to initiate the trial, provide the legislative mandate, and secure funding (Interviews Est7, Est11, Est12, Est15, and Est18). Politicians' focus on the electoral cycle, however, was viewed as a major hurdle in undertaking large-scale policy experiments. Furthermore, some of the interviewees pointed out that due to the frequent turnover of governments in Estonia (with each coalition lasting 1–2 years), the time horizons of politicians are even shorter than the electoral cycle of four years, with one respondent stating "the political cycle in Estonia is about a year—so you cannot actually assume that you have 4 years to carry out any project" (Interview Est9). That, in turn, has hindered the initiation of any longer-term experiments: "politicians are afraid that by the time the experiment is completed, they might not be in the office anymore, so they want to implement the proposed solution right away" (Interview Est18).

The electoral timeframes have implications also for the *timing* of new policy experiments. The Finnish experience with large-scale RCTs demonstrates that if politicians do want to launch them, they strongly prefer to do it at the *beginning* of the governmental term, to ensure that the experiments get completed within their time budget. An interviewee summarized the tendency as follows: "it's very difficult for politicians to accept that they have launched a certain experiment but not get the outcome during the same electoral period. This is one reason why experiments are often launched at the beginning of the governmental term." (Interview Fin12)

The Finnish interviews indicate that the Government Program is a key instrument for outlining which experiments the government plans to undertake but the composition of the Program is influenced by political timeframes. A Government Program is a 4-year action plan that a new coalition government agrees on at the beginning of the governmental term. If the experiments are decided during the negotiations over the Government Program and written into it, then they cannot be neglected later (Interviews Fin3, Fin12, Fin15, Fin16, Fin 17, Fin22, Fin23, Fin24, and Fin25). The time budget for compiling the Government Program, however, is usually quite short (from a couple of weeks to a couple of months) and that means that only limited time is available for considering what kinds of strategic experiments could be mentioned in the Program.

When politicians in Finland have decided to launch larger-scale experiments spanning several years, the clashes between the political and experimental timeframes have influenced the *design* of the experiments. While researchers and civil servants would have preferred longer timeframes for preparing the experiments and their actual duration, the political timeframes were much shorter due to the electoral cycle (Interviews Fin7, Fin8, Fin10, Fin11, Fin14, Fin15, Fin16, Fin20, Fin28, and Fin32). As observed by one of the interviewees, in the case of the UBI experiment, there was strong pressure from politicians to launch the experiment quickly: "the politicians wanted to have the experiment up and running in one month. But there were a lot of things that needed to be done before, so that was very unrealistic." (Interview Fin6) In the words of another interviewee: "the toughest part was the timeframe. The universal basic income experiment had to be conducted quite quickly because the politicians wanted to complete it before the next elections. It is difficult for them to understand how long it takes to prepare and implement a good experiment." (Interview Fin8) This provides evidence for the theoretical expectation that politicians' considerations of the electoral cycle can lead them to prefer shorter timeframes for policy experiments.

However, our interviews also indicate that there can be *political* reasons for why the launch of an RCT can take a long time. In the Finnish context, RCTs that treat people differently need a legislative basis. Passing the necessary legislation can take time and considerably delay the start of the experiment (Interviews Fin7, Fin14, and Fin21). Furthermore, in some instances, electoral considerations can lengthen (rather than shorten) the duration of the experiment. In the case of the pre-primary education trial, the politicians demanded the addition of one more cohort of pupils to the experiment, which extended its length by one year.[3] Although that meant that the trial could not be concluded anymore within the Government period, the interviewees suggested that the request to include an additional cohort was related to the political calculation: "they just thought that it would look nice for the people who are voting for them, the electorates." (Interview Fin13). Thus, contrary to our expectation that electoral cycles would shorten the timeframes of policy experiments, the Finnish case provides examples of situations where political considerations actually lengthen the experimental timeframes. In sum, our cases indicate that electoral cycles play a significant role in whether and when politicians prefer to launch policy experiments and which experimental timeframes they prefer. Next, we analyze how electoral cycles affect the learning from experiments.

[3]The RCT sought to assess the effects of offering pre-primary education to 5 year olds on children's development and learning (Ministry of Education and Culture 2021).

Learning from Experiments

We conjectured that, in addition to decisions on how long the experiment should last, the electoral cycle may influence *how quickly* the politicians draw lessons from an on-going experiment or whether they are willing to wait for them at all. As the Finnish interviews indicate, this has indeed happened in the case of a large-scale policy pilot that was intended to test whether delegating employment services from the central government to the municipalities would improve employment outcomes. The pilot was launched in early 2021 and was meant to last for 4 years (until the end of 2024). However, before the 2023 spring elections—2 years before the experimental evidence was supposed to come in—the Finnish government decided to scale up the policy solution being tested and delegate employment services to all municipalities, starting from 2025 (Interviews Fin9, Fin12, Fin13, and Fin15). As one of the interviewees explained, "the pilot wasn't given enough time to show any results before the politicians took the decision for this large-scale structural change. The first results only came in after the government had already decided to scale this policy up. The preliminary results showed that there was no significant difference between the pilot municipalities and the synthetic control groups, but by then the policy decision was already done." (Interview Fin9) This shows how pressures of the electoral cycle led politicians to become impatient waiting for the results of the policy experiment and decide to expand the reform before evidence about its impacts was available.

As the interviewees explained, the government wanted to use the opportunity when in office to get something done in the policy domain it considered ideologically and politically important. Since the coalition politicians could not be sure they would be in the next government, they wanted to leave a mark on the employment policy area (Interviews Fin9, Fin12, and Fin15). Similarly, in the case of the pre-primary education trial, the approaching general elections put the issue of pre-primary education on the political agenda and led to a political suggestion of scaling up the solution to all the children before any results from the trial had come in (Interview Fin16).

We suggested that the electoral cycles may lead to the disregarding of experimental evidence when a new government comes to office and has different policy priorities from the previous one. We witnessed some of these dynamics in the case of the UBI experiment in Finland, where the results of the experiment only came in after the term of the government (that had initiated it) had ended. Although the experiment did show some promising results regarding improving the well-being indicators of participants, the new government that came into office after elections had other policy priorities and did not pursue any further steps with universal basic income (Interviews Fin6, Fin20, Fin23, and Fin27). One of the interviewees explained that there is a higher chance for the results of the experiment to "endure" beyond the term of the government who initiated it and to influence the new government in office when the new coalition includes at least one party from the previous coalition (Interview Fin20). It was also suggested that if the issue being experimented with is considered important by all parties, the subsequent coalition

governments are more likely to consider its results (Interviews Fin9 and Fin14).

Another factor influencing whether the legacy of an experiment endures from one government to the other is whether the parliamentary opposition parties have been included in the discussions on the experiment and support it (Interviews Fin14 and Fin21). That way, the opposition parties are less likely to discard the experiment when they take office. Thus, involving opposition parties in decisions regarding an experiment can lengthen politicians' time horizons beyond the electoral cycle. Broad-based discussions and the building of consensus concerning an experiment can hence provide opportunities for longer experimental timeframes, which in turn improves the validity of the evidence experiments offer. In sum, our cases show that electoral cycles play a significant role in whether experiments are launched, how long they last, and how politicians draw conclusions from them. Next, we turn to the impacts of the problem cycles on policy experimentation.

Problem Cycle Versus Experimental Cycle

We conjectured that the short problem-cycles politicians follow can hamper the use of large-scale experiments in experimental policymaking, and the sense of policy urgency is likely to be amplified by mediatization and crises. The interviews from both Finland and Estonia corroborate these expectations.

Although the Finnish government has—to some extent—been able to overcome the short-termism pressures imposed by the problem-cycle 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 interviews put it, "during the covid crisis, nobody in their right mind could suggest experimenting with the solutions before scaling them up. We had to act immediately" (Interview Fin34). Thus, we can see that even in a country where politicians favor policy experimentation, the time pressures of mediatization and crises can hamper their willingness to experiment.

Our interviews suggest that policy experiments in a specific policy area may also be influenced by problem pressures in *neighboring* policy fields. In the case of the policy pilot that delegated employment services to the municipalities, the context was changed by a major reform of social and health services whereby the responsibility and budget for these services were moved from the municipal to the regional level, with one interviewee stating "at that stage, the politicians thought that we had to compensate for this change somehow to the municipalities and decided to give employment services from the state to the municipalities" (Interview Fin12). This consideration contributed to the decision to scale up the employment services reform before the policy pilot could deliver any results (Interviews Fin12 and Fin13).

On the other hand, the Finnish interviews also demonstrate how politicians use pilots and experiments to "buy time" in solving problems and to mitigate time pressures created by mediatized problem cycles. Several interviewees argued that since launching pilots requires less time (and also less money and effort) than a full-scale policy change, politicians often prefer it as a first step in addressing the problem (Interviews Fin2, Fin4, Fin8, Fin16, and Fin20), with one interviewee stating "launching pilots allows the politicians to go on TV and make an announcement about that" (Interview Fin2). It also emerged from the Finnish interviews that, at times, politicians use experiments to reach a compromise in a situation where reaching a "permanent" policy solution may be too difficult (e.g., Interviews Fin4, Fin8, Fin16, and Fin20), with one interviewee stating "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 the 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 and Fin12). Thus, our interviews suggest that initiating pilots is one way for politicians to postpone decision making. As one of the Finnish officials put it, "if politicians are not able to decide upon some policy, it's easy to say: 'Let's experiment with it'" (Interview Fin8).

The limitations imposed by problem cycles on large-scale policy experiments were perceived to be significant in the Estonian context. As the Estonian interviews indicate, delays in solving societal issues are often perceived as problematic, which, in turn, constrains the use of large-scale policy experiments (e.g., Interviews Est1, Est15, Est17, and Est27). In the words of the interviewees: "politicians feel that they have to put out the fire immediately. They live one day at a time. This makes it challenging to introduce policy experiments. …. They want to put the seed into the ground today and take out the carrot tomorrow" (Interview Est17). Another argued "we have to decide on the solution and then start running. Even then we may not be fast enough. … If we started to experiment before choosing direction, we would lose even more time" (Interview Est1). The Estonian interviews also confirm that problem cycles faced by politicians can be further shortened by media cycles and perceived crises, which amplify the time pressure in solving problems. As one of the interviewees explained, "when something happens, for example, the electricity price goes through the roof, the minister has to be on TV in 15 minutes and promise solutions. And then they go there, blurt out some solution that may not be reasonable actually and then we somehow have to make it work." (Interview Est4) This illustrates well how problem cycles—especially when amplified by mediatization and crises—impose pressures on politicians to come up with quick policy solutions, leaving no time to test them experimentally before scaling up.

In sum, both cases demonstrate how clashes between problem cycles and experimental timeframes can pose challenges for launching large-scale policy experiments. While in Finland, politicians have found ways to overcome the time pressures created by problem cycles and launch a number of large-scale experiments, in Estonia, the constrained time budgets created by problem cycles have been a major hurdle for experimental policymaking.

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 (e.g., Huitema et al. 2018; McGann et al 2018; Rogers-Dillon 2004), we lack systematic knowledge on how various political dimensions affect experimental policymaking and interact with the preconditions of creating reliable experimental evidence.

In this article, we sought to address a specific gap in the existing understanding of the politics of experimentation: how political timeframes influence experimental policymaking. As the growing literature on political timeframes has suggested, temporal features like electoral terms, time horizons, and time budgets are likely to influence policymaking (Goetz 2014; Goetz and Meyer-Sahling 2009; Howlett and Goetz 2014; Linz 1998; Schedler and Santiso 1998). In our study, we sought to shed light on the challenges that various political temporalities—the electoral cycles and problem cycles—pose for large-scale policy experiments. Drawing on theoretical discussions on experimental policymaking, public policy, electoral politics, and mediatization of politics, we synthesized expectations about how electoral and problem cycles may influence the timing, design, and learning from policy experiments. We then probed the plausibility of the theoretical conjectures using interview data from Estonia and Finland. Table 1 summarizes our findings.

First, we conjectured in the theoretical discussion that since electoral timeframes limit the time budgets and shorten the time horizons of politicians (Goetz 2014; Goetz and Meyer-Sahling 2009; Howlett and Goetz 2014; Linz 1998; Schedler and Santiso 19988), electoral cycles are likely to discourage politicians from undertaking large-scale policy experiments in the first place. Our interviews show that the electoral cycle has indeed constituted a major hurdle to conducting large-scale policy experiments in Estonia. In Finland, however, politicians have launched several large-scale policy experiments during the past decade despite the potential obstacles from the electoral cycle. This indicates that the electoral cycle does not always hinder large-scale policy experiments, and it would be fruitful to investigate further under which conditions this happens.

Second, drawing on studies on political time (Goetz 2014; Goetz and Meyer-Sahling 2009) and experimental policymaking (Nair and Howlett 2016; Stoker 2010), we suggested that even if politicians do decide to launch large-scale experiments, they may prefer shorter designs than warranted by scientific considerations. The Finnish experience with large-scale policy experiments—whereby politicians have pressured the implementers of the experiments to design shorter experiments—offers some evidence of such a tendency. At the same time, our interviews pointed to an example where politicians actually preferred to extend the duration of the experiment since it was electorally expedient.

Third, we expected that the electoral cycle may lead politicians to draw hasty conclusions or ignore the experiment's results altogether (e.g., Checkland et al. 2023; Rogers-Dillon 2004; Stoker 2010). The experience with several large-scale policy experiments in Finland provides evidence of such a propensity. Fourth, we conjectured that problem cycles are likely to shorten politicians' time horizons further as there is pressure to solve problems quickly (e.g., Clarke and Craft; Pearce and Raman 2014), making them reluctant to undertake policy experiments and to prefer immediate policy action instead. We expected the impact of such problem cycles to be amplified by media cycles (e.g., Blumler 2014; Strömbäck 2008; Strömbäck and Esser 2014) and the perceptions of crisis (Hulme and Hulme 2012; Lesch and Millar 2022). Our interviews indicate that the compressed timeframes imposed by the problem cycles and amplified by media attention and crises have indeed acted as a major constraint on launching large-scale experiments in Estonia. In Finland, however, even though these pressures are perceived to be present, politicians have launched numerous large-scale experiments.

Fifth, our theoretical discussion pointed to how politicians may be inclined to use large-scale experiments to cope with the political timeframes: either to tie the hands of the next government (Callander and Hummer 2014; Corduneanu-Huci, Dorsch and Maarek 2021) or to avoid taking painful decisions (Nair and Howlett 2016). While we did not find evidence of the former, the Finnish interviews did indeed demonstrate the use of experiments as a way of postponing or avoiding larger decisions in the face of time pressures imposed by both electoral *and* problem cycles. Furthermore, as the Finnish experience indicates, the temporary nature

Table 1. Summary of empirical findings.

	Finland	Estonia
Conjecture 1: Electoral cycles prevent the launch of large-scale experiments.	Although the electoral cycle is seen as a potential obstacle, several large-scale experiments have been launched	Confirmed
Conjecture 2: Electoral cycles shorten the duration of large-scale experiments.	Evidence of pressures towards shorter design, but also instances of lengthening the duration	No large-scale experiments undertaken
Conjecture 3: Electoral cycles hasten learning from experiments and can lead to ignoring experimental evidence	Confirmed	No large-scale experiments undertaken
Conjecture 4: Problem cycles prevent the launch of large-scale experiments.	Not confirmed	Confirmed
Conjecture 5: Politicians use large-scale experiments to cope with time pressures from electoral and problem cycles.	Confirmed	Not confirmed
Conjecture 6: Politicians use large-scale experiments to tie the hands of next governments.	Not confirmed	Not confirmed

of experiments may make it possible for the coalition partners to reach a joint decision at all, during their government term. Also, in situations where due to a mediatized environment, politicians are expected to offer "quick" solutions, the politicians can seek to "win time" by saying that they are launching a pilot as a response to the problem. We did not, however, witness such uses of experiments in Estonia.

In sum, even though our interviews did not fully corroborate all the conjectures, our theoretical framework offered useful lenses for uncovering important dynamics in experimental policymaking and could be used as a starting point in future theorizing and empirical studies. Furthermore, theorizing could focus on the conditions in which our expectations are likely to hold. In addition to providing evidence of the relevance of the electoral and problem cycles, our interviews also revealed *additional* political timeframes that can play a significant role in experimental policymaking. First, the Estonian case pointed to the potential importance of timeframes imposed by *government turnovers* in the middle of the electoral cycles. As the average duration of cabinets in Estonia is rather short (1–2 years), this appears to amplify the effects that the electoral cycle has on experimental policymaking. The shortened time horizons, due to frequent turnover of governments, are likely to act as a further deterrent to launching longer-lasting strategic policy experiments. Thus, when we are looking for reasons why the Finnish government has been able to launch large-scale strategic policy experiments spanning several years while Estonia has not, then alongside the awareness and knowledge about policy experiments (which may be lagging behind in Estonia), the swift turnover of governments may be acting as an additional obstacle to undertaking such long-lasting experiments. The more stable time horizons available to Finnish politicians (the expectation to stay in the office for the whole 4-year period) can be viewed as more conducive to large-scale experiments than the shorter time horizons Estonian politicians face in the context of frequent change of governments. In Finland, governments are better able to commit to strategic experiments owing to their longer time horizons.

Second, our cases point to the importance of the Government Program as an instrument for managing the constrained time resource in the context of coalition governments. In both countries, the Government Program is the key document for subsequent policy actions and serves as an important document for politicians in committing to certain experiments. As such, it structures both the choices of politicians as well as other stakeholders—through the period of time dedicated to the negotiations of the Government Program, the need to have the key policy goals formulated *before* the negotiations, and the *timing* of preparations for launching the experiments right after the adoption of the Program.

Third, as the Finnish case revealed, another additional timeframe that can be relevant for experimental policymaking is that of passing necessary legislation for the experiment. Large-scale RCTs which foresee differential treatment of individuals need a legislative basis. The passing of the necessary law(s), however, can considerably lengthen the preparation phase of the policy experiment. At the same time, parliamentary proceedings also build wider consensus around the experiments and through such legitimation increase their chance of survival beyond the ongoing electoral cycle.

In sum, our study shows that to understand experimental policymaking in democratic settings, it is crucial to pay attention to the potential clashes between political and experimental timeframes. In addition to the limitations posed by the electoral and problem cycles, experimental policymaking can be influenced by timeframes imposed by the frequency of government turnover, the limited time available for compiling coalition agreements, and the procedural requirements for passing legislation underpinning the experiments. Thus, our study shows that there appears to be a deep-rooted tension between the timeframes followed by politicians and the timeframes required by policy experiments. Given that in a democratic setting, politicians' temporalities cannot be subjected to the demands of policy experiments, this tension may, in fact, be *unresolvable*. This, in turn, has implications for the expectations we can have about experimental policymaking.

In addition to rendering the expectations more realistic, more scholarship is needed to uncover the factors that alleviate the tensions between political and experimental timeframes. For example, our cases suggest that conducting experiments in policy areas that are considered important by all parties, involving opposition parties in the discussion and design of experiments, and stipulating experiments in laws may lengthen the political time horizons. However, more research is needed on how different types of experimental designs, processes for preparing and conducting experiments, and the range of stakeholders involved in experimental policymaking influence the compatibility of political and experimental timeframes. Furthermore, it would be insightful to investigate whether the tensions between timeframes vary across policy areas and are influenced by the salience and complexity of different policy solutions being tested. If a policy issue is highly salient to a powerful interest group, the interactions between the political and experimental timeframes are likely to be different from when the tested policy has low salience and is not on the radar of strong interest groups.[4] For example, a strong interest group of beneficiaries may pressure politicians to undertake a shorter experiment in order to benefit from the scaled-up policy sooner. Conversely, experiments with low salience and diffused impacts may allow policymakers to consider longer experimental designs.

In our study, we have focused on how the timeframes of elected officials influence large-scale experiments. As existing research on various temporalities in policymaking has shown, however, different actors in the political system may have different time limits, time horizons, and time budgets (Goetz and Meyer-Sahling 2009), and these are likely to have implications for experimental policymaking as well. For example, unelected civil servants may have longer time horizons (Goetz 2014; Goetz and Meyer-Sahling 2009) and could hence have incentives to launch longer-lasting experiments, avoid premature interpretation of experimental evidence, or minimize the involvement of politicians. Future studies could analyze how the different timeframes of elected and unelected officials (and their interactions) influence experimental policymaking.

An important limitation of our study is that we probed the plausibility of our theoretical expectations in two parliamentary countries that have proportional electoral systems resulting in multi-party coalition governments. In future research, it

[4]We would like to thank the anonymous reviewer for drawing our attention to this issue.

would be fruitful to investigate how accurately our theoretical propositions hold in countries that are presidential, have majoritarian electoral systems, and/or have one-party governments.

In sum, our article confirms that politics can profoundly affect policy experiments. In order to better understand the politics of experimental policymaking, it is crucial to study how political and experimental timeframes clash and how these clashes influence whether, when, and how politicians experiment. We demonstrate that electoral cycles and problem cycles play an important role in politicians' motivations regarding policy experiments. There are, of course, additional aspects of the political setting that can influence politicians' preferences and decisions regarding policy experiments (e.g., ideological preferences, budgetary constraints, and blame avoidance). In order to advance our understanding of experimental policymaking it is important to address them in future research.

Supplementary material

Supplementary material is available at the *Journal of Public Administration Research and Theory* online (www.jpart.oxfordjournals.org).

Acknowledgments

This work was supported by the Estonian Research Council grant PRG1125. The authors are grateful for the helpful comments of the participants of the EGPA conference in Zagreb, 5–8 September 2023.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

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

References

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

Bédécarrats, F., I. Guérin, and F. Roubaud. 2019. "All That Glitters is Not Gold. The Political Economy of Randomized Evaluations in Development." *Development and Change* 50 (3): 735–62.

Beesley, C., D. Hawkins, and N. Moffitt. 2022. "What Leads Government Officials to Use Impact Evidence?" *Journal of Public Policy* 42 (1): 20–42.

Bernecker, A., P. C. Boyer, and C. Gathmann. 2021. "The Role of Electoral Incentives for Policy Innovation: Evidence from the US Welfare Reform." *American Economic Journal: Economic Policy* 13 (2): 26–57.

Blumler, J.G. 2014. "Mediatization and Democracy." In *Mediatization of Politics: Understanding the Transformation of Western Democracies*, edited by F. Esser and J. Strömbäck, 31–41. London: Palgrave Macmillan UK.

Boin, A., and P.'t Hart. 2022. "From Crisis to Reform. Exploring Three Post-COVID Pathways." *Policy and Society* 41 (1): 13–24.

Bravo-Biosca, A. 2020. "Experimental Innovation Policy." *Innovation Policy and the Economy* 20: 191–232.

Bueno, N. S. 2023. "The Timing Of Public Policies: Political Budget Cycles And Credit Claiming." *American Journal of Political Science* 67 (4): 996–1011.

Cairney, P. 2016. *The Politics of Evidence-Based Policy Making*. London: Springer.

Callander, S., and P. Hummel. 2014. "Preemptive Policy Experimentation." *Econometrica* 82 (4): 1509–28.

Checkland, K., J. Hammond, A. Coleman, J. Macinnes, R. Mikelyte, S. Croke, J. Billings, et al. 2023. "'Success' in Policy Piloting: Process, Programs, And Politics." *Public Administration* 101 (2): 463–80.

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

Cobb, P., J. Confrey, A. diSessa, R. Lehrer, and L. Schauble. 2003. "Design Experiments in Education Research." *Educational Researcher* 32 (1): 9–13.

Corduneanu-Huci, C., M. T. Dorsch, and P. Maarek. 2021. "The Politics of Experimentation: Political Competition and Randomized Controlled Trials." *Journal of Comparative Economics* 49 (1): 1–21.

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): 148–64.

Ettelt, S., N. Mays, and P. Allen. 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.

Goetz, K., and J. H. Meyer-Sahling. 2009. "Political Time in the EU: Dimensions, Perspectives, Theories." *Journal of European Public Policy* 16 (2): 180–201.

Goetz, K. H. 2014. "Time and Power in the European Commission." *International Review of Administrative Sciences* 80 (3): 577–96.

Harrington, J. E. 1993. "Economic Policy, Economic Performance, and Elections." *American Economic Review* 83 (1): 27–42.

Howlett, M., and K. H. Goetz. 2014. "Introduction: Time, Temporality and Timescapes in Administration and Policy." *International Review of Administrative Sciences* 80 (3): 477–92.

Huitema, D., A. Jordan, S. Munaretto, and M. Hildén. 2018. "Policy Experiments: Core Concepts, Political Dynamics and Impacts." *Policy Sciences* 51 (2): 143–59.

Hulme, R., and M. Hulme. 2012. "Policy Learning? Crisis, Evidence and Reinvention in the Making of Public Policy." *Policy & Politics* 40 (4): 473–89.

Jones, B. D. 2017. "Behavioral Rationality as a Foundation for Public Policy Studies." *Cognitive Systems Research* 43: 63–75.

Kangas, O., S. Jauhiainen, M. Simanainen and M. Ylikännö, eds. 2021. *Experimenting with Unconditional Basic Income: Lessons from the Finnish BI Experiment 2017–2018*. Cheltenham: Edward Elgar.

Katsimi, M., and V. Sarantides. 2012. "Do Elections Affect the Composition of Fiscal Policy in Developed, Established Democracies?" *Public Choice* 151 (1): 325–62.

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): 281–97.

Leino, H., and M. Akerman. 2022. "The Politics of Making Finland an Experimenting Nation." *Critical Policy Studies* 16 (4): 441–59.

Lesch, M., and H. Millar. 2022. "Crisis, Uncertainty and Urgency: Processes of Learning and Emulation in Tax Policy Making." *West European Politics* 45 (4): 930–52.

Linz, J. J. 1998. "Democracy's Time Constraints." *International Political Science Review* 19 (1): 19–37.

Majumdar, S., and W. S. Mukand. 2004. "Policy Gambles." *American Economic Review* 94 (4): 1207–22.

Martin, S., and I. Sanderson. 1999. "Evaluating Public Policy Experiments: Measuring Outcomes, Monitoring Processes or Managing Pilots?" *Evaluation* 5 (3): 245–58.

Mazzoleni, G., and W. Schulz. 1999. "Mediatization of Politics: A Challenge for Democracy?" *Political Communication* 16 (3): 247–61.

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: 161–87.
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): 249–67.
Millo, Y., and J. Lezaun. 2006. "Regulatory Experiments: Genetically Modified Crops and Financial Derivatives on Trial." *Science and Public Policy* 33 (3): 179–90.
Ministry of Education and Culture. 2021. *Two-year Pre-primary Education Trial 2021–2024*. Helsinki: Information booklet.
Mink, M., and J. De Haan. 2006. "Are There Political Budget Cycles in the Euro Area?" *European Union Politics* 7 (2): 191–211.
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.
Pearce, W., and S. Raman. 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 *The Experimenting Society. Essays in Honor of Donald T. Campbell*, edited by W. Dunn. New Brunswick, NJ: Transaction, 125–40.
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): 822–39.
Raudla, R., K. Sarapuu, E. Juuse, N. Harbuzova, K. Onno, J. Vallistu, and A. 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.
Rogers-Dillon, R.H. 2004. *The Welfare Experiments: Politics and Policy Evaluation*. Stanford, USA: Stanford University Press.
Rogoff, K. 1990. "Equilibrium Political Budget Cycles." *American Economic Review* 80 (1): 21–36.
Schedler, A., and J. Santiso. 1998. "Democracy and Time: An Invitation." *International Political Science Review* 19 (1): 5–18.
Sørensen, E., C. M. Hendriks, N. Hertting, and J. Edelenbos. 2020. "Political Boundary Spanning: Politicians at the Interface Between Collaborative Governance and Representative Democracy." *Policy and Society* 39 (4): 530–69.
Stoker, G. 2010. "Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and Policy Relevance." *Political Studies* 58 (2): 300–19.
——— and Evans, M., 2016. "Evidence-Based Policy Making and Social Science." In *Evidence-Based Policy Making in the Social Sciences*, edited by G. Stoker and Evans, M., 15–28. Bristol: Policy Press.
———, and P. John. 2009. "Design Experiments: Engaging Policy Makers in the Search for Evidence about What Works." *Political Studies* 57 (2): 356–73.
Strömbäck, J. 2008. "Four Phases of Mediatization: An Analysis of the Mediatization of Politics." *The International Journal of Press/Politics* 13 (3): 228–46.
——— and Esser, F. 2014. "Mediatization of Politics: Towards a Theoretical Framework." In *Mediatization of Politics: Understanding the Transformation of Western Democracies*, edited by J. Strömbäck, and F. Esser, 3–28. London: Palgrave Macmillan.
Strydom, W. F., N. Funke, S. Nienaber, K. Nortje, and M. Steyn. 2010. "Evidence-Based Policymaking: A Review." *South African Journal of Science* 106 (5): 1–8.
Torfing, J., and C. Ansell. 2017. "Strengthening Political Leadership and Policy Innovation Through the Expansion of Collaborative Forms of Governance." *Public Management Review* 19 (1): 37–54.
Trein, P., and T. Vagionaki. 2022. "Learning Heuristics, Issue Salience and Polarization in The Policy Process." *West European Politics* 45 (4): 906–29.

Appendix I

Reference(s)	Institution	Interview time(s)
Estonian interviewees		
Est1–Est15	Ministry of Finance	16.06.2022–06.01.2023
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

Appendix 4

Publication IV

Harbuzova, N. (2025). Choix de conception et perspectives de passage à l'échelle dans le pilotage des politiques: exploration des relations dans le domaine des politiques de création d'emplois (Original: Design choices and scaling prospects in policy piloting: exploring the relationship in the domain of job creation). *In:* Duverger, T. et al. (eds). *Territoires zéro chômeur de longue durée: de l'expérimentation à l'institutionnalisation*. Le Bord de L'eau, 45–72 (Ch. 3; tr. in French). (3.2)

Territoires zéro chômeur de longue durée : de l'expérimentation à l'institutionnalisation

Collection « Territoires de l'ESS »
dirigée par Timothée Duverger

www.editionsbdl.com
33300 Bordeaux

ISBN : 978-2-38519-140-5

Timothée Duverger, Victoria Bazurto,
Mathias Guérineau et Julien Kleszczowski
(sous la direction de)

Territoires zéro chômeur de longue durée : de l'expérimentation à l'institutionnalisation

LE BORD DE L'EAU

Chapitre III

Choix de conception et perspectives de passage à l'échelle dans le pilotage des politiques : exploration des relations dans le domaine des politiques de création d'emplois

Nastassia Harbuzova[1]

Les programmes innovants de création d'emplois gagnent du terrain. À la lumière de l'appel à la mise à l'essai pilote du FSE+ (annoncé en avril 2024), l'une des principales questions auxquelles les initiateurs potentiels se doivent de répondre est de savoir comment concevoir les pilotes : quels choix de conception intégrer pour maximiser les chances d'accroissement d'échelle du projet en cas de succès ? Si les recherches existantes indiquent que le mode de conception des pilotes politiques influe sur leurs perspectives de transposition à plus grande échelle, elles ne donnent aucune clé systémique pour comprendre comment cela se produit : les données probantes sont limitées, mitigées et éparpillées entre les différents domaines politiques.

Pour tenter de combler cette lacune, cette étude propose une analyse intégrée des implications que la conception de pilotes politiques peut avoir sur leurs perspectives de passage à l'échelle. Trois nouveaux programmes d'emploi correspondant à différents types de conception – limite *(boundary)* : Territoires zéro chômeur de longue durée, TZCLD (France), technocratique *(technocratic)* : Garantie d'emploi MAGMA (Autriche) et plaidoyer *(advocacy)* : Kinofelis Ergasia (Grèce) – sont comparés selon trois dimensions de conception – autonomie des ressources de l'initiateur, architecture du réseau et production et diffusion des connaissances – afin d'examiner l'incidence des différences de conception dans leurs trajectoires d'adoption.

1 Chercheure doctorante, Université de technologie de Tallinn.

Les résultats démontrent que, si les choix de conception individuels sont invariablement ambigus dans leurs implications, l'« avantage comparatif » de chaque programme est associé à une dimension de conception distincte : l'autonomie des ressources pour le plaidoyer Kinofelis, l'architecture de réseau pour la limite TZCLD et la production de connaissances pour le MAGMA technocratique.

À l'appui des enseignements tirés de l'analyse, l'étude se conclut par des recommandations pratiques sur les modes de conception, dans l'espoir d'informer d'autres lancements de pilotes, tant dans le domaine des politiques de création d'emplois que dans d'autres domaines.

Introduction

La reconnaissance croissante de l'expérimentation politique en tant que stratégie pour pallier les incertitudes modernes a alimenté les appels en faveur de la mise à l'essai pilote de politiques dans différents domaines (Checkland *et al.,* 2021 ; Neuwinger, 2022 ; Voß & Simons, 2018). Les pilotes politiques illustrent l'un des principaux types d'expérimentation politique (avec les expériences de conception, la législation expérimentale et les bacs à sable réglementaires) et peuvent être globalement définis comme des interventions limitées dans le temps et dans l'espace visant à tester de nouvelles solutions politiques et à utiliser les données pour décider de leur avenir (Bailey *et al.,* 2019 ; Raudla *et al.,* 2024). Selon la conception rationnelle de l'expérimentation politique qui domine encore dans la littérature (par exemple, selon Huitema *et al.,* 2018), le processus de mise à l'essai pilote est d'une simplicité attrayante et suit une logique linéaire : si des données expérimentales prouvent qu'une politique pilote est couronnée de succès, cette politique doit normalement bénéficier d'un accroissement d'échelle, c'est-à-dire être étendue dans le temps et/ou dans l'espace (Bravo-Biosca, 2020 ; Marston & Watts, 2003 ; Stoker, 2010).

Cependant, ce que l'on attend des pilotes politiques va souvent à l'encontre de ce que l'on constate dans l'élaboration des politiques dans le monde réel. Un nombre croissant d'études empiriques indique que de nombreux pilotes politiques fructueux ne sont jamais transposés à plus grande échelle (Koon *et al.,* 2020 ; Mauw *et al.,* 2023 ; Cote-Boileau *et al.,* 2019). Dans le même temps, les études existantes sur l'innovation sociale montrent que l'adoption des nouvelles solutions au-delà de l'expérimentation est largement influencée par les choix de conception intégrés dans cette dernière – en termes d'initiateurs des pilotes politiques, d'acteurs impliqués dans le pilotage, de répartition des pouvoirs et des rôles entre eux, de données

probantes collectées et de modes de collecte, d'évaluation et de diffusion des données probantes (par exemple, Torrens *et al.,* 2018 ; Nair&Howlett, 2016 ; Reymond *et al.,* 2020 ; Spicer *et al.,* 2014). Pourtant, si les données probantes abondent pour démontrer que les résultats des pilotes politiques sont influencés par leur conception, ces données probantes sont incohérentes, mélangées à d'autres facteurs (par exemple, la politique au sens large et un contexte d'expérimentation plus vaste) et éparpillées parmi les différents domaines politiques.

Cette étude contribue à combler les lacunes dans le domaine des politiques de création d'emplois. À l'heure actuelle, ce domaine n'est pas visible dans la littérature sur l'élaboration expérimentale de politiques, d'où la contribution théorique que ce chapitre espère apporter. Cependant, l'intérêt croissant pour les « garanties d'emploi » observé au niveau européen et dans le monde entier (Markowitsch & Scharle, 2024) augmente les chances que les résultats soient directement appliqués dans la pratique.

Pour combler cette lacune, ce chapitre propose une analyse empirique comparative des implications que les choix de conception intégrés dans les pilotes politiques peuvent avoir sur leurs trajectoires de passage à l'échelle.

L'objectif de l'étude est donc d'examiner comment la manière dont les pilotes politiques sont mis en place et gérés influe sur leurs perspectives d'adoption en comparant les forces et les faiblesses d'expérimentations politiques conçues différemment. Pour ce faire, l'étude s'appuie sur le cadre de la « conception de gouvernance » (McFadgen & Huitema, 2017a et 2017b), qui intègre des choix de conception autrement dispersés en trois types « idéaux » d'expérimentation politique : technocratique, limite et plaidoyer. Sur la base d'une étude de cas comparative de trois pilotes politiques novateurs de création d'emplois correspondant à différents types de conception – limite : Territoires zéro chômeur de longue durée, TZCLD (France), technocratique : Garantie d'emploi MAGMA (Autriche) et plaidoyer : Kinofelis Ergasia (Grèce) –, le chapitre identifie tout d'abord les dimensions clés des choix de conception qui différencient les programmes, puis compare la manière dont ces choix façonnent les trajectoires de passage à l'échelle des programmes.

Les recommandations issues de l'analyse sont proposées pour aider les praticiens à concevoir des politiques pilotes dans le domaine des politiques de l'emploi et au-delà.

Le chapitre est structuré comme suit : la deuxième section présente le cadre analytique et explique son développement ; la troisième section justifie l'échantillon d'études de cas et présente les méthodes de travail avec les données ; la quatrième section présente les résultats de l'analyse, suivie des conclusions, des recommandations et des limites de la recherche dans la cinquième section.

Cadre analytique

L'analyse de la littérature révèle que les relations entre la conception de la gouvernance des expérimentations politiques et leurs résultats n'ont pas, à ce jour, été étudiées de manière cohérente. Au contraire, des études distinctes indiquent que :

- le type d'agence initiatrice détermine le(s) objectif(s) des expérimentations lancées et le type de données probantes qu'elles tendent à produire (par exemple, Ettelt *et al.,* 2015 ; Lee & Ma, 2020 ; Cairney & Oliver, 2017),

- la présence d'acteurs spécifiques (par exemple, des décideurs politiques) parmi les parties prenantes des expériences influe sur leurs perspectives de passage à l'échelle (par exemple, Mauw *et al.,* 2023 ; Strassheim & Kettunen, 2014 ; Corduneanu-Huci *et al.,* 2021 ; Nair & Howlett, 2016),

- les approches de la collecte et de l'évaluation des données probantes adoptées dans le cadre des expérimentations politiques (qui, quoi, comment) influencent la perception de la crédibilité de leurs résultats (par exemple, Bedecarrats *et al.,* 2019 ; Checkland *et al.,* 2021 ; Farrelly, 2008),

- les dispositions régissant les relations entre les parties prenantes, les décideurs et les bailleurs de fonds de l'expérimentation influencent les chances de passage à l'échelle des solutions expérimentales. Ces dispositions comprennent les modèles de prise de décision, les approches managériales, les stratégies de communication et de diffusion adoptées dans les expériences (par exemple, Reymond *et al.,* 2020 ; Kuchenmüller *et al.,* 2022 ; Vreugdenhil, 2009 ; Guerineau *et al.,* 2024).

La seule exception trouvée dans la littérature au début de l'étude (2022) est la recherche de B. McFadgen et D. Huitema (McFadgen & Huitema, 2017a et 2017b, 2018), qui ont adopté une approche systématique des relations « conception-résultats » esquissées. Les auteurs ont mis au point un cadre qui intègre des choix de conception autrement dispersés dans la notion holistique de « conception

de gouvernance », en distinguant trois types « idéaux » de conception l'expérimentation politique :

- technocratique – initiatives politiquement neutres et (quasi) rationnelles,
- limite – initiatives ascendantes collaboratives et consensuelles et
- plaidoyer – initiatives descendantes autoritaires.

Le tableau 1 résume les principales caractéristiques de chaque type « idéal » de conception.

Tableau 1. Conception de la gouvernance des expérimentations politiques : comparaison des types « idéaux »

Type de conception / **Caractéristiques**	**Technocratique (*technocratic*)**	**Limite (*boundary*)**	**Plaidoyer (*advocacy*)**
Objectif escompté	Tester un nouvel outil et produire des données probantes pour l'élaboration de politiques	Parvenir à l'adoption d'un outil développé de manière consensuelle	Pousser l'action politique dans la direction souhaitée : prouver, atténuer les objections ou retarder la décision
Type d'initiateur et réseau de parties prenantes	Décideurs politiques/à leur demande, fonctionnaires	Base, généralement des ONG	Décideurs politiques
	Réseau fermé, accès sur invitation, des experts évaluent l'E	Réseau ouvert : différents types d'acteurs interagissent dans la conception, l'évaluation et la prestation de l'E	Réseau fermé, accès sur demande/invitation, les initiateurs contrôlent la conception, l'évaluation et la prestation de l'E
Formes de données probantes	Scientifiques, sans contexte, évaluées par rapport à certains critères prédéfinis	Diverses, contextualisées	Probablement symboliques et/ou utilisées de manière sélective
Rôle de la politique au sens large	La neutralité politique est supposée, aucun décideur politique ne participe	Projet à caractère politique : les décideurs politiques facilitent/participent	Projet intrinsèquement politique
E – Expérimentation			

Le type d'expérimentation politique est identifié en fonction des caractéristiques prédominantes dans six groupes de choix de conception, ou « règles ». Les « règles » décrivent le processus expérimental en termes de : qui est impliqué et exclu (règles de délimitation), qui fait quoi (règles de position), comment l'autorité est répartie entre les parties prenantes (règles de choix), comment les décisions sont prises (règles d'agrégation), quelles informations sont produites et comment elles sont diffusées (règles d'information) et comment l'initiative est financée (règles de rentabilité).

À ce jour, le cadre a été utilisé dans des recherches quantitatives pour étudier la manière dont la conception des expérimentations politiques influence leurs résultats d'apprentissage et leur capacité à éclairer les décideurs politiques (McFadgen & Huitema, 2017a et 2017b, 2018). La présente étude, en revanche, opte pour une approche désagrégée, « choix par choix », afin de discerner l'incidence que peuvent avoir les choix de conception individuels inclus dans la notion de « conception de gouvernance » sur les perspectives des pilotes politiques.

Partant, sur la base de l'analyse documentaire et des données empiriques recueillies dans les études de cas, le cadre original a été adapté de manière que les six groupes de choix de conception avec des critères sous-jacents complexes soient réagrégés en trois dimensions de conception, comme suit :

- Autonomie des ressources de l'initiateur : différencie les pilotes politiques en fonction de l'accès de leurs initiateurs à l'autorité et aux fonds nécessaires au lancement, à la mise en œuvre et à l'accroissement d'échelle des pilotes politiques.

- Architecture du réseau : différencie les pilotes politiques en fonction des acteurs impliqués et de leurs modes d'interaction.

- Production et diffusion de connaissances : différencie les pilotes politiques en fonction de leur approche de la collecte, de l'évaluation et de la diffusion de données probantes.

Les nouvelles dimensions de la conception génèrent une structure plus claire et plus simple pour l'analyse comparative des études de cas**.** Chaque dimension intègre les principaux choix de conception individuels qui différencient les pilotes politiques présentant différents types de conception. La figure 1 illustre le processus d'adaptation du cadre, en résumant les choix de conception qui sous-tendent chaque dimension.

Le cadre adapté sera ensuite utilisé pour : a) démontrer que les programmes sélectionnés pour l'analyse relèvent de trois types de conception différents, b) mettre en évidence les différences de leur conception et, par conséquent, c) structurer l'analyse des implications que ces différences ont pour les perspectives d'adoption des programmes – l'annexe 1 résume les résultats.

Figure 1. Adapter le cadre aux besoins de l'approche « choix par choix »

Cadre original : six blocs de choix de conception

Groupes de choix de conception	Description [critères agrégés]
1) Délimitation	Qui est impliqué et qui est exclu
2) Position	Qui fait quoi
3) Choix	Comment l'autorité est répartie entre les parties prenantes
4) Agrégation	Comment les décisions sont prises
5) Informations	Quelles informations sont produites
6) Rentabilité	Comment l'initiative est financée

Cadre adapté : trois dimensions de conception

Dimensions de conception	Choix de conception inclus [critères]
1 Autonomie des ressources de l'initiateur	Autorité d'origine
	Sources de financement
2 Architecture du réseau	Liaisons et institutions
	Mode de prise de décision
	Implication des décideurs politiques
3 Production et diffusion de connaissances	Stratégie d'apprentissage
	Activité de diffusion

Études de cas et collecte de données

Pour l'analyse empirique, les programmes suivants ont été sélectionnés :

• Garantie d'emploi MAGMA *(Modellprojekt Arbeitsplatzgarantie Marienthal)* : essai contrôlé randomisé (ECR) à petite échelle initié par le service public local de l'emploi (SPE) en étroite collaboration avec le monde universitaire ; en cours à Gramatneusiedl, en Basse-Autriche, de 2020 à 03.2024 ; terminé à la fin de la durée prévue de l'essai pilote, sans passage à l'échelle ; 150 résidents travailleurs.

• Territoires Zéro Chômeur de longue durée, TZCLD : expérimentation française en cours initiée par l'association d'ONG en 2016, essai non randomisé ; prolongée pour un second mandat (2016-2021, 2022-2026) et élargie de 10 à 75 territoires : 3 281 travailleurs (en septembre 2024).

• Kinofelis Ergasia (*ΚΟΙΝΩΦΕΛΗΣ ΕΡΓΑΣΙΑ,* ou *Bien commun*) : programme national pour l'emploi en cours et régulièrement mis en œuvre, initié par le ministère grec du Travail (MT) ; d'abord lancé en tant que projet pilote en 2011 (expérimentation de conception non randomisée) ; en 2016, la version améliorée du programme a été introduite, passage à l'échelle et adoptée de manière permanente ; accroissement d'échelle en trois étapes : de 17 (2011-2013) à 51 municipalités (2013-2015), puis à l'ensemble des 326 municipalités (2016) ; depuis 2017 : réduction à 25 000 - 45 000 inscriptions/an.

Les programmes constituent un échantillon comparatif solide qui obéit aux objectifs de l'étude et satisfait tous les critères établis pour trouver les pilotes éligibles. Plus précisément, les programmes :

a) ont été lancés récemment sous la forme d'essais pilotes de politiques (ou ont été officiellement désignés comme tels) – c'est-à-dire que leurs résultats sont pertinents et comparables dans le temps.

b) pilotent un outil politique similaire sur le plan conceptuel.
Les programmes sont cités comme des exemples concrets de la Garantie d'emploi (par exemple, Tcherneva & Lalucq, 2022) – une politique active du marché du travail qui entend éradiquer le chômage de longue durée en créant des opportunités d'emploi volontaires, adaptées aux besoins individuels, bénéfiques au niveau local, non concurrentielles par rapport à l'économie traditionnelle et accompagnées d'un salaire décent *a minima* (par exemple, Tcherneva, 2018). Bien que chaque projet pilote ait ses propres caractéristiques, tous les programmes répondent majoritairement à cette définition conceptuelle.

c) illustrent les différents types de conception de gouvernance.
Chaque programme adopte son propre mode de mise en œuvre politique (administration, évaluation, fonctionnement) et représente l'un des trois types « idéaux » de conception : MAGMA – technocratique, TZCLD – limite, Kinofelis – plaidoyer –, justifiés par le cadre adapté – voir le tableau 2.

Tableau 2. Aperçu comparatif des conceptions des études de cas

Dimension de conception	**Critères [choix de conception]**	**TZCLD**	**MAGMA**	**Kinofelis**
Autonomie des ressources de l'initiateur	Autorité d'origine Sources de financement	Non : - ONG - fonds publics [majoritairement], adhésion des parties prenantes [gestion uniquement]	Limitée : - SPE local - autofinancé [budget disponible]	Complète : - MT - fonds à disposition [FSE+, budget national]
Architecture du réseau	Liaisons et institutions	Innovante	Existante, bien établie	Existante, bien établie
	Mode de prise de décision	Consensuelle	Autoritaire et consensuelle	Autoritaire
	Implication des décideurs politiques	Forte facilitation politique, projet à caractère politique	Politiquement neutre [à l'origine]	Projet intrinsèquement politique [initiateur = décideur politique]
Production et diffusion de connaissances	Stratégie d'apprentissage	Apprentissage pour prouver le succès et adapter la mise en œuvre de la politique	Apprentissage pour tester la réussite	Apprentissage pour optimiser la mise en œuvre de la politique
	Activité de diffusion	Promotion internationale active dans les médias, résultats et mises à jour	Promotion internationale active, principalement des résultats*	Promotion limitée pendant le pilotage, active – après
CONCLUSION DU TYPE DE CONCEPTION		LIMITE (*BOUNDARY*)	TECHNOCRATIQUE (*TECHNOCRATIC*)	PLAIDOYER (*ADVOCACY*)
		Pilote collaboratif ascendant visant à adopter une solution consensuelle	Pilote local quasi normatif visant à tester une solution	Pilote autoritaire descendant visant à adopter une solution par étapes
* La seule caractéristique qui différencie MAGMA du type « idéal »				

Pour que l'échantillon reste comparable et propice à la collecte des données, la recherche a été limitée aux pays de l'Union européenne. Au moment où cette étude a commencé (début 2022), les programmes choisis semblaient être les seuls exemples éligibles.

L'étude s'appuie sur des méthodes de recherche qualitatives. Pour collecter les données, la recherche documentaire a été complétée par 18 entretiens semi-structurés avec les parties prenantes des programmes (5 pour TZCLD, 7 pour MAGMA, 6 pour Kinofelis). Les entretiens ont été menés en ligne et ont duré une heure en moyenne ; ils ont été enregistrés et transcrits. Afin de préserver l'anonymat des personnes interrogées, celles-ci seront identifiées par le code de leur pays (par exemple, « Fr1/Au1/Gr1 »).

Pour chaque étude de cas, les entretiens et les données secondaires ont été utilisés de manière interactive et itérative, pour combler les lacunes des uns et des autres et vérifier les conclusions tirées. Lors de la collecte des données, l'accent a été mis sur les détails de la conception de gouvernance, des défis posés par le pilotage, des perspectives d'accroissement d'échelle et des enseignements tirés.

Alors qu'au niveau international, les programmes analysés sont présentés comme une réussite (par exemple, Tcherneva&Lalucq 2022), leurs perspectives nationales ne sont pas aussi définitives et varient considérablement (TZCLD – un pilote étendu en cours, MAGMA – un pilote qui n'a pas été mis à l'échelle, Kinofelis – un pilote qui est devenu un programme national régulier). La section suivante élucide dans quelle mesure et comment la conception de la gouvernance du programme peut expliquer ces différences de transposition à plus grande échelle.

Résultats

Autonomie des ressources de l'initiateur

L'accès à l'autorité et aux fonds nécessaires au lancement, à la mise en œuvre et au passage à l'échelle des pilotes politiques s'est révélé constituer le premier aspect principal de la conception de la gouvernance des pilotes politiques qui différencie les études de cas et affecte leurs perspectives d'adoption.

Pas d'autonomie : TZCLD

Dans le cas de TZCLD, qui est piloté par la base, l'État est le principal pourvoyeur de fonds et des autorisations légales nécessaires au maintien et au développement du pilote. Par conséquent, les perspectives du projet dépendent des priorités de l'État en matière d'affectation des ressources (Fr2) et restent vulnérables à une série de facteurs spécifiques, à savoir :

- les réductions et interruptions de financement, y compris la réticence de l'État à augmenter les fonds proportionnellement à l'expansion du programme – ce qui exige des parties prenantes qu'elles soient vigilantes et capables de se mobiliser et de défendre les besoins du programme. Parmi les faits marquants récents, les personnes interrogées ont noté la nécessité de réclamer une augmentation adéquate dans le budget expérimental pour 2024 (satisfait) et les risques de réduction du financement en raison des Jeux olympiques de Paris (Fr5, discussion de groupe).

- une tension sur les ressources liée à l'incertitude du chevauchement fonctionnel entre le programme et le système de SPE nouvellement réformé (en cours par l'État parallèlement à l'expérimentation). La réforme oblige les parties prenantes à imaginer le futur fonctionnement de TZCLD dans le nouveau cadre institutionnel et à prouver sa compatibilité avec celui-ci pour poursuivre le projet sous sa forme actuelle ou adoptée (Fr1, Fr4).

- les questions de « déresponsabilisation » liées à la réticence à réaffecter des ressources publiques des autorités nationales aux collectivités locales (Fr4). Un accroissement de l'autonomie des acteurs locaux impliqués dans le pilotage peut potentiellement réduire les avantages des décideurs politiques au niveau national (y compris les réductions des fonds disponibles et le contrôle réduit sur les territoires expérimentaux) – ce qui peut priver le programme du soutien des principaux décideurs politiques.

- la remise en cause du *statu quo* ou la concurrence financière avec les politiques publiques conventionnelles dans le domaine. La « rigidité » des pratiques existantes est citée comme un obstacle potentiel à l'accroissement d'échelle (Fr2-4).

Dans le même temps, le manque d'autonomie des ressources et la conscience même d'une telle « vulnérabilité par défaut » ont encouragé les initiateurs à développer des adaptations efficaces pour accroître la résilience du pilote face aux priorités potentiellement évolutives du décideur – y compris les leviers du soutien

juridique (lancement du pilote par le biais de lois expérimentales), la facilitation politique et le renforcement des capacités (approche anticipative, vision à long terme, agilité, communication efficace).

Autonomie limitée : MAGMA

L'exemple de MAGMA montre comment une autonomie des ressources suffisante au lancement du pilote, mais pas à son passage à l'échelle, peut le placer dans un dilemme où l'absence de stratégie de passage à l'échelle se heurte à un agenda politique inopportun.

Le SPE local à l'origine de MAGMA est autorisé à piloter des innovations sans la permission du ministère ou du SPE fédéral tant que le coût reste dans les limites de son budget disponible (Au7). À première vue, ce mode d'autonomie pourrait placer MAGMA dans une position plus avantageuse que TZCLD, étant donné qu'une certaine marge de manœuvre pour des décisions politiquement neutres est prévue : par exemple, on pourrait s'attendre à ce que le programme soit prolongé au sein de la juridiction autorisée en cas de succès. Toutefois, ce cas démontre que les chances d'un passage à l'échelle politiquement neutre sont très limitées dans la pratique.

D'une part, le passage à l'échelle devient politique dès lors qu'un soutien extérieur est nécessaire. Bien que MAGMA se soit avéré un succès, il est apparu que le coût de son déploiement permanent dépassait le budget du SPE (Au7, Au4). En d'autres termes, une adhésion externe était déjà nécessaire pour prolonger le programme à une échelle similaire.

Dans ce contexte, l'agenda défavorable du gouvernement conservateur et le faible taux de chômage dans le pays ont empêché toute forme d'extension du programme au-delà de la période pilote convenue (Au3, Au5-6).

D'autre part, le passage à l'échelle du programme dépend également de la rotation des instances dirigeantes de l'agence initiatrice. Le risque de ne pas s'aligner sur les priorités du nouvel exécutif a été cité comme un défi potentiel aux progrès de MAGMA, au même titre que la contrainte budgétaire (Au7).

En outre, l'autonomie des ressources limitée de l'agence initiatrice a imposé les contraintes suivantes aux perspectives du pilote :

- absence de stratégie d'accroissement d'échelle et d'adaptations. La capacité de l'initiateur à lancer le pilote dans le cas de MAGMA n'a pas motivé le développement d'adaptations à l'incertitude politique. Plus précisément, le

mode d'extension du programme en cas de succès n'a pas été étudié (Au1-2, Au4, Au7) et l'option de l'accroissement d'échelle n'a pas été jugée légitime, ce qui a laissé le programme « pieds nus » face à un agenda politique défavorable.

• préoccupations concernant la validité externe des résultats du pilote : la taille du budget disponible pour l'expérimentation a été citée parmi les raisons expliquant la petite taille du pilote (150 résidents) (Au7), ce qui risque au final de limiter les perspectives du programme. Si les données probantes produites ont été jugées valables sur le plan interne (Au1-2), les parties prenantes ont émis des doutes quant à leur potentiel de généralisation (Au1), arguant qu'il valait mieux comprendre le contexte avant de chercher à étendre le programme à la région ou à l'ensemble du pays (Au4).

Autonomie complète : Kinofelis

Lancé par le ministère grec du Travail, Kinofelis a profité de l'engagement politique « renforcé ». L'autonomie des ressources dont disposait l'initiateur a permis l'adoption du programme à l'échelle nationale après la phase pilote, mais s'est révélée être un privilège à court terme dépendant d'un changement de gouvernement. Le caractère court-termiste de l'engagement politique est considéré comme un défi majeur pour le maintien de Kinofelis sous les gouvernements successifs (Gr1, Gr4-5). Après le remplacement d'un gouvernement de gauche radicale par un gouvernement de centre-droit, le programme a été spectaculairement réduit et sa qualité s'est détériorée (Gr4).

Néanmoins, malgré l'absence de soutien politique pour les deux cycles suivants de gouvernement de droite, le programme n'a pas été interrompu en raison du coût politique élevé imposé par le dirigeant pilote dans un tel scénario. Une légitimation appropriée du programme à l'échelle et une demande stable de la population créée par des investissements dans l'optimisation de la mise en œuvre de la politique ont été citées comme raisons expliquant la poursuite des inscriptions au programme Kinofelis (Gr4).

Architecture du réseau

L'architecture du réseau des pilotes politiques – c'est-à-dire les modes d'interaction établis entre les parties prenantes impliquées dans le pilote ainsi que les

caractéristiques des parties prenantes elles-mêmes – est la deuxième unité identifiée de la conception de gouvernance qu'il importe de prendre en compte dans l'analyse. Les trois aspects de l'architecture du réseau qui se sont révélés différencier les études de cas et influencer leurs perspectives seront examinés plus en détail : les types d'institutions et de liens impliqués dans le processus de pilotage politique, le mode de prise de décision adopté dans le pilote et la présence de décideurs politiques parmi les parties prenantes.

Institutions innovantes versus *institutions existantes :* *TZCLD* versus *MAGMA et Kinofelis*

Le fait qu'un programme pilote s'appuie sur des liens existants et des institutions existantes ou qu'il faille en créer de nouveaux s'avère avoir une incidence sur le coût de son passage à l'échelle. Le coût de l'adoption du programme à grande échelle est considéré comme un obstacle majeur à l'extension de TZCLD (Fr2-4), où la mise en œuvre et la gestion du programme reposent sur les nouvelles entités créées par le projet. Par conséquent, pour mettre le programme à l'échelle, il faut se mettre d'accord sur la manière de légitimer ces structures innovantes (Fr1, Fr2), de les intégrer dans le système des institutions existantes (Fr2, Fr4) et de les faire fonctionner à grande échelle (Fr2-3). Par conséquent, une adoption « tronquée » du programme devient un scénario probable (Fr4).

En revanche, MAGMA et Kinofelis présentent un avantage en termes de coûts pour ce qui est de l'extensibilité, étant donné que tous deux ont adopté des liens et des installations bien établis pour mettre en œuvre les programmes (Gr6). Dans les deux cas, les bureaux locaux des SPE sont les principaux exécutants du programme : le fonctionnement de MAGMA repose sur une collaboration horizontale entre l'AMS Niederösterreich (PSE initiateur) et son partenaire habituel it.works – un prestataire de services du marché du travail à but non lucratif (responsable de l'orientation et de la formation des participants) ; Kinofelis, pour sa part, repose sur une interaction verticale entre le MT (qui définit les conditions et rembourse les coûts salariaux des emplois créés) et les bureaux municipaux du SPE qui créent et gèrent directement les projets d'emploi.

Simultanément, cependant, le cas de Kinofelis démontre qu'un accroissement d'échelle efficace peut nécessiter des changements dans les structures institutionnelles existantes, étant donné qu'elles peuvent comporter des risques de corruption et de favoritisme. Plus précisément, pour protéger le programme de la « mainmise politique », l'initiateur a dû remplacer la médiation de l'emploi sur place assurée par les centres syndicaux municipaux par un processus de candidature électronique transparent (Gr1).

Prise de décision consensuelle versus *prise de décision autoritaire : TZCLD* versus *MAGMA et Kinofelis*

Le cas de TZCLD montre comment le processus décisionnel inclusif et collaboratif adopté dans le cadre du pilote encourage sa transposition à plus grande échelle. Le mode consensuel de prise de décision dans lequel les parties prenantes du pilote ayant des rôles et des affiliations politiques différents conservent les moyens de poursuivre le développement du programme a transformé la diversité politique des parties prenantes de TZCLD en avantage stratégique pour le programme, lui permettant de s'étendre sur des territoires français politiquement hétérogènes, tant intérieurs et qu'extérieurs. Dans ce processus, initialement associé à la politique de gauche, le programme neutralise son affiliation politique en démontrant que la gauche et la droite peuvent « faire le projet ensemble » (Fr4).

Parallèlement, le réseau dans lequel la base et l'organisme de gestion de l'expérimentation (y compris les responsables de la mise en œuvre et de la recherche) peuvent interagir directement (Fr1, Fr4-5) nourrit un sentiment d'engagement politique partagé entre les parties prenantes (Fr1). Une telle attitude permet au pilote de mobiliser une coalition de soutien « de l'intérieur » et « en cours de route », ce qui renforce sa résilience et l'aide à accéder aux ressources nécessaires pour soutenir et étendre le projet à de nouveaux territoires. Des tensions peuvent toutefois se faire jour en cas de changement d'échelle : par exemple, sur la participation volontaire des demandeurs d'emploi, alors que les systèmes d'obligation et de sanction se durcissent par ailleurs.

Dans MAGMA et Kinofelis, les relations entre les parties prenantes sont « soumises à un protocole » et le pouvoir de décision est soit distribué aux nœuds fonctionnels (MAGMA), soit fortement concentré entre les mains de l'instance dirigeante (Kinofelis). Dans de tels contextes, le sentiment d'engagement commun à l'égard de la politique n'a aucune raison d'évoluer parmi les parties prenantes, de sorte que les programmes sont considérés comme « un projet à court terme de plus » (Au4**,** Gr2).

Implication des décideurs politiques : MAGMA versus *TZCLD*

L'ambiguïté des implications que l'engagement politique peut avoir sur les perspectives des pilotes dépendant des ressources est explicitée dans les exemples de MAGMA et de TZCLD. Bien qu'ils reconnaissent tous deux l'importance de la politique au sens large pour leur accroissement d'échelle, en admettant que la décision de faire d'une politique expérimentale dont le succès a été prouvé un outil permanent est une décision politique (Fr2-5, Au1-6, Gr1, Gr4-6), la nécessité d'impliquer

les décideurs politiques dans les expérimentations est considérée comme une opportunité dans le cas de TZCLD et comme un défi dans le cas de MAGMA – selon le point de vue : crédibilité des données probantes ou facilitation de la politique.

Dans TZCLD, la facilitation politique fait figure de levier essentiel qui permet de lancer l'expérimentation et l'aide à accéder et à mobiliser les ressources nécessaires (Fr1, Fr5).

Dans MAGMA, la non-implication des décideurs politiques est citée comme une condition préalable au lancement du programme et comme son principal avantage en termes de crédibilité (Au7). Toutefois, dans la pratique, il s'agit d'une médaille à deux faces : alors que le fait de s'assurer que les données probantes n'ont pas d'affiliation politique était un point essentiel pour l'initiateur (Au7), son propre plaidoyer de gauche – à la lumière de sa transition de haut fonctionnaire à président des sociaux-démocrates (parti de centre-gauche) en Basse-Autriche – devient maintenant un levier pour poursuivre le programme après le pilotage, sans le soutien du gouvernement et l'autonomie des ressources nécessaires pour un passage à l'échelle politiquement neutre (Au4, Au7).

Production et diffusion de connaissances

Les stratégies d'apprentissage que les pilotes politiques adoptent en termes de connaissances à produire et de modalités de production constituent la troisième unité d'analyse principale, qui présente des opportunités et des défis distincts en termes d'accroissement d'échelle pour les programmes étudiés. Bien que les politiques aient été évaluées dans le cadre de chaque programme, les aspects mis en avant et les méthodes appliquées diffèrent, ce qui démontre que le choix de la stratégie d'apprentissage est une question normative qui dépend de l'objectif du pilotage.

Apprentissage pour optimiser la mise en œuvre de la politique : Kinofelis

Pour Kinofelis, l'objectif de l'expérimentation était d'optimiser le système avant de l'adopter à l'échelle nationale (Gr1-2, Gr4-5). Bien que qualifié de « pilote » (par exemple, OIT 2018), le processus s'apparentait davantage à un prototypage de politique, car le passage à l'échelle avait déjà été décidé sur la base de l'engagement personnel de l'initiateur et de données probantes issues de recherches antérieures (Gr1).

Par conséquent, l'évaluation au cours des « phases pilotes » n'était pas axée sur la collecte de données probantes, mais sur l'identification des goulets d'étranglement du système aux fins de leur résolution (Gr1-2, Gr4-5). Cette approche pratique a permis de renforcer les capacités municipales nécessaires à la mise en œuvre efficace du programme à grande échelle et, par conséquent, de le soutenir à long terme en créant une demande stable chez la population grecque (Gr4).

Apprentissage pour prouver le succès et adapter la mise en œuvre de la politique : TZCLD

Les principales motivations du lancement de TZCLD étaient de démontrer que le système fonctionne, de le reproduire en fonction des besoins des différents territoires français et, enfin, de parvenir à son adoption légale à l'échelle nationale (Fr1-2, Fr5). Comme dans le cas de Kinofelis, les initiateurs du pilote prévoyaient son succès dès son lancement, mais, en ce cas, le succès devait être prouvé à l'État (Fr1).

Pour répondre à ces deux exigences – prouver et reproduire le système –, deux évaluations simultanées sont menées dans le cadre du projet. D'une part, le programme est évalué « de l'intérieur » par les parties prenantes du pilote. Les initiateurs ont adopté une approche large de la collecte de données probantes : de multiples types de connaissances (notamment des données probantes scientifiques, des apprentissages pratiques, des retours d'information de la base) sont générés et reconnus comme ayant la même valeur (Fr1, Fr4). Bien que cette évaluation ait permis de promouvoir l'« extensibilité » du programme – en expliquant comment il devrait être développé plus avant pour fonctionner efficacement dans divers territoires (Fr1), les données probantes qu'elle produit ne suffisent pas à convaincre l'État (Fr2-4).

D'autre part, le pilote est évalué par le Comité spécial régi par les lois expérimentales promulguant le programme TZCLD (2016, 2022). Cette évaluation, pour sa part, reflète les priorités de l'État et se concentre donc sur la collecte de données probantes scientifiques et l'analyse du rapport coût-bénéfice du projet (Fr3).

En conséquence, TZCLD obtient des données probantes contradictoires lorsque les résultats de l'évaluation interne contredisent les conclusions de l'évaluateur désigné, de sorte que le succès établi dans le cadre du pilote risque d'être dévalorisé par l'État (Fr4). En outre, si le Comité a été conçu à l'origine comme une solution contre l'évaluation dépendante, sa propre neutralité peut être sujette à débats. Comme l'ont expliqué les personnes interrogées, l'État dispose d'une marge de manœuvre pour « ajuster » les résultats de l'évaluation en choisissant les effets que le Comité doit ou non inclure dans le rapport coût-bénéfice du programme (Fr4). Par

conséquent, la solution est susceptible de déplacer, à défaut de bloquer, la possibilité pour les intérêts politiques de décider de l'avenir du pilote.

Apprentissage pour tester la réussite : MAGMA

Dans le cadre de MAGMA, l'objectif premier du pilote était de le tester véritablement et de produire des connaissances susceptibles d'éclairer d'éventuelles décisions politiques (Au2, Au7). À cet égard, de nombreux efforts ont été déployés pour garantir la crédibilité des données probantes expérimentales, notamment l'adoption d'un modèle de recherche « étalon » rigoureux pour évaluer le programme (essai contrôlé randomisé préenregistré), la participation d'évaluateurs réputés (universitaires d'Oxford et de Vienne) et le maintien de la neutralité politique du pilote, de sorte que les décideurs politiques ne puissent pas influencer les résultats de l'évaluation (Au4). Comme l'a précisé une personne interrogée, cette approche globale a permis au pilote d'influencer l'élaboration de politiques internationales malgré sa petite taille et son faible attrait politique au niveau local (Au4). En particulier, le succès de MAGMA a contribué à façonner l'argumentaire des rapports des Nations Unies sur la garantie de l'emploi (Au2, Schutter 2023) et à lancer l'appel à la mise à l'essai pilote de politiques de création d'emplois par le Fonds social européen Plus (Markowitsch&Scharle 2024).

En même temps, le programme a été exposé aux critiques politiques locales pour les résultats financiers qu'il a générés – qui n'ont pas été inclus dans l'évaluation expérimentale (Au1, Au3-4). Comme l'a expliqué une personne interrogée, l'absence de normativité concernant « ce qui doit être évalué » a laissé le champ libre au gouvernement conservateur pour dévaloriser le succès du pilote, si bien que les parties prenantes de MAGMA n'ont guère été surprises d'apprendre que le programme qui s'était avéré efficace ne serait pas mis à l'échelle (Au 4).

Le cas de MAGMA démontre donc que la grande crédibilité des résultats expérimentaux ne suffit pas à convaincre les décideurs politiques dès lors que l'adoption de l'outil pilote va à l'encontre de leurs priorités politiques.

Activité de diffusion

La communication autour des succès des politiques pilotes s'est révélée être un « multiplicateur d'impact » efficace qui permet aux pilotes de compenser leur échelle limitée et d'améliorer leurs perspectives d'adoption lorsque le soutien politique national et l'autonomie des ressources sont insuffisants pour continuer à les mettre en œuvre autrement. Les études de cas révèlent trois façons de procéder.

Premièrement, la promotion active de la réussite d'une politique pilote peut accroître ses chances d'adoption locale en augmentant le coût de résiliation. Une sensibilisation sociale accrue envers l'expérimentation en cours rend sa suspension politiquement inacceptable pour l'État (Fr4, Au7).

Deuxièmement, la diffusion internationale peut déboucher sur d'autres scénarios d'accroissement d'échelle de la politique. Par exemple, malgré un soutien politique ambigu en France, TZCLD a stimulé la mise à l'essai pilote de programmes similaires à travers l'Europe, notamment des initiatives en Belgique et en Italie (Fr4).

De même, la promotion internationale des programmes a alimenté l'appel du FSE+ à piloter des programmes de création d'emplois et a nourri l'ambition de la garantie européenne de l'emploi (Tcherneva&Lalucq 2022), ce qui, dans les deux cas, peut être considéré comme un « transfert d'échelle ».

Troisièmement, la promotion de pilotes politiques peut être un moyen d'imaginer d'autres leviers de facilitation : par exemple, en gagnant un soutien politique à l'étranger ou en renforçant l'argumentaire politique par la création d'une coalition internationale (telle que la Coalition européenne pour la garantie de l'emploi promue par TZCLD) (Au3, Gr1, Fr1, Fr4).

Conclusions

Cette étude a mené une analyse systémique des implications que le mode de conception des pilotes politiques peut avoir sur leurs perspectives d'extension dans le domaine des politiques de l'emploi. Les conclusions sont organisées en deux blocs.

Type de conception et « avantage comparatif » des pilotes politiques

Si les choix individuels pratiqués lors de l'initiation, de la mise en œuvre et de l'évaluation des programmes sont invariablement ambigus dans leurs implications, l'« avantage comparatif » d'un programme apparaît être associé à différentes dimensions de conception.

Les pilotes de plaidoyer peuvent tirer parti de l'autonomie des ressources accordée à leurs initiateurs à court terme. Dans le cas de Kinofelis, l'accès à l'autorité et aux fonds nécessaires pour accroître l'échelle du programme a permis à l'initiateur non seulement de l'adopter à l'échelle nationale, mais aussi d'imposer un coût politique élevé pour résilier le programme en l'absence de soutien politique de la part des gouvernements suivants. Les investissements réalisés pour renforcer les capacités municipales en vue d'une mise en œuvre efficace du programme à grande échelle

ont permis de garantir auprès de la population locale une demande élevée et durable pour Kinofelis. Conjugué à une légitimation appropriée, le programme a fini par être trop bien ancré dans le système des services de l'emploi conventionnels pour qu'il y soit mis fin.

Les pilotes limites peuvent bénéficier d'un réseau de parties prenantes très collaboratif et inclusif, ainsi que d'une prise de décision consensuelle. Dans le cas de TZCLD, de tels contextes facilitent l'extension du programme sur des territoires politiquement différents et renforcent la résilience aux changements d'agenda politique en gagnant le soutien politique de la droite comme de la gauche. La notion d'engagement politique commun qu'un tel réseau promeut contribue également à renforcer la résilience du projet.

Les pilotes dont la conception est essentiellement technocratique sont susceptibles d'être reconnus pour les données probantes scientifiquement rigoureuses et politiquement neutres qu'ils produisent, lesquelles peuvent ensuite être transformées en levier politique pour promouvoir le pilote à l'international et compenser son échelle limitée – comme l'illustre le cas de MAGMA.

Deux dimensions de l'accroissement d'échelle : adaptations de la conception pour de meilleures perspectives

Dans le même temps, l'analyse démontre que, à long terme, l'accroissement d'échelle des politiques demeure problématique pour tous les programmes, quel que soit le type de conception auquel ils répondent. En d'autres termes, aucun des types de conception n'est « plus fort » ou « plus faible » que l'autre : par exemple, si les pilotes de plaidoyer fructueux peuvent bénéficier de la protection de l'« engagement politique renforcé » contre la suspension, le maintien de cette politique contre les cycles gouvernementaux soulève des inquiétudes, tandis que les pilotes limites « vulnérables par défaut » peuvent, au contraire, apparaître comme les plus résilients à long terme.

Dans cette optique, il est essentiel de garder à l'esprit que la transposition à plus grande échelle d'une politique pilote est un processus à deux dimensions – le temps et la portée – et que les initiateurs doivent bien prendre en compte ces deux aspects lors de la conception du pilote :

- les questions relatives à l'extension de la politique, ou comment adopter une politique pilote à grande échelle (dimension de portée)
- les questions relatives au maintien de la politique, ou comment soutenir la politique à long terme (dimension de temps).

Les enseignements tirés de l'analyse suggèrent les adaptations suivantes au niveau de la conception. Comment adopter des politiques pilotes à grande échelle ?

► Promulguer des politiques pilotes par le biais de la loi et légitimer l'accroissement d'échelle en tant que résultat réalisable de l'expérimentation

L'adoption de pilotes par la voie législative – de sorte que les conditions et les résultats de l'expérimentation soient précisés – réduit le risque que l'initiative couronnée de succès soit suspendue pour des raisons politiques.

► Anticiper les problèmes d'« extensibilité » – et réduire le coût de l'adoption des politiques en :

- exploitant des structures institutionnelles existantes pour la mise en œuvre de la politique : l'accroissement d'échelle de projets fondé sur des liens et des entités bien établis coûte moins cher. Cela étant, le risque de corruption doit également être pris en compte.

- incluant la dimension de la mise en œuvre (connaissances pratiques) et le retour d'information de la base (connaissances profanes) dans l'évaluation pilote afin de déterminer les itérations du programme à effectuer et les capacités à renforcer pour optimiser la mise en œuvre du programme à l'échelle.

► Définir un dispositif adéquat de facilitation politique

Dans le cas où les initiateurs de pilotes sont privés d'autonomie des ressources, les leviers de facilitation doivent être bien pensés, en termes de :

- source et canaux de facilitation : rechercher un soutien externe ou le développer « de l'intérieur » (par exemple, lorsqu'une partie prenante se transforme en décideur politique)

- discours de la facilitation : identifier un discours approprié pour la promotion des politiques, notamment en termes d'affiliation politique (par exemple, neutraliser l'affiliation ou l'utiliser comme levier de plaidoyer).

► Diffuser les connaissances produites au niveau international

Une large diffusion de la réussite du pilote peut augmenter ses chances d'adoption au niveau national (en créant un coût de résiliation par voie de sensibilisation sociale), débloquer d'autres scénarios d'accroissement d'échelle (étendre le pilote à l'étranger,

motiver la mise à l'essai pilote au niveau supranational) ou concevoir d'autres leviers de facilitation (coalition internationale, soutien politique à l'étranger).

Comment maintenir les politiques pilotes à long terme ?

Indépendamment du type de conception dans lequel s'inscrit une politique pilote, ses perspectives d'adoption restent vulnérables à la dynamique de l'agenda politico-économique, à la « valse » des gouvernements et à la rotation des dirigeants au sein de l'agence initiatrice. Si, dans le cas de l'autonomie des ressources, une solution pilotée peut être exemptée de la nécessité de préparer le terrain à son adoption permanente, sa viabilité à long terme exige que des protections supplémentaires soient élaborées à l'avance.

► Assurer un engagement durable ou créer le coût de résiliation comme substitut.

Dans la pratique, lorsque l'engagement politique tend à être limité par la durée d'un cycle électoral, l'élaboration de mesures qui augmentent le coût de résiliation des politiques pourrait être une alternative plus réaliste pour soutenir la politique à l'échelle. Les solutions suivantes en sont des exemples :

- légitimation « de bout en bout » de la politique, couvrant tous les éléments de sa mise en œuvre et de son application : le coût associé à la procédure requise pour « défaire » l'ancrage juridique de l'outil complique son abandon par les gouvernements qui ne le soutiennent pas

- création d'une demande durable pour la politique en investissant dans la qualité de sa mise en œuvre (renforcement des capacités institutionnelles, managériales et techniques)

- maintien de la visibilité de la politique dans les médias afin de rendre sa suspension politiquement inacceptable.

Limitations

Les conclusions de ce chapitre sont de nature inductive et sont donc soumises à au moins trois limitations : la taille limitée de l'échantillon de l'étude de cas (N=3), la diversité contextuelle des programmes comparés (les programmes sont mis en œuvre dans trois pays différents, c'est-à-dire que leurs contextes politiques, institutionnels et socioéconomiques varient considérablement), et les limites de la collecte de données (manque d'accès à certaines personnes interrogées, décalage entre

les entretiens et le programme (Kinofelis)). Par conséquent, les recommandations ci-dessus ne prétendent pas être universellement applicables et peuvent manquer de pertinence dans certains contextes.

D'autres études empiriques – à la fois dans le domaine des politiques de l'emploi et au-delà – seraient utiles pour vérifier et affiner les résultats.

Bibliographie

Bailey, S., Checkland, K. & Hodgson, D. (2019). « Pilots as projects : policymaking in a state of exception », *in* Hodgson, D., Fred, M., Bailey, S. & Hall, P. (eds.), *The projectification of the public sector*. Abingdon on Thames, Royaume-Uni : Routledge, p. 130-148.

Bédécarrats, F. *et al.* (2019). « All that glitters is not gold : the political economy of randomized evaluations in development », *Development and Change*, *International Institute of Social Studies*, 50(3), p. 735-762.

Bravo-Biosc, A. (2020). « Experimental innovation policy », *Innovation Policy and the Economy*, 20, p. 191-232.

Cairney, P., Oliver, K. (2017). « Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? », *Health Res Policy Sys*, 15, 35.

Checkland, K. *et al.* (2021). « Success' in policy piloting: Process, programs, and politics », *Public Administration*, 101(2), p. 463-480.

Corduneanu-Huci, C., Dorsch, M., Maarek, P. (2021). « The politics of experimentation: Political competition and randomized controlled trials », *Journal of Comparative Economics*, 49(1), p. 1-21.

Cote-Boileau, É. *et al.* (2019). « The unpredictable journeys of spreading, sustaining and scaling healthcare innovations: a scoping review », *Health Res Policy Syst,* 17(1), 84. Decentralized Wastewater Treatment Systems in Indonesia. Rapport. Première édition. WSP (Banque mondiale), Indonésie.

Ettelt, S., Mays, N., Allen, P. (2015). « 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), p. 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), p. 7-12.

Guerineau, M., Jacob, F., Kleszczowski, J. (2024). « Codesign in Action: Design Principles to Successfully Manage Transformative Social Innovation », *IEEE Transactions on Engineering Management*, 71, p. 12037-12052.

Hughes, S., Yordi, S., Besco, L. (2020). « The role of pilot projects in urban climate change policy innovation », *Policy Studies Journal*, 48(2), p. 271-297.

Huiteman, D. *et al.* (2018). « Policy experiments: core concepts, political dynamics and impacts », *Policy Sciences*, 51 (2), p. 143-159.

ILO (2018). « Towards a stronger Kinofelis: Recommendations for the strengthening of future phases », *Rapport*. Commission européenne, Genève.

Koon, A. *et al.* (2020). « A scoping review of the uses and institutionalisation of knowledge for health policy in low- and middle-income countries », *Health Res Policy Sys*, 18, 7.

Kuchenmüller, T. *et al.* (2022). « Domains and processes for institutionalizing evidence-informed health policy-making: a critical interpretive synthesis », *Health Res Policy Sys* , 20, 27.

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), p. 281-297.

Markowitsch, J., Scharle A. (2024). « Towards zero long-term unemployment in the EU: Job guarantees and other innovative approaches », *Social Innovation Initiative*, Rapport pour la Commission européenne.

Marston, G., Watts, R. (2003). « Tampering With the Evidence: A Critical Appraisal of Evidence-Based Policymaking », *The Drawing Board: An Australian Review of Public Affairs*.

Mauw, T., Smith, S., Torrens, J. (2023), « Sustainability transitions in Los Angeles' water system: the ambivalent role of incumbents in urban experimentation », *Journal of Environmental Policy and Planning*, 25(4), p. 368-385.

McFadgen, B., Huitema, D. (2017a). « Stimulating learning through policy experimentation: a multi-case analysis of how design influences policy learning outcomes in experiments for climate adaptation », *Water*, 9(9).

McFadgen, B., Huitema, D. (2017b). « 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), p. 1765-1784.

McFadgen, B., Huitema, D. (2018). « Experimentation at the interface of science and policy: A multi-case analysis of how policy experiments influence political decision-makers », *Policy Sciences*, 51, p. 161-187.

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

Neuwinger, M. (2022). « The revolution will not be randomized: Universal basic income, randomized controlled trials, and "evidence-based" social policy », *Global Social Policy*, 22(1), p. 27-45.

Raudla, R. *et al.* (2024). « It is about time! Exploring the clashing timeframes of politics and public policy experiments », *Perspectives on Public Management and Governance,* 7(4), p. 137–148.

Reymond, P., Chandragiri, R., Ulrich, L. (2020). « Governance Arrangements for the Scaling Up of Small-Scale Wastewater Treatment and Reuse Systems – Lessons From India », *Frontiers in Environmental Science*.

Schutter, O. de (2023). « The employment guarantee as a tool in the fight against poverty », Urapport de l'ONU sur l'extrême pauvreté et les droits de l'homme.

Spicer, N. *et al.* (2014). « Scaling-up is a craft not a science: Catalysing scale-up of health innovations in Ethiopia, India and Nigeria », *Soc Sci Med*.

Stoker, G. (2010), « Exploring the promise of experimentation in political science: Micro-Foundational Insights and Policy Relevance », *Political Studies*, 58(2), p. 300-319.

Strassheim, H., Kettunen, P. (2014). « When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms », *Evidence and Policy*, 10(2), p. 259–277.

Tcherneva, P. (2018). « The Job Guarantee: Design, Jobs, and Implementation », Levy Economics Institute, Working Papers Series No. 902.

Tcherneva, P., Lalucq, A. (2022). « A Job Guarantee for Europe », *Policy Brief,* 20 p. *The Foundation for European Progressive Studies*.

Torrens, J., Johnstone, P., Schot, J. (2018). « Unpacking the formation of favourable environments for urban experimentation: The case of the Bristol energy scene », *Sustainability*, 10(3), p. 879.

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

Vreugdenhil, H. *et al.* (2009). « The next step in policy transitions: Diffusion of pilot project », IRSPM XIII. Éditeur / Carsten Greve. Copenhague : IRSPM, p. 1-25.

Annexe 1. Choix de conception et implications pour l'accroissement d'échelle : comparaison d'études de cas de différentes conceptions. Résumé

<table>
<tr><th>Pilote
Dimension de conception</th><th>TZCLD, limite (boundary)</th><th>MAGMA, technocratique (technocratic)</th><th>Kinofelis, plaidoyer (advocacy)</th></tr>
<tr><td>Autonomie des ressources de l'initiateur</td><td>La dépendance à l'égard de l'État a obligé à concevoir des adaptations qui ont finalement contribué à la croissance du pilote dans un contexte de priorités politiques changeantes</td><td>L'autonomie limitée [suffit pour lancer les pilotes, mais pas pour les mettre à l'échelle] a laissé le programme « pieds nus » face à un agenda politique défavorable [conservateurs], ce qui a conduit à son arrêt après une réussite avérée</td><td>L'autonomie complète a permis un accroissement d'échelle nationale avant la fin de l'évaluation du pilote, mais a rendu le programme vulnérable aux changements de gouvernement ; cependant, le « coût de résiliation » proactif imposé par l'initiateur a atténué la vulnérabilité, permettant de poursuivre les inscriptions limitées « politiquement inacceptables » du programme</td></tr>
<tr><td rowspan="2">Architecture du réseau</td><td rowspan="2">Ambigüité de l'ouverture et de la diversité politique : l'accroissement d'échelle nécessite de concilier les intérêts des parties prenantes de gauche et de droite, mais la prise de décision collaborative en fait une opportunité stratégique pour accroître la résilience aux changements d'agenda politique
Le coût élevé de l'adoption à l'échelle remet en question la faisabilité d'un accroissement d'échelle nationale</td><td>Ambigüité de la neutralité politique : a renforcé la crédibilité des données probantes, mais a laissé le pilote sans leviers pour le plaidoyer local [nécessitant le développement de tels leviers « de l'intérieur »]</td><td>Ambigüité des institutions existantes : facilite l'accroissement d'échelle, mais peut comporter un risque de corruption/patronage [qui a dû être traité en premier lieu]
La prise de décision autoritaire ne favorise pas la création d'un engagement politique durable</td></tr>
<tr><td colspan="2">L'utilisation des liens existants et des infrastructures institutionnelles existantes simplifie l'accroissement d'échelle des pilotes politiques</td></tr>
<tr><td rowspan="2">Production et diffusion de connaissances</td><td>La polyvalence de l'évaluation interne facilite l'extensibilité du programme, mais les données probantes produites ne sont pas perçues comme crédibles et peuvent être dévalorisées par l'État
La coexistence d'évaluations internes et d'évaluations « assignées » fait la part belle aux spéculations politiques sur les résultats du pilote</td><td>La crédibilité d'une évaluation rigoureuse agit comme un « multiplicateur d'impact » [les données probantes influencent l'élaboration des politiques à l'étranger], mais ne protège pas contre la critique politique intérieure</td><td>L'évaluation « pour optimiser » plutôt que « pour tester » a garanti un accroissement d'échelle efficace, mais a suscité des doutes quant à la « base probante » de l'élaboration de politiques fondées sur des données probantes dans le cadre de l'autonomie des ressources</td></tr>
<tr><td colspan="3">La diffusion en tant que catalyseur d'accroissement d'échelle : augmente les chances des pilotes d'être adoptés au niveau national (sensibilisation sociale en tant que « coût de résiliation »), « débloque » des scénarios alternatifs d'accroissement d'échelle (extension du pilotage à l'étranger, pilotage au niveau supranational), crée d'autres leviers de facilitation (soutien politique à l'étranger, coalition internationale, données probantes collectives)</td></tr>
</table>

Table des matières

Partie 2 • 71
Le territoire : enjeux, ressources, valeurs

Partie 3 • 131
Garantie d'emploi territorialisée et politiques publiques

Partie 4 • 185
Regards sur l'institutionnalisation de TZCLD

Composition : Le Bord de L'eau éditions

Cet ouvrage a été achevé d'imprimer en juillet 2025
pour le compte des éditions Le Bord de l'eau
par Présence graphique (37260 Monts).

Dépôt légal : Septembre 2025

Imprimé en France

This paper is an English-language original of Chapter III "Choix de conception et perspectives de passage à l'échelle dans le pilotage des politiques: exploration des relations dans le domaine des politiques de création d'emplois" in: Duverger, T. et al. (eds). *Territoires zéro chômeur de longue durée: de l'expérimentation à l'institutionnalisation*. Le Bord de L'eau, 45–72 (translated and published in French).

Design Choices and Scaling Prospects in Policy Piloting: Exploring the Relationship in the Domain of Job Creation

Nastassia Harbuzova
Doctoral Researcher, Tallinn University of Technology

Abstract

Innovative job-creation schemes are gaining momentum. In light of the ESF+ call for piloting such (as announced in April 2024), one of the key puzzles facing potential initiators is how to design the pilots: what design choices should be incorporated to maximise the chances of the project for being scaled-up in case of success? While the existing research indicates that the way policy pilots are designed affects their scaling prospects, there is no systemic understanding of how: the evidence is limited, mixed, and scattered across policy domains.

To help fill the gap, this study offers an integrated analysis of implications that the design of policy pilots may have for their scaling prospects. Three novel employment programs that fit into different design types – boundary Zero Long-Term Unemployment Territories, TZCLD (France), technocratic MAGMA Job Guarantee (Austria), and advocacy Kinofelis Ergasia (Greece) – are compared along three design-dimensions – resource autonomy of initiator, network architecture, and knowledge production and dissemination – to study how the differences in their design have affected their adoption trajectories.

The results demonstrate that while individual design choices are invariably ambiguous in their implications, the "comparative advantage" of each program is associated with a different design dimension: resource autonomy for advocacy Kinofelis, network architecture for boundary TZCLD, and knowledge production for technocratic MAGMA.

Based on the lessons learnt from the analysis, the study concludes with hands-on recommendations on "how to design better" – hoping to inform further pilot launches, both in the domain of job-creation policies and beyond.

1. Introduction

The growing recognition of policy experimentation as a strategy for dealing with modern uncertainties has fuelled the calls for policy piloting across various policy domains (Checkland et al. 2021; Neuwinger 2022; Voß&Simons 2018). Policy pilots exemplify one of the core types of policy experimentation (along with design experiments, experimental legislation, and regulatory sandboxes) and can be broadly defined as time- and scope-limited interventions intended to test novel policy solutions and use the data to decide on their future (Bailey et al. 2019; Raudla et al. 2024). According to the rational understanding of policy experimentation that still dominates in the literature (e.g., according to Huitema et al. 2018), the process of piloting is attractively straightforward and follows a linear logic: in case experimental evidence proves a pilot policy to be successful, the policy is expected to be scaled-up – i.e., extended in time, space, or both (Bravo-Biosca 2020; Marston&Watts 2003; Stoker 2010).

However, what the rationale expects policy pilots to be often run counter what the real-world policymaking shows they are. A growing number of empirical studies indicates that many successful policy pilots are never scaled up (Koon et al. 2020; Mauw et al. 2023; Cote-Boileau et al. 2019). At the same time, the existing scholarship on social innovation demonstrates that whether the novel solutions will be adopted beyond the experimentation is largely affected by

design choices embedded in the latter – in terms of who initiate the pilots, which actors are involved in piloting, how authority and roles are distributed among them, what evidence and how is collected, evaluated and disseminated (e.g., Torrens et al. 2018; Nair&Howlett 2016; Reymond et al. 2020; Spicer et al. 2014). Yet, while there is ample evidence that the outcomes of policy pilots are shaped by their design, this evidence is inconsistent, mixed with other factors (e.g., politics and a broader context of experimentation) and scattered across policy domains.

To help fill in the gap, this chapter offers a systematic empirical analysis of the implications that the design choices embedded in policy pilots may have for their scaling-up trajectories. The research draws on the comparative case study in the domain of job-creation policies. Currently, the domain is not visible in the literature on experimental policymaking – which encourages theoretical contribution that this study hopes to make. Yet, the rapidly growing interest in "job guarantees" observed in Europe and worldwide (Markowitsch&Scharle 2024) increases the chances of the findings being directly applied in practice.

The aim of the study is to explore how the way policy pilots are set-up and governed shape their adoption prospects by comparing the strengths and weaknesses of differently designed policy experiments. The analysis relies on the framework of "governance design" (McFadgen&Huitema 2017a-b) that integrates otherwise scattered design choices into three "ideal" types of policy experimentation: technocratic, boundary and advocacy**.** Based on comparative case-study of three novel job-creation pilots that fit into different design types – boundary Zero Long-Term Unemployment Territories (France), technocratic MAGMA Job Guarantee (Austria), and advocacy Kinofelis Ergasia (Greece) – the chapter first identifies the key dimensions of the design choices that differentiate the programs, and then compares how these choices shape the programs' scaling-up trajectories.

The recommendations distilled from the analysis are offered to assist practitioners in designing pilot policies in the domain of employment policies and beyond.

The chapter is structured as follows. Section two introduces the analytical framework and explains its development. Section three justifies case-study sample and introduces methods for working with data. Section four presents the results of the analysis followed by the conclusions, recommendations and research limitations outline in Section five.

2. Analytical Framework

The literature review reveals that the relations between governance design of policy experiments and their outcomes so far have not been consistently studied. Rather, there are separate studies indicating that:

- the type of the initiating agency shapes the purpose(s) of the launched experiments and the type of evidence they tend to produce (e.g., Ettelt et al. 2015; Lee&Ma 2020; Cairney&Oliver 2017);
- the presence of specific actors (e.g., politicians) among the stakeholders of the experiments influence their scaling-up prospects (e.g., Mauw et al. 2023; Strassheim&Kettunen 2014; Corduneanu-Huci et al. 2021; Nair&Howlett 2016);
- the approaches to evidence collection and evaluation adopted in policy experiments (who, what, how) affect how credible their deliverables are perceived (e.g., Bedecarrats et al. 2019; Checkland et al. 2021; Farrelly 2008);

- the arrangements regulating the relationships between the stakeholders, decision-makers and donors of the experimentation are shown to influence the scaling chances of experimental solutions. Such arrangements include: decision-making models, managerial approaches, communication and dissemination strategies adopted in the experiments (e.g., Reymond et al. 2020; Kuchenmüller et al. 2022; Vreugdenhil 2009; Guerineau et al. 2024).

The only exception found in the literature by the start of the study (2022) was the research of B. McFadgen and D. Huitema (McFadgen&Huitema 2017a-b, 2018) who adopted a systematic approach to the outlined "design-outcomes" relations. The authors developed a framework that integrates otherwise scattered design choices into the holistic notion of "governance design", distinguishing between three "ideal" design types of policy experimentation:

- technocratic – politically neutral, (near-)rational initiatives,
- boundary – collaborative, consensual bottom-up initiatives, and
- advocacy – authoritative top-down initiatives.

The type of policy experiment is identified based on the prevailed characteristics along six groups of design choices, or "rules". The "rules" describe the experimental process in terms of: who is involved and excluded (boundary rules), who does what (position rules), how authority is distributed among stakeholders (choice rules), how decisions are made (aggregation rules), what information is produced and how disseminated (information rules), and how the initiative is funded (pay-off rules).

To date, the framework has been used in quantitative research to explore how the design of policy experiments influences their learning outcomes and ability to enlighten policymakers (McFadgen&Huitema 2017a-b, 2018). This study, by contrast, is interested in taking a disaggregated, "choice-by-choice" approach to discern how individual design choices included in the notion of "governance design" may affect the prospects of policy pilots.

In this regard, based on the literature review and empirical data collected from the case-studies, the original framework has been adapted so that the six groups of design choices with complex underlying criteria have been re-aggregated into the three design dimensions as follows:

- Resource autonomy of initiator: differentiates policy pilots in terms of access their initiators have to authority and funds required to launch, implement and scale-up the pilots;
- Network architecture: differentiates policy pilots in terms of stakeholders involved and the ways they interact with one another;
- Knowledge production and dissemination: differentiates policy pilots based on their approach to evidence collection, evaluation and dissemination.

The new design dimensions altogether provide a clearer and simpler structure for comparative case-study analysis**.** Each dimension incorporates the key individual design choices that differentiate policy pilots of different design types. Figure 1 illustrates the process of framework adaptation, summarising the design choices underlying each dimension.

The adapted framework will be further utilised to: a) justify that the programs selected for the analysis fit in three different design types, b) highlight the differences of their design, and consequently, c) structure the analysis of implications that these differences have for the programs' adoption prospects – Appendix 1 summarises the results.

Figure 1. Adapting the framework to the needs of the "choice-by-choice" approach

Original framework: six blocks of design choices

Groups of design choices	Description [aggregated criteria]
1 Boundary	Who is involved and who is excluded
2 Position	Who does what
3 Choice	How authority is distributed among stakeholders
4 Aggregation	How decisions are made
5 Information	What information is produced
6 Pay-off	How the initiative is funded

Modified framework: three design dimensions

Design dimensions	Design choices included [criteria]
1 Resource autonomy of initiator	Initiating authority Funding sources
2 Network Architecture	Liaisons and institutions Decision-making mode Involvement of politicians
3 Knowledge production and dissemination	Learning strategy Dissemination activity

3. Case-studies and data collection

For empirical analysis, the following programs have been selected:

- MAGMA Job Guarantee (*Modellprojekt Arbeitsplatzgarantie Marienthal*): a small-scale randomized control trial (RCT) initiated by the local public employment service (PES) in close collaboration with academia; underway in Gramatneusiedl, Lower Austria from 2020 to 03.2024; finished according to the planned term of piloting, no scale-up; 150 employed residents;
- Zero Long-term Unemployment Territories, TZCLD (*Territoires Z´ero Ch^omeur de Longue Dur´ee*): an ongoing French experimentation initiated by the association of NGOs in 2016, non-randomized trial; prolonged for the second term (2016-2021, 2022-2026), expanded from 10 to 75 territories: 3 281 of employed (as of 09.2024);
- Kinofelis Ergasia (*ΚΟΙΝΩΦΕΛΗΣ ΕΡΓΑΣΙΑ*, or *Of Common Benefit*): an ongoing regularly implemented national employment program initiated by the Greek Ministry of Labour (MOL); first launched as a pilot in 2011 (non-randomized design experiment); in 2016, the improved version of the program was introduced, scaled-up and adopted permanently; scaled in three steps: from 17 (2011-2013) to 51 (2013-2015) and to all 326 (2016) municipalities; since 2017: downscaled to 25,-45,000 enrolments/year.

The programs make up a strong comparative sample that aligns with the purposes of the study and meets all the criteria set by the author to find the eligible pilots. Specifically, the programs:

a) have been recently launched in the form of policy pilots (or were officially labelled so) – i.e., their outcomes are relevant and comparable in time;

b) pilot a conceptionally similar policy tool

The programs are cited as real-world examples of the Job Guarantee (e.g., Tcherneva&Lalucq 2022) – an active labour market policy aimed to eradicate long-term unemployment by creating job-opportunities that are voluntary, individually-tailored, locally beneficial, non-competitive with traditional economy and provide at least a living wage (e.g., Tcherneva 2018). While each pilot scheme has its specific characteristics, all programs majorly meet this conceptual definition;

c) exemplify different types of governance design

Each program adopts its own way of policy implementation (administration, evaluation, operation) and represents one of three "ideal" design types: MAGMA – technocratic, TZCLD – boundary, Kinofelis – advocacy – as justified with the adapted framework – see Table 1.

To keep the sample comparable and within reach for data collection, the scope of search has been limited to the European Union states. At the time this study began (early 2022), the chosen programs appeared to be the only eligible examples.

Table 1. Comparative overview of case-study designs

Design dimension	**Criteria [design choices]**	**TZCLD**	**MAGMA**	**Kinofelis**
Resource autonomy of initiator	Initiating authority Funding sources	No: - NGOs - public funds [majorly], stakeholders' buy-in [management only]	Limited**:** - local PES - self-funded [disposable budget]	Full: - MOL - funds at disposal [ESF+, national budget]
Network Architecture	Liaisons and institutions	Innovative	Existing, well-established	Existing, well-established
	Decision-making mode	Consensual	Authoritative-consensual	Authoritative
	Involvement of politicians	Strong political facilitation, politically driven project	Politically neutral [originally]	Intrinsically political project [initiator = policymaker]
Knowledge production and dissemination	Learning strategy	Learning to prove success and tailor policy delivery	Learning to test if successful	Learning to optimise policy delivery
	Dissemination activity	Active international promotion in media, outcomes and updates	Active international promotion, mostly outcomes*	Limited promotion during piloting, active – after
DESIGN TYPE CONCLUSION		**BOUNDARY** collaborative bottom-up pilot with a purpose to adopt a consensual solution	**TECHNOCRATIC** near-normative local pilot with a purpose to test a solution	**ADVOCACY** authoritative top-down pilot with a purpose to adopt a solution phase-wise
* the only characteristic that diverges MAGMA from the "ideal" type				

The study draws on qualitative research methods. To collect data, desk research was complemented with 18 semi-structured interviews with the programs' stakeholders (5 for TZCLD, 7 for MAGMA, 6 for Kinofelis). The interviews were conducted online and lasted an hour on average; recorded and transcribed. To preserve anonymity of the interviewees, they will be identified by the country codes (e.g., "Fr1/Au1/Gr1").

For each case-study, the interviews and secondary data were used interactively and iteratively: to cover each other's blanks and check the inferences made. In collecting data, details on the governance design, piloting challenges, scale-up prospects and lessons learnt were put in focus.

While at the international level all three programs are promoted as successful (e.g., Tcherneva & Lalucq 2022), their domestic outcomes are not so definitive and vary. The sample includes an ongoing extended pilot (TZCLD), a pilot that remained unscaled (MAGMA), and a pilot

that became a regular national program (Kinofelis). The next section elucidates to what extent and how the program's governance design has shaped their scaling journeys.

4. Results

4.1 Resource autonomy of initiator

Access to authority and funds required to launch, implement and scale-up policy pilots has proven to be the first principal aspect of the pilots' governance design that differentiates the case-studies and affects their adoption prospects.

No autonomy: TZCLD

In case of grassroots-driven TZCLD, the State is a primary provider of funds and legal permissions needed to maintain and grow the pilot. Consequently, the project's prospects depend on the State's priorities in resource allocation (Fr2) and remain vulnerable to a range of specific threads as follows:

- funding cuts and interruptions including the State's reluctance to increase the funds proportionally to the program expansion – which requires stakeholders to be vigilant and able to mobilise and advocate for the program's needs. Among the recent highlights, the interviewees noted the need to claim an adequate increase in the experimental budget for 2024 (satisfied) and the risks of funding cuts due to the Paris Olympics (Fr5, group discussion);

- resource tension linked to the uncertainty of functional overlap between the program and the newly reformed PES-system (underway by the State in parallel with the experimentation). The reform makes the stakeholders to envision how TZCLD will work within the renewed institutional framework and prove its compatibility with such to continue the project in its current or adopted form (Fr1, Fr4).

- issues of "disempowerment" linked to the reluctant reallocation of public resources from national- to local-level authorities (Fr4). An increase in autonomy of local actors involved in piloting potentially curtails the benefits of the national-level policymakers (including cuts in disposable funds and reduced control over experimental territories) – which may leave the program unsupported by the key policymakers.

- status-quo challenge, or resource competition with conventional public policies in the domain. The "stickiness" of existing practices is cited as a potential scaling hurdle (Fr2-4).

At the same time, the lack of resource autonomy and the very awareness of such a "vulnerability by default" encouraged the initiators to develop effective adaptations for growing the pilot resilient to potentially changing priorities of the decision-maker – including the levers of legal support (launching the pilot through Experimental Laws), political facilitation and capacity building (anticipatory approach, long-term vision, agility, effective communication).

Limited autonomy: MAGMA

The example of MAGMA illustrates how resource autonomy sufficient to launch but not to scale-up the pilot may leave it in a quandary where the lack of scaling strategy collides with an unwelcome political agenda.

The local PES initiating MAGMA is authorised to pilot innovations without asking permission from the Ministry or Federal PES as long as the cost remains within its disposable budget (Au7). At first sight, such autonomy-mode could place MAGMA in a more advantageous position compared to TZCLD as some room for politically neutral decisions is expected: e.g.,

one could expect the program to be prolonged within the authorised jurisdiction in case of success. However, the case demonstrates that the chances for politically neutral scale-up in practice are very limited.

On the one hand, scaling-up turns political once external buy-in is required. Although MAGMA proved successful, the cost of its permanent roll-out appeared to go beyond the PES' budget (Au7, Au4). I.e., an external buy-in was required already to prolong the program on a similar scale.

In such settings, the unfavourable agenda due to Conservative government and a low unemployment rate in the country proved to impede any form of the program's scale-up beyond the agreed pilot term (Au3, Au5-6).

On the other hand, the program's scale-up is also contingent on the leadership turnover in the initiating agency. The risk of not aligning with the new executive's priorities was cited as a potential challenge for moving MAGMA forward along with the budgetary constraint (Au7).

Besides, the limited resource autonomy of the initiating agency imposed the following constraints on the pilot prospects:

- lack of scaling strategy and adaptations. The capacity of the initiator to launch the pilot in case of MAGMA did not motivate developing adaptations to political uncertainty. Specifically, it has not been thought in advance how to scale-up the program in case of success (Au1-2, Au4, Au7), nor has the scale-up option been secured as legitimate – leaving the program "barefoot" against adverse political agenda;
- concerns about external validity of the pilot outcomes: the size of the budget available for experimentation was cited among reasons for the small size of the pilot (150 residents) (Au7) which ultimately risks limiting the program's prospects. While the produced evidence was highlighted as internally valid (Au1-2), the stakeholders doubted its generalizability (Au1) saying that a better contextual awareness should be developed before seeking to expand the program to the region or nationwide (Au4).

Full autonomy: Kinofelis

Initiated by Greek Ministry of Labour, Kinofelis took advantage of the "empowered" policy commitment. The resource autonomy available to the initiator secured the program's nationwide adoption after the piloting yet proved to be a short-term privilege contingent on a government turnover. The short-termism of political commitment is highlighted as a principal challenge for maintaining Kinofelis under successive governments (Gr1, Gr4-5). After the government changed from radical left to centre-right, the program was dramatically downscaled and deteriorated in quality (Gr4).

Nevertheless, despite the lack of political support for the next two cycles of right-wing government, the program was not terminated due to the high political cost imposed by the pilot leader on such a scenario. A proper legitimation of the program at scale and a stable demand from the population created through investments in optimising policy delivery were cited as the reasons why Kinofelis continues enrolments (Gr4).

4.2 Network Architecture

Network architecture of policy pilots – i.e., the ways of interaction established between the stakeholders involved in the piloting as well as the characteristics of the stakeholders themselves – is the second identified unit of governance design that is important to have in the analysis. The three aspects of network architecture that proved to differentiate the case-studies and influence their prospects will be considered further: type of institutions and ties involved

in the policy piloting process, decision-making mode adopted in the pilot, and presence of politicians among its stakeholders.

Innovative vs existing institutions: TZCLD vs MAGMA and Kinofelis

Whether a pilot program relies on existing ties and institutions or requires creating additional ones proves to affect its scaling cost. The cost of program adoption at scale is highlighted as a principal obstacle for scaling TZCLD (Fr2-4) where both implementation and management of the program rely on the novel entities created by the project. Accordingly scaling the program requires agreeing on how to legitimize these innovative structures (Fr1, Fr2), how to integrate them into the system of existing institutions (Fr2, Fr4), and how to operate them at scale (Fr2-3). Therefore, a "truncated" adoption of the program becomes a likely scenario (Fr4).

By contrast, MAGMA and Kinofelis have a cost-advantage in scalability since both have adopted well-established ties and facilities to implement the programs (Gr6). In both cases, local PES offices are the key program executors: the operation of MAGMA was based on horizontal collaboration between the AMS Niederösterreich (initiating PES) and its regular partner it.works – a non-profit provider of labour market services (responsible for counselling and training participants); Kinofelis, in turn, relies on vertical interaction between the MOL (defines the terms and refunds the wage-costs of created employment) and the municipal PES offices that directly create and administer employment projects.

Simultaneously, however, the case of Kinofelis demonstrates that effective scaling-up may require changes in existing institutional facilities as they may involve risks of corruption and patronage. Specifically, to protect the program from "political capture", the initiator had to replace the on-site employment mediation of municipal trade-union centres with a transparent electronic application process (Gr1).

Consensual vs authoritative decision-making: TZCLD vs MAGMA and Kinofelis

The case of TZCLD demonstrates how the inclusive and collaborative decision-making adopted in the pilot encourages its scale-up. The consensual decision-making mode where the pilot's stakeholders of different roles and political affiliations have access to further program development has turned political diversity of TZCLD stakeholders into the program's strategic advantage enabling it to scale-up across politically heterogeneous – internally and to each other – French territories. In that process, initially associated with the left-wing politics, the program is neutralizing its political affiliation by demonstrating that left and right can "make the project together" (Fr4).

Along with it, the network where the grassroots and the managing body of the experimentation (including the leaders of implementation and research) can directly interact with each other (Fr1, Fr4-5) fosters a sense of shared policy commitment among stakeholders (Fr1). Such an attitude empowers the pilot to mobilise a coalition of support "from within" and "on-the-go" – which adds to its resilience and helps in accessing resources to sustain and expand the project into new territories.

In Magma and Kinofelis, the relations among stakeholders are "subject to protocol" and the decision-making authority is either distributed to functional nodes (MAGMA) or highly concentrated in the hands of the leader (Kinofelis). In such settings, a sense of shared commitment to the policy has no incentive to evolve among stakeholders leading the programs to be seen as "another short-term project" (Au4**,** Gr2).

Involvement of politicians: MAGMA vs TZCLD

The ambiguity of implications that political involvement may have for the prospects of resource-dependent pilots is made explicit in the examples of MAGMA and TZCLD. While

both acknowledge the importance of politics for their scaling, agreeing that whether the piloted policy proven successful eventually becomes a permanent tool is a political decision (Fr2-5, Au1-6, Gr1, Gr4-6), the need to involve politicians in experiments is regarded as an opportunity in case of TZCLD and as a challenge in case of MAGMA – depending on the vantage point: evidence credibility or policy facilitation.

In TZCLD, political facilitation is called an essential lever that has enabled to launch the experimentation and helps it access and mobilize the resources needed (Fr1, Fr5).

In MAGMA, the non-involvement of politicians is cited as a prerequisite for the program's launch and its key advantage credibility-wise (Au7). However in practice, it proves to be a two-sided coin: while ensuring that the evidence has no political affiliation was a principal point for the initiator (Au7), his own left-wing advocacy – in light of his transition from a senior civil servant to the chairman of social democrats (centre-left party) in Lower Austria – is now becoming a lever to continue the program after piloting, without government support and resource autonomy needed for politically neutral scale-up (Au4, Au7).

4.3 Knowledge production and dissemination

The learning strategies that policy pilots adopt in terms of what knowledge to produce and how is the third principal unit of analysis that poses different scaling opportunities and challenges for the studied programs. While policy evaluation was conducted within each program, the aspects put in focus and the methods applied prove to be different – demonstrating that the choice of learning strategy is a normative question contingent on the piloting purpose.

Learning to optimise policy delivery: Kinofelis

For Kinofelis, the purpose behind piloting the scheme was to optimize it before adopting nationwide (Gr1-2, Gr4-5). While labelled as a "pilot" (e.g., ILO 2018), the process was more akin to policy prototyping as the scale-up had already been decided based on personal commitment of the initiator and evidence from prior research (Gr1).

Consequently, the evaluation during the "pilot phases" was not focused on collecting evidence but on identifying bottlenecks of the scheme to fix them (Gr1-2, Gr4-5). Such a hands-on approach allowed to build the municipal capacities needed for effective program delivery at scale and therefore, helped sustain it in the long run by creating stable demand from the Greek population (Gr4).

Learning to prove success and ensure effective policy delivery: TZCLD

The primary motives for launching TZCLD were to demonstrate that the scheme works, iterate it according to the needs of diverse French territories and ultimately achieve its legal adoption nationwide (Fr1-2, Fr5). Similarly to Kinofelis, the initiators of the pilot foresaw its success already at launch, yet in this case, the success had to be proven to the State (Fr1).

To accommodate both demands – to prove and iterate the scheme – two simultaneous evaluations are conducted in the project. On the one hand, the scheme is evaluated "from within" by the pilot's stakeholders. The initiators adopted a broad approach to collecting evidence: multiple types of knowledge (including scientific proofs, hands-on learnings, grassroots-feedback) are generated and recognised as equally valuable (Fr1, Fr4). Although this evaluation proved to promote the "scalability" of the scheme elucidating how it should be further developed to operate effectively across diverse territories (Fr1), the evidence it produces does not suffice to convince the State (Fr2-4).

On the other hand and relatedly, the pilot is evaluated by the special Committee regulated within the Experimental Laws enacting TZCLD (2016, 2022). This evaluation, by contrast,

reflects the priorities of the State and consequently focuses on collecting scientific evidence and analysing the project's cost-benefit ratio (Fr3).

As a result, TZCLD faces the challenge of conflicting evidence when the outcomes of internal evaluation contradict the conclusions made by the assigned evaluator, so that the success established in the pilot risks being devalued by the State (Fr4). Besides, while originally the Committee was conceived as a solution against dependent evaluation, its own neutrality is a subject of debate. As interviewees explained, the State has room for "tweaking" the evaluation outcomes by choosing what effects the Committee should and should not include in the program's cost-benefit ratio (Fr4). Hence, the solution is likely to shift rather than block the opportunity for political interest to decide on the pilot's future.

Learning to test if successful: MAGMA

In MAGMA, the primary purpose of piloting the scheme was to genuinely test it and produce knowledge that would be eligible for informing potential policy decisions (Au2, Au7). In that respect, multiple efforts were made to ensure high credibility of experimental evidence, including the adoption of rigorous, "gold-standard" research design to evaluate the scheme (pre-registered RCT), involvement of reputable evaluators (academics from Oxford- and Vienna Universities) and maintenance of the pilot's political neutrality – so that politicians could not affect the outcomes of evaluation (Au4). As an interviewee further elaborated, such a comprehensive approach allowed the pilot to influence international policymaking despite its own small-scale and low political appeal at the local level (Au4). In particular, the success of MAGMA became instrumental in shaping the argument for United Nations report on the job guarantee (Au2, Schutter 2023) and launching the call for job-creation pilots by the European Social Fund Plus (Markowitsch&Scharle 2024).

At the same time, the program was exposed to local political critique from the point of financial outcomes it generated – which were not included in the experimental evaluation (Au1, Au3-4). As an interviewee explained, a freedom of normativity upon "what should be evaluated" left room for the Conservative government to devalue the pilot's success and consequently, it was not a surprise for MAGMA stakeholders that the program proven successful would remain unscaled (Au 4). The case, therefore, suggests that the ability of experimental evidence "to convince" is likely to be an intrinsically political issue: if empowered policymakers are not in favour of the pilot tool, hardly it could be recognised as successful, irrespective of the rigour and approach to its evaluation.

Dissemination activity

Communicating pilot policy success has proved to be an effective "impact multiplier" that enables the pilots to compensate for their limited scale and improve their adoption prospects when domestic political support and resource autonomy are insufficient to continue implementing them otherwise. The case-studies reveal three ways of how.

First, active promotion of pilot policy success may increase its chances for local adoption by adding to the cost of termination. A raise in social awareness about the ongoing experimentation makes its curtailment politically unpalatable for the State (Fr4, Au7).

Second, international dissemination may unlock alternative scaling scenarios for the policy. E.g., despite the ambiguous political support within France, TZCLD has spurred piloting of similar schemes across Europe, including the initiatives in Belgium and Italy (Fr4).

Similarly, international promotion of the programs has fuelled the ESF+ call for piloting job-creation schemes and informed the ambition of the European Job Guarantee (Tcherneva&Lalucq 2022) – which in both cases can be regarded as a "scale-up transfer".

Third, pilot promotion may be a way to devise alternative facilitation levers: e.g., by acquiring political support abroad or reinforcing political argument by building international coalition (such as European Job Guarantee Coalition promoted by TZCLD) (Au3, Gr1, Fr1, Fr4).

5. Conclusions

This study has conducted a systemic analysis of the implications that the way policy pilots are designed may have for their scaling prospects in the domain of employment policies. The conclusions are organised in two blocks.

5.1 Design type and "comparative advantage" of policy pilots

While individual design choices made when initiating, implementing and evaluating the programs are invariably ambiguous in their implications, the program's "comparative advantage" proves to be associated with different design dimensions.

Advocacy pilots may take advantage from the resource autonomy granted to their initiators in the short run. In case of Kinofelis, access to authority and funds needed to scale-up the program allowed the initiator not only to adopt it nationwide but also to impose a high political cost of terminating the program in the absence of political support from subsequent governments. The investments made to building municipal capacities for an effective program delivery at scale has ensured a sustainably high demand for Kinofelis from the local population. Coupled with a proper legitimisation, the program ended up being too well-entrenched in the system of conventional employment services to terminate it.

Boundary pilots may benefit from a highly collaborative and inclusive network of stakeholders, and consensual decision-making. In case of TZCLD, such settings facilitate the extension of the program across politically diverse territories and grow resilience to changes in political agenda by winning political support from both right and left. The sense of shared policy commitment that such a network incentivises, also contributes to building resilience of the project.

Pilots with a predominantly technocratic design are likely to get credit for producing scientifically rigorous and politically neutral evidence that can later be turned into a political leverage for promoting the pilot internationally and compensating for its limited scale – as the case of MAGMA elucidates.

5.2 Two dimensions of scaling-up: design adaptations for better prospects

At the same time, the analysis demonstrates that in the long run, policy scaling remains a challenge for all programs irrespective of the design type they relate to. I.e., neither of the design types is "stronger" or "weaker" than the other: e.g., while the successful advocacy pilots may enjoy the protection of the "empowered policy commitment" against being terminated, the maintenance of this policy against government cycles raises concerns, whereas a "vulnerable by default" boundary pilots may, on the contrary, appear to be the most resilient in the long run.

In that light, it is essential to bear in mind that scaling a pilot policy is a process with two dimensions - time and scope, - and the issues of both should be well-considered by initiators while designing the pilot:

- the issues of policy extension, or how to adopt a pilot policy at scale (scope-dimension);
- the issues of policy maintenance, or how to sustain the policy in the long-run (time-dimension).

The lessons learnt from the analysis suggest the following design adaptations.

How to adopt pilot policies at scale?

● Enact policy pilots through law and legitimate scaling-up as a feasible outcome of experimentation

Adopting pilots through law – so that the terms and outcomes of the experimentation are specified, – reduces the risk that the initiative proven successful will be terminated for political reasons.

● Anticipate "scalability" issues – and reduce the cost of policy adoption by:

- harnessing existing institutional facilities for operating the policy: scaling projects that relies on well-established ties and entities costs less. Yet, the risk of such to be corrupted is to be addressed as well;
- including implementation dimension (practical knowledge) and grassroots-feedback (lay knowledge) in the pilot evaluation to learn what iterations in the scheme should be done and what capacities should be built to optimise program delivery at scale.

● Define a proper set-up of political facilitation

In case pilot initiators do not have resource autonomy, facilitation levers are to be well-thought, in terms of:

- source and channels of facilitation: to seek external support or to develop it "from within" (e.g., when a stakeholder turns into a politician);
- narrative of facilitation: identify a proper narrative for policy promotion, especially in terms of its political affiliation (e.g., neutralizing the affiliation or employing it as an advocacy lever).

● Disseminate knowledge produced internationally

Wide dissemination of the pilot's success may increase its chances for domestic adoption (creating termination cost by raising social awareness), unlock alternative scaling-scenarios (expanding piloting abroad, motivating piloting at supra-national level) or devise alternative facilitation levers (international coalition, political support abroad).

How to maintain pilot polices in the long run?

Irrespective of the design type a pilot policy fits in, its adoption prospects remain vulnerable to the dynamics of the polit-economic agenda, government turnover and leadership rotation in the initiating agency. While in case of resource autonomy, a piloted solution may be exempted from the need to pave its way to permanent adoption, its sustainability in the long run requires additional protections to be developed in advance.

● Secure a lasting commitment or create the cost of termination as a substitute.

In practice, where policy commitment tends to be limited by the length of an electoral cycle, developing measures that increase the cost of policy termination could be a more feasible alternative for sustaining the policy at scale. Examples include the following solutions:

- "end-to-end" legitimation of the policy including all elements of its implementation and delivery: the cost associated with the procedure required to "undo" the legal entrenchment of the tool complicates its termination by non-supportive governments;

- creation of a lasting demand for the policy by investing in the quality of its delivery (building proper institutional, managerial, and technical capacities);
- maintaining visibility of the policy in media to make its curtailment politically unpalatable.

5.3 Limitations

The findings of this chapter are inductive in nature and are, therefore, subject to at least three limitations: limited size of case-study sample (N=3), contextual diversity of the programs compared (the programs are implemented in three different countries, i.e., their political, institutional and socio-economic settings considerably vary), and data collection limitations (lack of access to certain interviewees, time-gap between the interviews and the program (Kinofelis)). Consequently, the recommendations above do not claim to be universally applicable and may not work in certain contexts. Further empirical studies, both within and beyond the domain of employment policies, would be useful to verify and refine the findings.

References

1. BAILEY Simon, CHECKLAND Kath, HODGSON Damian, 2019, "Pilots as projects: policymaking in a state of exception". In: Hodgson, D., Fred, M., Bailey, S. & Hall, P. (Eds.) *The projectification of the public sector*. Abingdon on Thames, UK: Routledge, p. 130-148.
2. BÉDÉCARRATS Florent et al., 2019, "All that glitters is not gold: the political economy of randomized evaluations in development", *Development and Change*, *International Institute of Social Studies*, 50(3), p. 735-762.
3. BRAVO-BIOSCA Albert, 2020, "Experimental innovation policy", *Innovation Policy and the Economy*, 20, p. 191-232.
4. CAIRNEY Paul, OLIVER Kathryn, 2017, "Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy?", *Health Res Policy Sys*, 15, 35.
5. CHECKLAND Kath et al.,2021, "'Success' in policy piloting: Process, programs, and politics", *Public Administration*, 101(2), p. 463-480.
6. CORDUNEANU-HUCI Cristina, DORSCH Michael, MAAREK Paul, 2021, "The politics of experimentation: Political competition and randomized controlled trials.", *Journal of Comparative Economics*, 49(1), p. 1-21.
7. COTE-BOILEAU Élizabeth et al., 2019, "The unpredictable journeys of spreading, sustaining and scaling healthcare innovations: a scoping review", *Health Res Policy Syst,* 17(1), 84. Decentralized Wastewater Treatment Systems in Indonesia. Report. first ed. WSP (World Bank), Indonesia.
8. ETTELT Stefanie, MAYS Nicholas, ALLEN Pauline, 2015. "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), p. 319-337.
9. FARRELLY Ross, 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), p. 7-12.
10. GUERINEAU Mathias, JACOB Florence, KLESZCZOWSKI Julien, 2024, "Codesign in Action: Design Principles to Successfully Manage Transformative Social Innovation," *IEEE Transactions on Engineering Management*, 71, p. 12037-12052.
11. HUGHES Sara, YORDI Samer, BESCO Laurel, 2020, "The role of pilot projects in urban climate change policy innovation", *Policy Studies Journal*, 48(2), p. 271-297.
12. HUITEMA Dave et al., 2018, "Policy experiments: core concepts, political dynamics and impacts", *Policy Sciences*, 51 (2), p. 143-159.
13. ILO, 2018, "Towards a stronger Kinofelis: Recommendations for the strengthening of future phases", *Report*. European Commission, Geneva.
14. KOON Adam et al., 2020, "A scoping review of the uses and institutionalisation of knowledge for health policy in low- and middle-income countries", *Health Res Policy Sys*, 18, 7.
15. KUCHENMÜLLER Tanja et al., 2022, "Domains and processes for institutionalizing evidence-informed health policy-making: a critical interpretive synthesis", *Health Res Policy Sys* , 20, 27.

16. LEE Celia, MA Liang, 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), p. 281-297.
17. MARKOWITSCH Jörg, SCHARLE Ágota, 2024, “Towards zero long-term unemployment in the EU: Job guarantees and other innovative approaches”, Social Innovation Initiative, Report for the European Commission.
18. MARSTON Greg, WATTS Rob, 2003, “Tampering With the Evidence: A Critical Appraisal of Evidence-Based Policymaking”, The Drawing Board: *An Australian Review of Public Affairs*.
19. MAUW Tessa, SMITH Shaun, TORRENS Jonas, 2023, “Sustainability transitions in Los Angeles’ water system: the ambivalent role of incumbents in urban experimentation”, *Journal of Environmental* Policy and Planning, 25(4), p. 368–385.
20. MCFADGEN Belinda, HUITEMA Dave, 2017a, “Stimulating learning through policy experimentation: a multi-case analysis of how design influences policy learning outcomes in experiments for climate adaptation”, *Water*, 9(9).
21. MCFADGEN Belinda, HUITEMA Dave, 2017b, “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), p. 1765-1784.
22. MCFADGEN Belinda, HUITEMA Dave, 2018, “Experimentation at the interface of science and policy: A multi-case analysis of how policy experiments influence political decision-makers”, *Policy Sciences*, 51, p. 161-187.
23. NAIR Sreeja, HOWLETT Michael, 2016, “Meaning and power in the design and development of policy experiments”, *Futures*, 76, p. 67-74.
24. NEUWINGER Malte, 2022, “The revolution will not be randomized: Universal basic income, randomized controlled trials, and ‘evidence-based’ social policy”, *Global Social Policy*, 22(1), p. 27-45.
25. RAUDLA Ringa et al., 2024, “It is about time! Exploring the clashing timeframes of politics and public policy experiments”, *Perspectives on Public Management and Governance,* 7(4), p. 137–148.
26. REYMOND Philippe, CHANDRAGIRI Rohit, ULRICH Lukas, 2020, “Governance Arrangements for the Scaling Up of Small-Scale Wastewater Treatment and Reuse Systems – Lessons From India”, *Frontiers in Environmental Science*.
27. SCHUTTER Olivier de, 2023, “The employment guarantee as a tool in the fight against poverty”, UN Report on Extreme Poverty and Human Rights
28. SPICER Neil et al., 2014, “Scaling-up is a craft not a science: Catalysing scale-up of health innovations in Ethiopia, India and Nigeria”, *Soc Sci Med*.
29. STOKER, Gerry, 2010, “Exploring the promise of experimentation in political science: Micro-Foundational Insights and Policy Relevance”, *Political Studies*, 58(2), p. 300-319.
30.
31. STRASSHEIM Holger, KETTUNEN Pekka, 2014, “When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms”, *Evidence and Policy*, 10(2), p. 259–277.
32. TCHERNEVA Pavlina, 2018, “The Job Guarantee: Design, Jobs, and Implementation”, Levy Economics Institute, Working Papers Series No. 902.
33. TCHERNEVA Pavlina, LALUCQ Aurore, 2022, “A Job Guarantee for Europe”, Policy Brief, 20 p. *The Foundation for European Progressive Studies*.
34. TORRENS Jonas, JOHNSTONE Phillip, SCHOT Johan, 2018, “Unpacking the formation of favourable environments for urban experimentation: The case of the Bristol energy scene”, *Sustainability*, 10(3), p. 879.
35. VOß Jan-Peter, SIMONS Arno, 2018, “A Novel Understanding of Experimentation in Governance: Co- producing Innovations between ‘Lab’ and ‘Field’”, *Policy Sciences*, 51(2), p. 213-229.
36. VREUGDENHIL Heleen et al., 2009, “The next step in policy transitions: Diffusion of pilot projects”, IRSPM XIII. editor / Carsten Greve. Kopenhagen: IRSPM, p. 1-25.

Appendix 1. Design choices and implications for scaling-up: comparing case-studies of different designs. Summary

<table>
<tr><th>Pilot Design dimension</th><th>TZCLD, boundary</th><th>MAGMA, technocratic</th><th>Kinofelis, advocacy</th></tr>
<tr><td>Resource autonomy of initiator</td><td>Dependence on the State forced to devise adaptations that ultimately helped growing the pilot under changing polit-priorities</td><td>Limited autonomy [suffices to launch yet not to scale-up pilots] left the program “barefoot” against unfavourable political agenda [Conservatives] leading to its termination after proven successful</td><td>Full autonomy allowed for nationwide scale-up before the pilot-evaluation was completed, yet left the program vulnerable to government-turnover; yet, proactive “termination cost” imposed by initiator alleviated the vulnerability allowing to continue the program’s limited “polit-unpalatable” enrolments</td></tr>
<tr><td rowspan="2">Network Architecture</td><td rowspan="2">Ambiguity of openness and political diversity: scaling-up requires reconciling interests of the Left and Right stakeholders, yet collaborative decision-making turns it into a strategic opportunity for growing resilient to changes in polit-agenda; High adoption cost at scale questions the feasibility of nationwide rollout</td><td>Ambiguity of polit-neutrality: strengthened evidence credibility yet left the pilot without levers for local advocacy [necessitating the development of such “from within”]</td><td>Ambiguity of existing institutions: facilitate scaling-up, yet may involve a risk of corruption/patronage [that had to be addressed first]; Authoritative decision-making does not support the creation of lasting policy commitment</td></tr>
<tr><td colspan="2">Using existing ties and institutional facilities simplifies scaling-up policy pilots</td></tr>
<tr><td rowspan="2">Knowledge production and dissemination</td><td>Versatility of internal evaluation facilitates scalability of the scheme, yet evidence produced is not perceived as credible and can be devalued by the State; co-existence of internal and “assigned” evaluations provides room for polit- speculation over the pilot’s outcomes</td><td>Credibility of rigorous evaluation acts as “impact-multiplier” [evidence influences policymaking abroad] but does not protect from domestic polit-critique</td><td>Evaluation “to optimise” rather than “to test” ensured effective scale-up yet raised doubts about the “evidence-base” of evidence-based policymaking under resource autonomy</td></tr>
<tr><td colspan="3">Dissemination as a scale-up catalyst: increases pilots’ chances for domestic adoption (social awareness as “termination cost”), “unlocks” alternative scaling-scenarios (extending piloting abroad, piloting at supranational level), devises additional facilitation-levers (polit-support abroad, international coalition, collective evidence)</td></tr>
</table>

Curriculum vitae

Personal data

Name:	Nastassia Harbuzova
Date of birth:	20.11.1996
Place of birth:	Minsk, Belarus
Citizenship:	Belarus

Contact data

E-mail:	nastassia.harbuzova@taltech.ee, harbuzova.nastassia@gmail.com

Education

11.2021–2026	Tallinn University of Technology (TalTech), PhD in Public Policy and Innovation
2018–03.2021	Russian State Academy of Intellectual Property (RSAIP), MA (Hons) in Innovation Governance (Thesis: 5/5)
2019–2019	Belarusian State University (BSU), MSc in Mathematical and Instrumental Methods in Economics (Thesis: 9/10; core subject (Econometrics): 10/10)
2014–2018	BSU, BSc in Economic Theory/Policy (Thesis: 10/10)

Language competence

Russian, Belarusian	Native
English	Fluent
Estonian	Basic

Professional employment

11.2021–2025	TalTech, School of Business and Governance, R. Nurkse Dep. of Innovation and Governance, Doctoral Researcher
2018–2021	BSU, Faculty of Economics, Dep. of International Political Economy, Lecturer (Macroeconomics, Economic Theory, Economy of Innovation; History of Economic Thought (Olympiad curriculum))

(Co)-supervised and Opposed Master's Theses (TalTech)

2024 – MA thesis "Challenges and Adaptations of Estonia's Energy Sector", M. Terentjev
2023 – MA thesis "The failure of mobile money as a public policy in Ecuador: An explanatory evaluation", C.O. Echeverria
2022 – MA thesis "Nigeria's Monetary Policy and Economic Development: Experience of the Cashless Society and an Appraisal of the CBDC", A. Tiamiyu
2022 – MA thesis "Implementing Central Bank Digital Currency (CBDC) into the Nation's Monetary System: a Case of Estonia", K. Moor
2022 – MA thesis "Challenges of ICT-enabled Financial Inclusion in Pakistan", Sh. Bizinjo (co-supervised with Prof. R. Raudla)

Honours and Grants (selected)

2025

FEED UK grant supporting participation in the workshop "Where Social Economics Meets Institutional Economics: On Norms, Values and Ethics", University "Lucian Blaga" Sibiu, 08.11.2025

AFEE grant for session chairing at the AHE Conference; King's College London, 18-20.06.2025

"Research Article of the year 2025" from the School of Business and Governance, TalTech, for "The politics of experimental policymaking: the influence of blame avoidance and credit claiming", with Raudla, R. et al.

2024

"Research Article of the year 2024" from the School of Business and Governance, TalTech, for "It is about time! Exploring the clashing timeframes of politics and public policy experiments", with Raudla, R. et al.

"Best Comparative Paper"; NISPAcee-2024 Conference for "The politics of experimental policymaking: the influence of blame avoidance and credit claiming", with Raudla, R. et al.

2023

AFEE grant supporting participation in the workshop week, Washington DC; 08.2023

2022

J. Monnet European Centre of Excellence grant "Economics and Institutions of EU, Module on Economics and Institutions in Europe", Summer Module held at the University of Trento, Italy; 09.2022

Erasmus+ grant "Leadership in Small States": visiting student at the University of Iceland; Center for Small States Studies; 06.2022

2021 and earlier

Silver Prize at the International Olympiad in History of Economic Thought awarded to a co-supervised BSU student team; Financial University, Moscow; 05.2021

3rd Presidential Prize at the XXV Republican Contest of Student and PhD Scientific Works, for BA thesis "Institutions and Institutional Change in the New Economy: the case of the Republic of Belarus"; 2019

Best Thesis at the National Contest of Research Works in Intellectual Property (IP); National Center of IP, Belarus, 2018; relatedly – grant of the Confederation of Rightsholders' Societies of Europe and Asia covering

MA-studies at RSAIP; 2018-2021*****

***** *BSc, MSc, and PhD-studies have been funded by public scholarships*

Publications (selected)

Harbuzova, N. (2026). Why successful policy pilots are (not) scaled up: The case of European employment-creation schemes. *Halduskultuur: The Estonian Journal of Administrative Culture and Digital Governance*, 25(1), 24–49

Harbuzova, N. (2025). Design choices and scaling prospects in policy piloting: exploring the relationship in the domain of job-creation. In: Duverger, T. et al. (eds). *Territoires zéro chômeur de longue durée: de l'expérimentation à l'institutionnalisation*. Le Bord de L'eau, 45–72 (Ch. 3; tr. in French)

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

Raudla, R., Sarapuu, K., Douglas, J. W., Vallistu, J., Harbuzova, N. (2025). Budgeting for innovation: How funding strategies enable or constrain public sector experimentation. *Public Budgeting & Finance*, 45(4), 28–40

Raudla, R., Sarapuu, K., Vallistu, J., Harbuzova, N. (2024). It is about time! Exploring the clashing timeframes of politics and public policy experiments. *Perspectives on Public Management and Governance*, 7(4), 137–148

Raudla, R., Sarapuu, K., Juuse, E., Harbuzova, N., Vallistu, J., Onno, K., 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

Harbuzova, N. (2022). *Macroeconomics: Task Book*. Belarusian State University Press (in Russian) / Гарбузова, А. Д. (2022) *Макроэкономика: практикум*. *Учебно-методическое пособие*. Минск: Белорусский государственный университет

Talks and Presentations (selected, 2023-2025)

Invited Talk “How to design a scalable/scaled-up Job Guarantee: Lessons from European Job Guarantee Interventions”; expert workshop “European Job Guarantee Projects in Practice: Challenges and Opportunities of Public Job Creation in the Fight against Long-Term Unemployment”, Austrian Federal Chamber of Labour, Vienna, 11-12.06.2025

Presentation "How policy pilots' design shapes their adoption prospects: the case of job-creation programs"; Annual Conference of the European Group for Public Administration, Panteion University of Social and Political Sciences; Athens, 05.09.2024

Presentation "Why successful policy pilots are (not) scaled: the case of differently designed ‘job guarantees’"; 26th AHE Conference, University of the West of England; Bristol, 11.07.2024

Invited Talk “Bridging the gap between successful and scaled: the case of Job Guarantee policy pilots”; research workshop of Territoires zéro chômeur de longue durée; Paris (MACIF Headquarter), 21.03.2024

Presentation “Lessons learnt from piloting employment-creation programs”, practitioner workshop “The main promises and challenges of public sector experiments”; workshop for civil servants from Estonia, Latvia and Finland; PRG1125; 13.10.2023

Presentation “Policy piloting as a way for heterodox solutions to become a common practice. Discussing how and when it works”; AFEE week of workshops, Washington DC, 14-18.08.2023

Presentation “Policy Learning from Job Guarantee Pilots: evidence of how the design choices affect the lessons learned”; Society for the Advancement of Socio-Economics Conference, Rio-de Janeiro, 22.07.2023

Elulookirjeldus

Isikuandmed

Nimi: Nastassia Harbuzova
Sünniaeg: 20.11.1996
Sünnikoht: Minsk, Valgevene
Kodakondsus: Valgevene

Kontaktandmed

E-post: nastassia.harbuzova@taltech.ee,
harbuzova.nastassia@gmail.com

Hariduskäik

11.2021–2026	Tallinna Tehnikaülikool, doktorantuur riigivalitsemise ja innovatsiooni erialal
2018–03.2021	Russian State Academy of Intellectual Property (RSAIP), magistrikraad (*Hons*) innovatsioonijuhtimise erialal (magistritöö hinne: 5/5)
2019–2019	Belarusian State University (BSU), magistrikraad erialal matemaatilised ja instrumentaalsed meetodid majandusteaduses (magistritöö hinne: 9/10; põhiaine (ökonomeetria): 10/10)
2014–2018	BSU, bakalaureusekraad erialal majandusteooria / majanduspoliitika (bakalaureusetöö hinne: 10/10)

Keelteoskus

Vene ja valgevene keel	emakeel
Inglise keel	kõrgtase
Eesti keel	algtase

Teenistuskäik

11.2021–2025	TalTech, Majandusteaduskond, Ragnar Nurkse innovatsiooni ja valitsemise instituut, doktorant
2018–2021	BSU, Majandusteaduskond, rahvusvahelise poliitökonoomia instituut, lektor (makroökonoomika, majandusteooria, innovatsiooniökonoomika; majandusteaduse ajalugu)

Kaasjuhendatud või oponeeritud magistritööd (TalTech, Majandusteaduskond)

2024 – Magistritöö „Challenges and Adaptations of Estonia's Energy Sector", M. Terentjev
2023 – Magistritöö „The failure of mobile money as a public policy in Ecuador: An explanatory evaluation", C.O. Echeverria
2022 – Magistritöö „Nigeria's Monetary Policy and Economic Development: Experience of the Cashless Society and an Appraisal of the CBDC", A. Tiamiyu
2022 – Magistritöö „Implementing Central Bank Digital Currency (CBDC) into the Nation's Monetary System: a Case of Estonia", K. Moor

2022 – Magistritöö „Challenges of ICT-enabled Financial Inclusion in Pakistan", Sh. Bizinjo (kaasjuhendaja koos professor R. Raudlaga)

Autasud ja grandid (valitud)

2025

FEED UK grant toetamaks osalemist töötoas „Where Social Economics Meets Institutional Economics: On Norms, Values and Ethics", University "Lucian Blaga" Sibiu, 08.11.2025

AFEE grant toetamaks sessiooni juhatamist konverentsil AHE Conference; King's College London, 18-20.06.2025

„Aasta parim teadusartikkel 2025" TalTech, Majandusteaduskond, teemal „The politics of experimental policymaking: the influence of blame avoidance and credit claiming", Raudla, R. et al.

2024

„Aasta parim teadusartikkel 2024" TalTech, Majandusteaduskond, teemal „It is about time!

Exploring the clashing timeframes of politics and public policy experiments", koos autoritega Raudla, R. et al.

Parim võrdlev teadustöö; NISPAcee-2024 konverents, töö teemal „The politics of experimental policymaking: the influence of blame avoidance and credit claiming", Raudla, R. et al.

2023

AFEE grant toetamaks osalemist töötubade nädalal, Washington DC; 08.2023

2022

J. Monnet European Centre of Excellence grant „Economics and Institutions of EU, Module on Economics and

Institutions in Europe", suvemoodul, mis toimus Trento Ülikoolis Itaalias; 09.2022

Erasmus+ grant „Leadership in Small States": külalisüliõpilane Islandi Ülikoolis, väikeriikide uuringute keskuses; 06.2022

2021 ja varem

Hõbemedal rahvusvahelisel olümpiaadil „International Olympiad in History of Economic Thought", mille sai kaasjuhendatud BSU õpilastiim; Financial University, Moskva; 05.2021

Auhind „3rd Presidential Prize" võistlusel „XXV Republican Contest of Student and PhD Scientific Works", bakalaureusetöö eest teemal „Institutions and Institutional Change in the New Economy: the case of the Republic of Belarus"; 2019

Parim uurimistöö intellektuaalse omandi uurimistööde riiklikul võistlusel „National Contest of Research Works in Intellectual Property (IP)"; National Center of IP, Valgevene, 2018; korduvalt – grant asutuselt Confederation of Rightsholders' Societies of Europe and Asia, mis kattis magistriõpingud RSAIPis; 2018-2021*****

***** *Bakalaureuse-, magistri- ja doktoriõpinguid on rahastanud riiklikud stipendiumid*

Publikatsioonid (valitud)

Harbuzova, N. (2026). „Why successful policy pilots are (not) scaled up: The case of European employment-creation schemes". *Halduskultuur: The Estonian Journal of Administrative Culture and Digital Governance*, 25(1), 24–49

Harbuzova, N. (2025). „Design choices and scaling prospects in policy piloting: exploring the relationship in the domain of job-creation“. Teoses: Duverger, T. et al. (eds). *Territoires zéro chômeur de longue durée: de l'expérimentation à l'institutionnalisation*. Le Bord de L'eau, 45–72 (Ch. 3; tr. prantsuse keeles)

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

Raudla, R., Sarapuu, K., Douglas, J. W., Vallistu, J., Harbuzova, N. (2025). „Budgeting for innovation: How funding strategies enable or constrain public sector experimentation“. *Public Budgeting & Finance*, 45(4), 28–40

Raudla, R., Sarapuu, K., Vallistu, J., Harbuzova, N. (2024). „It is about time! Exploring the clashing timeframes of politics and public policy experiments“. *Perspectives on Public Management and Governance*, 7(4), 137–148

Raudla, R., Sarapuu, K., Juuse, E., Harbuzova, N., Vallistu, J., Onno, K., 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

Harbuzova, N. (2022). *„Macroeconomics: Task Book“*. Belarusian State University Press (vene keeles) / Гарбузова, А. Д. (2022) *Макроэкономика: практикум*. *Учебно-методическое пособие*. Минск: Белорусский государственный университет

Esinemised ja presentatsioonid (valitud, 2023–2025)

Kutsutud esineja teemal „How to design a scalable/scaled-up Job Guarantee: Lessons from European Job Guarantee Interventions”; ekspertide töötuba „European Job Guarantee Projects in Practice: Challenges and Opportunities of Public Job Creation in the Fight against Long-Term Unemployment”, Austrian Federal Chamber of Labour, Viin, 11-12.06.2025

Presentatsioon „How policy pilots' design shapes their adoption prospects: the case of job-creation programs"; Annual Conference of the European Group for Public Administration, Panteion University of Social and Political Sciences; Ateena, 05.09.2024

Presentatsioon „Why successful policy pilots are (not) scaled: the case of differently designed 'job guarantees'"; 26th AHE Conference, University of the West of England; Bristol, 11.07.2024

Kutsutud esineja teemal „Bridging the gap between successful and scaled: the case of Job Guarantee policy pilots”; Territoires zéro chômeur de longue durée uurimistöö töötuba; Pariis (MACIFi peakontoris), 21.03.2024

Presentatsioon „Lessons learnt from piloting employment-creation programs”, praktikute töötuba „The main promises and challenges of public sector experiments”; töötuba Eesti, Läti ja Soome riigiametnikele; PRG1125; 13.10.2023

Presentatsioon „Policy piloting as a way for heterodox solutions to become a common practice. Discussing how and when it works”; AFEE töötubade nädal, Washington DC, 14-18.08.2023

Presentatsioon „Policy Learning from Job Guarantee Pilots: evidence of how the design choices affect the lessons learned”; Society for the Advancement of Socio-Economics Conference, Rio-de Janeiro, 22.07.2023

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-553-4 (PDF)