

EXPLORING THE PATHWAYS: REGULATORY EXPERIMENTS FOR SUSTAINABLE DEVELOPMENT – AN INTERDISCIPLINARY APPROACH

Dierk Bauknecht^{*}, Thore Sören Bischoff^{**}, Kilian Bizer^{***},
Martin Führ^{****}, Peter Gailhofer^{*****}, Dirk Arne Heyen^{*****},
Till Proeger^{*****}, Kaja von der Leyen^{***}

^{*} Öko-Institut (Institute for Applied Ecology); Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

^{**} Corresponding author, Institute for Small Business Economics at the University of Goettingen, Goettingen, Germany

Contact details: Institute for Small Business Economics, Heinrich-Düker-Weg 6, 37073 Goettingen, Germany

^{***} Chair of Economic Policy and SME Research, University of Goettingen, Goettingen, Germany

^{****} University of Applied Sciences, Darmstadt, Germany

^{*****} Öko-Institut (Institute for Applied Ecology), Freiburg, Germany

^{*****} Institute for Small Business Economics at the University of Goettingen, Goettingen, Germany



Abstract

How to cite this paper: Bauknecht, D., Bischoff, T. S., Bizer, K., Führ, M., Gailhofer, P., Heyen, D. A., Proeger, T., & von der Leyen, K. (2020). Exploring the pathways: Regulatory experiments for sustainable development – An interdisciplinary approach. *Journal of Governance & Regulation*, 9(3), 49-71. <http://doi.org/10.22495/jgrv9i3art4>

Copyright © 2020 The Authors

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). <https://creativecommons.org/licenses/by/4.0/>

ISSN Print: 2220-9352

ISSN Online: 2306-6784

Received: 18.06.2020

Accepted: 17.08.2020

JEL Classification: L51, O31, Q58

DOI: 10.22495/jgrv9i3art4

The Sustainable Development Goals (SDGs) adopted by the United Nations provide normative orientation for many national and regional governments as well as actors from industry and other parts of the civil society. There is a growing consensus that the corresponding transformation processes needed – e.g., in the field of production and consumption patterns (SDG 12) – have to be fostered by a corresponding institutional framework. Properly designed experiments that generate a learning system for all actors involved may be an important building block. Based on an interdisciplinary approach, we provide an overview of the various terminologies for experimentation currently discussed in the social sciences, derive common criteria for a broader approach to the concept of “regulatory experimentation” in reflexive governance structures and present a novel conceptual framework for analysing empirical studies of regulatory experiments.

Keywords: Regulation, Regulatory Experimenting, Sustainable Development, Reflexive Governance, Regulatory Learning

Authors' individual contribution: Conceptualization – D.B., K.B., M.F., D.A.H., and T.P.; Methodology – T.S.B., K.v.d.L., and D.B.; Formal Analysis – T.S.B. and K.v.d.L.; Writing – Original Draft – T.S.B., K.v.d.L., D.B., P.G., and D.A.H.; Writing – Review & Editing – D.B., T.S.B., K.v.d.L., K.B., M.F., and T.P.; Visualization – T.S.B.; Supervision – K.B.; Project Administration – K.v.d.L.; Funding Acquisition – D.B., K.B., M.F., D.A.H., and T.P.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

Acknowledgements: The present paper is one result of the research project RERA-GI – Regulatory experiments for the reflexive and adaptive governance of innovation. The research project is funded by the German Federal Ministry of Education and Research.

1. INTRODUCTION

This paper aims to facilitate the inclusion of experimenting for sustainable development in future

governance structures. The core meaning of the term sustainable development is still captured by the 1987 Brundtland Commission definition: “Development that meets the needs of the present without

compromising the ability of future generations to meet their own needs” (WCED, 1987). The United Nations also adopted a very broad approach to sustainable development when defining the seventeen Sustainable Development Goals (SDGs), which take into account the debate on the “planetary boundaries” (Rockström et al., 2009)¹. Under the title “*Transforming our World*”, the Agenda 2030 lists more than 170 action points, most of which are ultimately intertwined². National translations of the global approach such as the German government’s sustainability strategy also follow a broad approach regarding what constitutes sustainable development (German Federal Government, 2016, updated 2017).

From the SDGs, it is clear that the transformation of consumption and production processes requires more than classical technical innovations. In order to bring about a fundamental change towards avoiding emissions and resource-efficient production and consumption, social and organisational innovations are essential since actors with diverging interests have to cooperate along global supply chains. An example of this required cooperation is the European chemicals regulation REACH. It implemented the “no data, no market” principle and thus forced producers to unveil (at least partly) their formerly exclusive knowledge on substance properties, which holds value to downstream users, authorities, academic researchers, and the public, including competitors (Führ & Bizer, 2007). This regulation simultaneously targets “a high level of protection of human health and the environment” as well as “competitiveness and innovation” (REACH, Ch. 1, Art. 1).

An encompassing view on sustainable development means that every political initiative, legislative and regulatory proposal must balance a set of conflicting issues in terms of goal attainment. Although the SDGs can guide choices to be made, many of them can be in conflict and require legislators and regulators to accept trade-offs. The latter has been studied in the literature on policy integration (e.g., Tosun & Lang, 2017). The SDGs themselves recognise this challenge and highlight the importance of governance mechanisms and the cooperation of actors as the solution. SDG 17 on partnerships between governments, the private sector, and civil society emphasises the necessity to “enhance policy coherence for sustainable development (SDG 17.14)”.

Several authors emphasize that sustainable development is not a final state but rather a continuous process (Erdmann, 2005; Meppem & Gill, 1998; Waas, Hugé, Verbruggen, & Wright, 2011). This process is likely to exhibit strong path dependencies and its outcomes that are hard to predict ex-ante. This long and complicated process can benefit from reiterated search and learning phases by industry actors, civil society as well as among administrative executives and politicians involved in drafting legal and regulatory initiatives³.

In this process, experiments in reflexive governance structures can help to guide regulatory choices. Reflexive governance systems use information obtained from experimenting with existing policies and regulations to improve them.

Reflexive governance in general and experimentation, in particular, are a challenging field of study. One reason for this is that experimentation can occur in various forms. An important distinction can be made between experimentation used as a research method or as an instrument in the regulatory practice. In the first case, experiments aim to reveal the causal effects of treatment using randomization and statistical analysis. In the latter case, experiments are used as an approach to governance to test several options when solving societal problems and to develop regulation in practice (Huitema, Jordan, Munaretto, & Hildén, 2018). The literature uses various terminologies for the different forms of experimentation in the social sciences, yet the extent to which they overlap or depict different artefacts has not yet been considered. While case studies of single experiments already exist, a general approach of categorising experiments suitable as an empirical basis for general recommendations for widespread use of the tool is lacking. This, in turn, hampers its further use by policy-makers, regulators, and researchers working on specific governance problems.

This paper provides an interdisciplinary discussion of experimentation as a reflexive governance tool and addresses the following questions:

- Which different concepts of experimenting with regulation exist?
- What are the differences and similarities between the various concepts?
- What could be a common understanding of regulatory experimentation that includes the various concepts?
- How can a comprehensive analytical framework for analysing empirical examples of regulatory experiments look like?

We explicitly anchor our analysis of regulatory experiments in reflexive governance structures with the aim of fostering sustainable development.

The remainder of this paper is structured as follows. Section 2 provides a review of the literature on reflexive governance and experimentation in the social sciences. We argue that sustainable development requires reflexive governance and examines the concepts used for experimentation in the social science literature and introduce our concept of regulatory experimentation. Section 3 derives an analytical framework for analyzing existing regulatory experiments that facilitates systematic data collection, and Section 4 illustrates the use of this framework through a content analysis of 3 empirical cases of regulatory experiments based on publicly available documents. Section 5 discusses practical challenges to experimentation, and Section 6 concludes.

¹ Climate change, biodiversity loss, biogeochemical, ocean acidification, land use, freshwater, ozone depletion, atmospheric aerosols, and chemical pollution.

² Information about the 2030 Agenda for Sustainable Development and the Sustainable Development Goals can be found online: http://ec.europa.eu/environment/sustainable-development/SDGs/index_en.htm

³ In addition, SDG 17.16 reads as follows “Enhance the Global Partnership for Sustainable Development, complemented by multi-stakeholder partnerships that mobilize and share knowledge, expertise, technology, and financial resources, to support the achievement of the Sustainable Development Goals in all countries, in particular developing countries”.

2. LITERATURE REVIEW

2.1. Reflexive governance for sustainable development

2.1.1. Reflexive governance: Why and how?

According to a basic definition of governance building on theories of government and the state, the term refers to public institutions that provide public goods and aggregates society's interest with respect to which public goods are distributed to whom (Gisselquist & Niño-Zarazúa, 2015). Governance has also been seen as encompassing all forms of a collective - or societal - coordination, independent from the actor initiating it (government, business, or other societal stakeholders) and from its form (hierarchical, free market-based, or coordinative), see Benz and Dose (2010) and Mayntz (2004).

According to the latter and broader approach, governance involves a multitude of actors. Legislators have the power to pass laws (typically state secretaries and members of parliament) whereas regulatory bodies conduct diverse activities such as advising legislators, imposing requirements, conditions, or restrictions on the practical implementation of laws as well as enforcing laws and obtaining compliance. For the purpose of this paper, we mostly refer to both policy-makers and regulators jointly. This has two main advantages: first, both the making of laws and their implementation processes shape governance; second, the distinction between legislators and regulators is country-specific and (in reality) often difficult to draw. Most governance interventions considered in this paper involve both changes in legislation and regulation. Civil society - as a third actor - is a heterogeneous group whose actions are influenced by the decisions of policy-makers and regulators. It should, therefore, be in their interest to monitor and adapt laws and regulations to secure progress towards societal goals.

With this latter fact in mind, we take the broader approach to governance as a starting point, but also expand it by focusing on reflexive governance. Reflexive governance is an over-arching concept that encompasses various conceptual and practical approaches to societal problem-handling (Voss, Kemp, & Bauknecht, 2006). The concept builds on the reflexive modernization discourse (Beck, Giddens, & Lash, 1994) and has mainly been developed in the environmental governance literature. A recent overview of the concept and its development can be found in Feindt and Weiland (2018). Reflexive governance understands the interaction between the regulator and its addressees as two-way communication. According to Voss et al. (2006), it implies constantly calling into question the foundations of governance itself, i.e., concepts, practices, and institutions.

Our starting premise is that a responsive and reflexive approach to governance is necessary for sustainable developments in economic and social systems. Sustainable development defined in a broad sense is multi-faceted, which the definition of as many as seventeen SDGs illustrates. A transformation towards sustainability must take into account the linkages between the various

dimensions. The complexity further increases since each dimension comprises many diverse but linked challenges. Policy-makers' and regulators' toolkits to lead societies towards sustainable developments are large and span setting legal standards, incentivising systems such as taxes, subsidies, and tradeable permits, as well as "softer" interventions such as nudges, voluntary schemes, and information campaigns. However, often, one policy or regulatory option is chosen without knowing its impacts on the desired outcome, as ex-ante evaluation is challenging (Greenstone, 2009). Furthermore, the existence of path dependencies in governance (e.g., Briglauer, Vogelsang, & Camarda, 2017; Modell, Jacobs, & Wiesel, 2007) can lead to an ever-increasing number of potentially ineffective policies and regulations, since choosing an inappropriate tool today will increase the chance of inefficient choices in the future.

In contrast, reflexive governance systems produce the knowledge necessary for policies and regulations towards societal sustainability by assessing the intended as well as unintended effects of regulation and integrate this knowledge into future regulation. As such, reflexive governance approaches address the shortcomings of the problem-solving approach which is characterized by blinding out uncertainty, ambivalence, and limited control capabilities. Eliminating real-world complexities in this way typically only leads to unintended side-effects of governance, which then need to be addressed as second-order problems.

2.1.2. Learning

The end-goal in reflexive governance systems is always to achieve some form of learning. Radaelli (2009) defines learning in public policy analysis as "a process of updating beliefs about key components of policy (such as problem definition, results achieved at home or abroad, goals, but also actors' strategies and paradigms)" (pp. 1046-1047). Learning in governance can take many different shapes and the proposition of this paper is that all types of learning are valuable. The following discussion on different types of learning largely draws on existing literature on knowledge bases, economic geography, and firm innovation.

When experimenting to discover the effects of policies and regulations, one might discover that they work as intended. However, they may just as well not work as intended, which is also valuable learning for governance as knowledge about ineffective measures save tax-payers money that can be better employed elsewhere.

In the economic literature on firms, authors have distinguished between vertical and horizontal learning (Li, 2014; Li, 2017). Vertical learning refers to how firms accumulate knowledge through their input-output relations with other firms (Li, 2017). A typical example of this is how producers and suppliers in value chains collaborate in the design of new products. In this strand of literature, horizontal learning "relates to a process in which knowledge is shared and created, directly or indirectly, among individuals and firms that have the same areas of expertise and are conducting similar activities" (Li, 2017, p. 392). The underlying rationale for such

learning can be economic, social, psychological, or strategic (Li, 2017).

For the purpose of this paper, we find it useful to translate this distinction between vertical and horizontal learning to governance. We consider that vertical learning in reflexive governance structures arise when policy-makers and regulators learn from the effects of policies and regulations on various societal stakeholders and continuously adjust them accordingly. Horizontal learning is also likely to be present in governance structures. According to Li (2014), "as 'social animals', individuals exchange information naturally" and why should this not be true for governance processes? Indeed, similar concepts already exist. "Relational learning", which "relates to issues such as trust-building, changes in the ability to cooperate, and changes in the ability to understand another party's goals and preferences" (Munaretto & Huitema, 2012, p. 19). Tosun and Lang (2017) explain "horizontal governance" with "supplementing or replacing government by establishing and strengthening horizontal networks of different governmental units to increase policy coordination, collaboration, and shared responsibility" (p. 559).

Greenstone (2009) provides one of the likely many ways to generate vertical learning in governance structures through a four-step institutional framework: first, the implementation of policies and regulations must make it possible to evaluate their causal effects; second, one must secure financing for independent evaluations of existing regulations (i.e., by establishing independent regulatory review boards). Third, he recommends that all new policies and regulations should contain an evaluation date accompanied by automatic sunset and expansion provision to generalise policies and regulations found to work and drop those found not to work. Finally, he recommends developing a code of ethics to improve potentially negative public associations with the proposed framework.

A horizontal learning process is usually characterized by self-governance (Li, 2017). The author finds that horizontal learning takes place through the existence of social structures, such as friendship and family ties, labor mobility, interaction with - and monitoring of - others in a spirit of rivalry as well as collective invention. The literature provides us with some clues concerning the conditions under which individuals interact in constructive ways. Results from spatial clustering analysis show that geographic proximity is important for interactive learning. While it is clear that some of the mechanisms in Li (2017) for horizontal learning such as social structures and labor mobility are local, this argument goes beyond the practical matter of easy travel distances. Regions are areas for common norms and values that facilitate interactions (Grillitsch, Martin, & Srholec, 2017); and they contain specific firms, infrastructure, knowledge, and institutions that tie firms, customers, research institutions, and local authorities to each other (Malmberg, 1996). One should nevertheless not exclude the possibility of interregional knowledge exchange (Vang & Chaminade, 2007; Fitjar & Huber, 2015; Grillitsch & Nilsson, 2015). In multi-level governance systems, intra- and interregional exchange is crucial for

learning. Furthermore, global scale horizontal learning processes are facilitated by the development of information and communication technologies (Li, 2017).

The literature is less clear when it comes to how the public sector can play a role in free-market horizontal learning structures (Li, 2017). Yet, to encourage firm innovations that emerge through collaborations involving several stakeholders instead of innovations that occur isolated in one firm it is likely that the role of the governance in these constellations should relate to building up appropriate governance structures and possibly by providing public funds for innovation support.

The signature feature of reflexive governance systems is that all actors engaged in these vertical and horizontal processes learn - through trial and error - both how to generate innovative policy and regulatory processes, and how to foster innovative developments in the market.

Learning can come about by several means. The most basic approach is systematically evaluating the outcomes of the legal and regulatory framework in place to identify shortcomings. Furthermore, findings from laboratory experiments can offer additional insights into the effects of specific policies, regulations, or institutional arrangements. Moreover, through scenario processes, one can gain a better understanding of medium- and long-term developments initiated by policies and regulations (e.g., Kleihauer, Führ, & Schenten, 2019). Moreover, simulation games involving experts and practitioners in the field as well as desk officers enable testing the interplay of actors under an amended legal and regulatory framework (e.g., Führ, Dopfer, & Bizer, 2018). Finally, experiments can generate learning.

Experimentation or flexibility clauses in-laws are one tool for experimentation (Maaß, 2003). Experimentation clauses authorise the executive to deviate from the existing law by a predefined degree. They allocate legal flexibilities or financial support for socio-technical or administrative innovations (Schwartzing, 2003), thus enabling the administration to carry out innovative projects, which may subsequently become a permanent part of the governance framework (Maaß, 2003). In Germany, municipal law, traffic law, laws on childcare and school legislation, all contain several examples of experimentation clauses. For instance, BMWi (2019) provides a detailed overview of a selection of different types of experimentation clauses to test (digital) innovations and how they are implemented in Germany.

While experimentation clauses are well suited to test specific socio-technological innovations, we suggest that it is helpful to take a broader view on experimentation to guide sustainable development processes. Social transformation processes require testing alternative governance options against one another. Furthermore, these are not linear processes and the design of governance structures should therefore facilitate changes when new technological developments occur. Experimentation for sustainable development, therefore, extends beyond the opportunity to test procedures to facilitate much broader systemic innovations including technical and social dimensions and new business models.

2.2. Experimentation in the social sciences

2.2.1. Two approaches to experiments

This section will show that there is a plethora of terms under which experimentation is discussed in the social sciences. To facilitate a comprehensive scientific debate necessary to operationalize real-world experimentation, it is, therefore, useful to review different ways of characterizing experiments that exist in the literature, which is the objective of this part of the paper. It is however useful to start by defining a common ground to the object of this study, namely the act of experimenting for reflexive governance.

For this, we use a modified version of the formulation in Morton and Williams (2008) that “the defining characteristic of experimental research is intervention by the researcher in the data-generating process” (p. 3). The modification stems from our focus on governance. This means that we are of course also interested in cases where policy-makers or regulators intervene in the data-generating process, both consciously and unconsciously. The latter case are natural experiments. The focus on governance also means that we do not cover experiments that do not entail active participation from the policy-maker or regulation although they need not set the impulse or be the prime stakeholder in the experiment.

To review the literature on such experiments, we find it useful to start by differentiating between two general ways of approaching experimentation in the social sciences. The first focuses on the deductive nature of an experiment. It may be likened to a government-centered approach where the focus lies on the service delivery (Tosun & Lang, 2017). Only in this case, the focus lies on the efficiency of the instrument tested in the experiment. A central feature is the necessary presence of a control group that is unaffected by the considered policy or regulation to ensure causal relationships between the treatment and its behavioural effects. Delimitations in terms of time, range, problem situation, and/or subject matter are all (in theory) possibilities to demark the treated from the control group.

In this framework, an ideal experiment introduces alternative policies and regulations in different parts of the same entity for a limited amount of time before evaluating their consequences and deciding on the best approach. This way, one can be as sure as possible in the social sciences to compare the relative effects of the policy/regulation while minimizing the interference of confounding factors. However, this approach is unrealistic for most governance processes due to both political and legal hurdles. Examples of such hurdles are the public’s acceptance of policies and regulations and its likely reticence to unequal treatment as well as the legal principle of equality (see Section 5 of this paper for a more detailed discussion of the challenges to experimentation).

According to Morton and Williams (2008) “the aspect of control that is most important is [...] that the researcher can control confounding variables [...] in order to make the comparison meaningful” (p. 4). Individuals within the experimental area often deviate from those outside it to some extent in ways that cannot be observed by the evaluator. Random

assignment of the experimental policies and regulations is, therefore, one solution as randomization implies that “in large samples the treatment indicator and the covariates are independent” (Imbens & Woolridge, 2009, p. 13), implying that there are no confounding factors. The use of randomization is however also difficult for reasons of public acceptance (see Section 5 of this paper for a detailed discussion). Our proposition is that, despite the near impossibility of a perfect control group, one can surely accept many alternative forms of control groups as the second best. However, the evaluation experiments should address this issue.

A second approach focuses on the participatory nature of learning and collaborative governance arising in experiments. It is similar to the concept of policy integration where actors from two or more different policy domains take each other’s aims and concerns into consideration (Tosun & Lang, 2017). Only here, the challenge is not on the different policy domains but on the (possibly) different, even sometimes conflicting, preferences of stakeholders. According to Candel and Bisbroek (2016), the adaptation of procedural rather than substantive policy instruments characterizes policy integration. Translating this into our research endeavor, it becomes important to ensure broad participation. This means emphasizing cooperation between various stakeholders, among which the regulator is naturally an important player in the experiment. As the output of collaborative governance processes (i.e., improved regulation) is a public good, the policy-maker or regulator may also want to consider compensating participating actors, especially if they are confronted with the additional costs as a result of the experiment. Finally, the procedural elements and mechanisms for ensuring transparency to the public are here important.

Regardless of the approach taken to experimentation, the aim is always to generate new knowledge and learning. The expected knowledge gain is however somewhat different according to the approach taken. When focusing on the deductive nature of an experiment, vertical learning as defined in the previous section of the paper becomes the outcome of interest and the key terms are internal and external validity of the experiment. When focusing on the participatory nature of social learning and collaborative governance, the outcome of interest is the interactions of the relevant actors in the field and the horizontal learning, as defined in the previous section that arises through these interactions.

In this paper, we consider both approaches to the experimentation of the corresponding knowledge gained equally as important for reflexive governance in order to achieve sustainable development. Our broad definition of what we refer to as “regulatory experimentation” is placed in one of the next sub-sections below.

2.2.2. Overview of terminology

In the social sciences, one encounters various terms for the act of experimenting with policy and regulation. It is, therefore, useful to create an overview of some common concepts, which is the objective of this part of the paper. All of the

concepts reviewed are relevant for sustainable development in line with our starting assumption that sustainable development is closely linked to reflexive governance and experimentation. The overview however also shows clear differences between the concepts. For instance, some concepts are clearly theoretical discussing how the political system can induce experimentation, whereas others have emerged from applied projects. We emphasize such conceptual differences have a positive (not a normative) character, as we do not pertain that one concept is better than the others. Nevertheless, the differences are important to keep in mind both in academic and policy discussions on experimentation. They showcase the need to always carefully define the object of consideration to make sure all discussants speak the same language.

Policy experimentation is one commonly-used concept, although what it actually means remains open to debate. In the discussion of policy experimentation in the social sciences, Huitema et al. (2018) highlight the need for some common understanding of the concept as it is not helpful to categorise every policy as an “experiment”, and yet to date, the conceptual clarity is lacking. Moreover, Ansell and Bartenberger (2016) highlight that “even a quick scan of the [...] literature reveals that they do not necessarily mean the same thing when they use the term ‘experiment’” (p. 64).

According to Campbell (1997), randomisation is the defining feature of policy experimentation. For McDermott (2002), policy experiments are cases in which the investigator has control over the recruitment, assignment to random conditions, treatment, and measurement of subjects. Castán Broto and Bulkeley (2013) consider policy experiments as novel, purposive initiatives emerging outside a formal policy process. Both Farrelly and Brown (2011) and Bos and Brown (2012) propose a similar understanding of policy experiments as alternative policy processes that enact new ideas with the aim of identifying ways to upscale them. The authors highlight that policy experimentation is a deductive procedure whereby an underlying theory is proven correct or incorrect. As such, the existence of a hypothesis to be tested is a necessary condition for a policy experiment. McFadgen and Huitema (2018) define policy experimentation as “a temporary, controlled field-trial of a policy-relevant innovation that produces evidence for subsequent policy decisions” (p. 1768). Two conditions are necessary for a novel policy to fall into this category, namely, it must test an explicit hypothesis and involve some form of novelty.

Other similar concepts to policy experiments used in the literature are *democratic experimentalism* and *laboratory federalism*. The core idea of democratic experimentalism is that many different local units experiment in parallel. Through explicit monitoring and evaluation, one can identify best practices to inform future policies and regulations. The concept of laboratory federalism is similar to the main takeaway from this literature is that federal systems have a significant advantage over unitary systems. Federal systems are divided into local units, which can be used for experimentation. One can loosely understand the theoretical concept of laboratory federalism as innovative governance introduced in a federal

multi-level system that can induce experimentation, learning, and competition (Kerber, 2005; Kerber & Eckardt, 2007).

The *sandbox concept* originated in the domain of financial services, especially in the United Kingdom (FCA, 2015). The energy sector later adopted the concept, e.g., with Ofgem’s sandbox scheme in the United Kingdom (Ofgem, 2018). Furthermore, the IEA International Smart Grid Action Network (ISGAN) discusses sandboxes that focus on enabling companies to test new products, services, or technologies by providing them with regulatory exemptions (ISGAN, 2019).

Instigated by subnational governments, *urban laboratories* are “mechanisms that mobilize place to generate economic wealth and stimulate more resilient urban conditions, both through the creation of new landscapes and the retrofitting of existing ones” (Evans & Karvonen, 2014, p. 413). Experimentation within urban laboratories involves three key elements, namely situatedness, change-orientation, and contingency (Karvonen & van Heur, 2014). The first refers to the existence of some form of the border, the second relates to the existence of some dynamic process that involves a “new rule of conduct” and a “new definition of the situation” in the sense of Park (1929, p. 17), whereas the last term means that laboratories are always associated with incertitude and the possibility of failure.

Several concepts related to experiments in *real-world niches* have emerged over recent years in the realm of the transdisciplinary and transformative research agenda (see Schöpke et al., 2017 for an overview and comparison). Here, the role of the policy-maker/regulator regardless of government level is very limited. These experimental concepts include strategic niche management (Hoogma, Kemp, Schot, & Truffer, 2002), transition experiments (van den Bosch, 2010; Loorbach & Rotmans, 2010), living laboratories/labs (Voytenko, McCormick, Evans, & Schliwa, 2016; Liedtke, Baedeker, Hasselkuß, Rohn, & Grinewitschus, 2015), urban transition labs (Nevens, Frantzeskaki, Gorissen, & Loorbach, 2013), social innovation labs (Westley, Antadze, Riddell, Robinson, & Geobey, 2014), and real-world laboratories (Schneidewind & Singer-Brodowski, 2013; Schöpke et al., 2017; WBGU, 2016). The latter concept focus on sustainability-related innovations in a broad sense (social, cultural, technical, and economic innovations) (Parodi, 2019).

Another related concept is that of *regulatory innovation zones* (RIZs), which nevertheless clearly specifies a single experiment, which must be temporarily and spatially limited (Bauknecht, Heinemann, Stronzik, & Schmitt, 2015). Furthermore, the object of RIZs more explicitly relates to policies and regulations.

The *real-world laboratory concept* in particular has also captured the regulators’ attention in Germany. A recent report commissioned by the German Federal Ministry for Economic Affairs and Energy⁴ considers what is referred to as “regulatory sandboxes” in English (in German “*Reallabore*”) as experimental areas – which are delimited in time and space as well as legally protected – that permit

⁴ The same ministry has launched a network for real-world laboratories (“Reallabore”). Website (in German): <https://www.bmwi.de/Redaktion/DE/Dossier/reallabore-testraeume-fuer-innovation-und-regulierung.html>

testing the interplay between innovation and regulation under real-world conditions (BMW, 2018). In their understanding of the concept, the regulatory component concerns only investigating the rules and regulations needed for a given new technology. A selection of their examples of regulatory sandboxes used for illustrative purposes includes testing an automatised parcel delivery robot in the city of Hamburg, testing a platform for remote medical services in the German region of Baden-Württemberg and testing autonomous cars in the capital Berlin (BMW, 2019). There are no clear design criteria for taking part in the Ministry's initiative, however, the initiative from the Ministry itself has a clear orientation towards digital innovations.

To sum up, our approach to what we refer to as *regulatory experimentation* for reflexive governance, we consider the following aspects important: testing explicit hypotheses, the existence of interactions between different actors, the presence of a control group that need not be a perfect counterfactual, as well as monitoring processes to ensure learning. Successful experimentation becomes even more important in the context of sustainable development and such experiments must be inclusive and should lead to both vertical and horizontal learning.

Hence, we favour a broad understanding of the concept of regulatory experimentation. Since our purpose is to describe real cases and develop a framework for analysis, we consider the disadvantage of including less relevant cases in our analysis to be much smaller than the disadvantage of excluding relevant ones.

3. RESEARCH METHODOLOGY – DERIVATION OF AN ANALYTICAL FRAMEWORK

This section zooms in on our concept of regulatory experiments. We do not design or conduct a regulatory experiment by ourselves but rather derive a comprehensive analytical framework for analysing and comparing such experiments in practice. This framework can be used by researchers for future research as well as practitioners planning to introduce regulatory experiments. Along with the four core features of regulatory experiments – 1) clear hypothesis, 2) interaction between actors, 3) causality, and 4) monitoring and learning – we outline a number of variables that any analysis of regulatory experiments should include. The derivation of this analytical framework builds on existing theoretical and empirical literature on experimentation in the social sciences. Subsequently, we illustrate the application of this framework by conducting a qualitative content analysis (Mayring, 2010) of three examples of regulatory experiments based on publicly available documents.

The literature has already attempted to categorise different types of policy experiments. Starting from the most general typology, Huitema et al. (2018) distinguish between approaching policy experimentation as a research method and an approach to governance. As a research method, experiments are treatments seeking to produce causality through randomisation and statistical analysis. As an approach to governance, experimentation is a means to test several options when solving societal problems and to develop regulation in practice in order to draw plausible

conclusions from the respective learning processes. They note that the most important differences between the various approaches to policy experimentation are whether experimentation is one specific method or a composition of several types of methods, as well as the extent to which they require a strict experimental design as well as solely quantitative data collection.

Ludwig, Kling, and Mullainathan (2011) distinguish between policy evaluations and mechanism experiments. Policy evaluations test the effect of a certain policy by implementing it on a small scale using randomisation procedures to form treatment and control groups. Mechanism experiments focus on discovering specific causal mechanisms that link a policy to given outcomes.

Howe (2004) distinguishes between two forms of “experimentalism”, namely neoclassical and mixed methods. The former relies exclusively on quantitative methods, whereas the latter opens up for the use of qualitative evaluation methods.

Ansell and Bartenberger (2016) distinguish between three experimental logics: controlled (identifying causality ideally in randomised controlled trials), Darwinian (enhancing systemic innovation through continuous trial and error), and generative (a process of generating and iteratively refining a solution to a social problem).

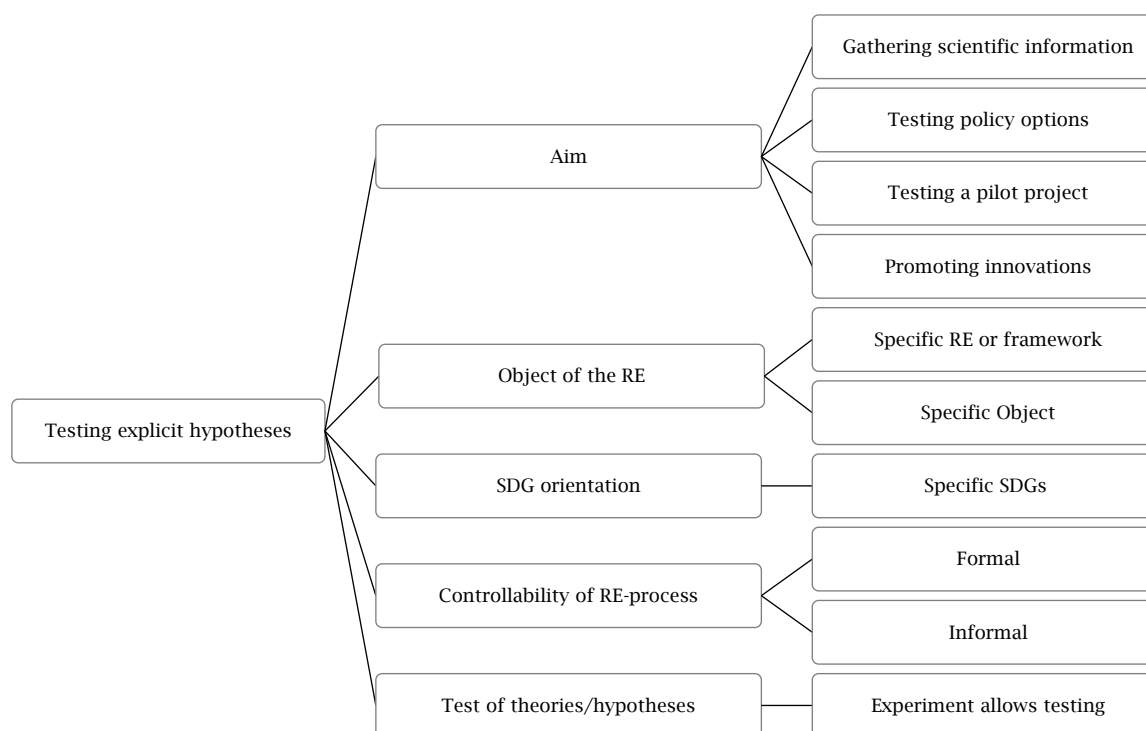
Finally, experiments have also been categorised according to one specific feature; for instance, based on the role of science in policy-making (Pielke, 2007), governance design (Weber, 1968; Dryzek, 1987) and learning outcomes (McFadgen & Huitema, 2018).

While all of these are important contributions, they only look at a limited number of variables defining experiments. Therefore, in our view, these are more suited for specific aspects of experimentation and less appropriate as a tool for surveying and analysing a heterogeneous sample of regulatory experiments according to our broad definition. For this, we need a categorisation that covers a multitude of aspects of regulatory experimentations in order to enable comparisons.

Hence, in the following, we outline a new analytical framework that describes existing regulatory experiments based upon our four criteria of regulatory experimentation. Both for hypotheses testing, participation, causal framework, and learning/monitoring, we define variables that are the features of regulatory experiments within each category that we consider to be relevant. For each variable, we further define which form a given regulatory experiment can take. Depending on the specific variable, characteristics can have the form of either yes or no answers, numbers, two or more possible answers that do not necessarily have to be mutually exclusive and can even be an open question.

3.1. Test of explicit hypotheses

The first part of the analytical framework contains aspects of regulatory experiments related to hypotheses testing. More precisely, the variables specify the aim, object, and SDG orientation of the regulatory experiment, the controllability of the experimental process and, finally, whether the regulatory experiment serves to falsify specific hypotheses. Figure 1 depicts each of the variables on “testing explicit hypotheses”.

Figure 1. Variables for testing explicit hypotheses

Note: RE = regulatory experiment; SDG = Sustainable Development Goal.

The variable *Aim* captures the overall purpose of the regulatory experiment. Based on McFadgen and Huitema (2018), our analytical framework differentiates between the aims to “gather scientific information”, “test policy options”, and “test a pilot project”. We add a fourth aim of “promoting innovations” since most regulatory experiments relate to innovations (Bernstein & Hoffmann, 2018). Löher and Schneck (2018) argue that firm-initiated real-world laboratories can offer the opportunity to test innovations and their social acceptance. This element might be also relevant for regulatory experiments.

The variable *Object of the regulatory experiment* specifies what is explicitly considered in the regulatory experiment and it is divided into two questions. The first question reveals whether the regulatory experiment describes a specific experiment or whether it is a framework for several experiments. The second question asks for a detailed description of the object. As our approach to regulatory experimentation focuses on its contribution to sustainable development, we include the variable *SDG orientation*, which illustrates whether the regulatory experiments aim at a high level of protection for humans and the environment by identifying (when present) SDG goals in accordance with the aim of the regulatory experiment.

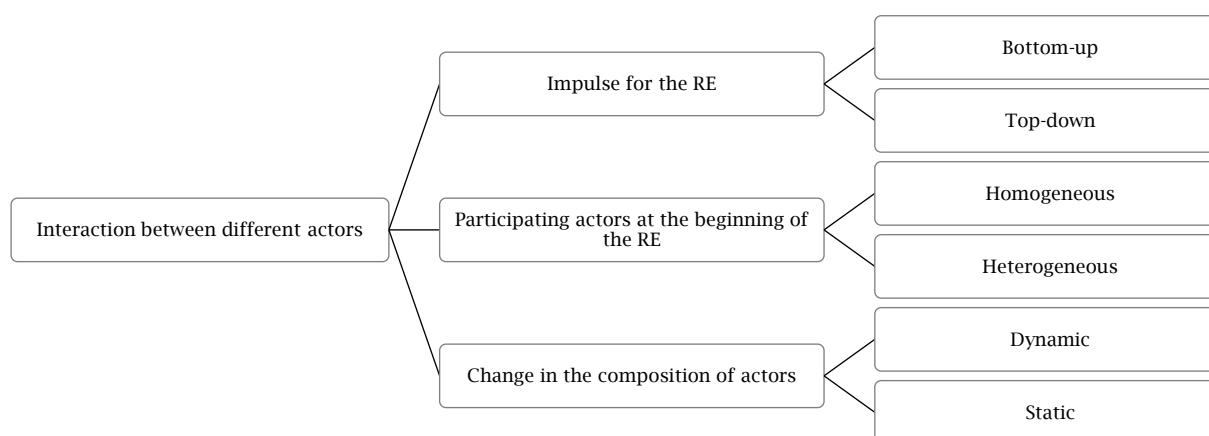
The variable *Controllability* illustrates the degree of formality of the experimental approach, which is important for testing explicit hypotheses. Existing experiments in voluntary commitment systems in the field of climate governance inspire the distinction between “formal” and “informal” experimentation (see. e.g., Abbott, 2017). In formal experiments, the experimental design lives up to scientific standards as far as it is possible in real-world contexts. This is not the case for informal experiments, which may also arise without an explicit experimental intention from the regulator.

Finally, the variable *Test of theories/hypotheses* evaluates whether the regulatory experiment has led to an explicit refutation of hypotheses and theories being tested and is characterised by a simple “yes” or “no”.

3.2. Interaction between different actors

The second part of the analytical framework examines the interaction between actors participating in the regulatory experiment. It covers who initiated the regulatory experiment, which actors participated in the implementation of the regulatory experiment and controls whether the composition of actors has changed during the process. Figure 2 depicts each of the variables profiled in the “interaction between different actors” part.

Figure 2. Variables of the part “interaction between different actors”



Note: RE = regulatory experiment.

The variable *Impulse* depicts how the regulatory experiment came about. Here, we want to capture whether the regulatory experiment originated as a top-down or a bottom-up process. Policies and regulations form the institutional framework within which market actors operate. How they are shaped will hence influence innovation incentives and capacities in the market. Vice versa, individual behaviour and market developments determine the needs and outcomes of policies and regulations. In the case of top-down governance, policy-makers and regulators can adjust the institutional framework to influence individual behaviour. Inversely, bottom-up governance processes materialise when firms try to alter the national institutional setup if it does not suit their needs (Crouch, Schröder, & Voelzkow, 2009).

The literature supports mapping out relevant actors in regulatory experiments. Castán Broto and Bulkeley (2013) analyse the mix of heterogeneous actors in urban climate change experiments. Additionally, McFadgen and Huitema (2018) argue that different actors can initiate policy experiments: an expert elite seeking scientific knowledge (technocratic experiment), a collaboration of actors developing different policy options (boundary experiment), an organiser (often policy-makers), and other actors with the same problem perception wanting to establish particular actions (advocacy experiment).

Therefore, the next two variables cover the type of actors involved in the regulatory experiment and possible dynamics in their composition based on the work of Simon et al. (2018), who establish criteria to classify different types of real-world laboratories. The variable *Participating actors at the beginning of a regulatory experiment* can take on the characteristics of “homogeneous” when a single actor initiates the regulatory experiment or “heterogeneous” when a collaboration of different actors stands behind the regulatory experiment. The variable *Change in the composition of actors* focuses on the implementation process of the experiment. The characteristics “dynamic” and “static” indicate whether the composition of actors has changed.

3.3. Causality

The third part of the analytical framework investigates whether the design of the regulatory experiment allows measuring causal effects. Variables in this part capture the geographical scope and time frame of the regulatory experiment, as well as several topics addressing the design in detail. Figure 3 depicts each of the variables profiled in this part.

Two questions inform about the *Geographical scope* of the regulatory experiment: whether the regulatory experiment was ex-ante deliberately limited in space and which ex-post geographical demarcation characterises the regulatory experiment. The latter is an open question because one cannot list all possible geographical scopes. The variable *Timeframe* illustrates how the experiment is limited in time (number of years).

The next variable *Target group* identifies the sub-populations affected by the regulatory experiment. Again, in order to cover all possibilities that may arise in practical applications, this is the answer to the open question “What is the target group of the regulatory experiment?”.

Our understanding of regulatory experiments implies that it is important to look for the presence of some kind of control group. The control group can be either explicitly defined ex-ante of the regulatory experiment or becomes apparent ex-post. Regardless of whether it was the purpose at the launch, in a first step, we identify whether a control group exists; and in a second step, whether it differs from the treated group in time, space or individuals.

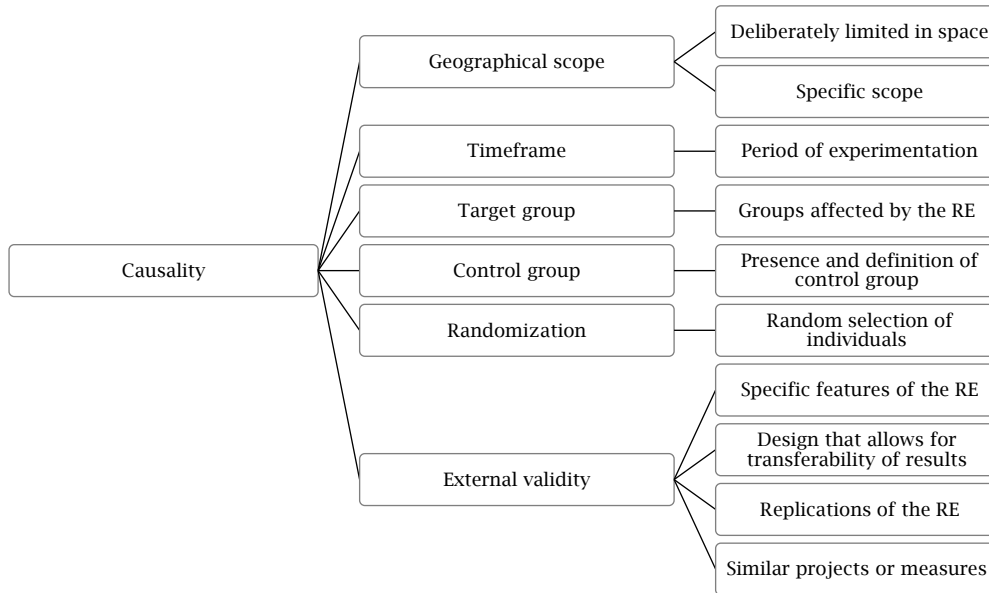
In laboratory experiments, researchers use randomisation as an assignment mechanism to define the treatment and control group based on the target population. We suspect that many real-world experiments do not (fully) fulfil this condition; and hence, we ask the question “Whether units of observation were selected randomly?”.

Finally, an important issue is whether the learning obtained through a regulatory experiment would be the same in different settings (Ludwig et al., 2011; Banerjee & Duflo, 2009). We include the final variable for this part *External validity of results* in our analytical framework. Four characteristics

depict this feature of the regulatory experiment. The first characteristic details the general features of the regulatory experiment that reduce the transferability of results. The remaining three characteristics review design elements that reduce problems of external validity issues, i.e., whether the design allows or simplifies the transferability results,

whether replications are possible, whether replication studies already exist, and finally, whether there are other projects or measures that test similar relations. Replication studies or the evaluation of related regulatory experiments can help to reduce uncertainty about the external validity of the results (Banerjee & Duflo, 2009).

Figure 3. Variables for causality

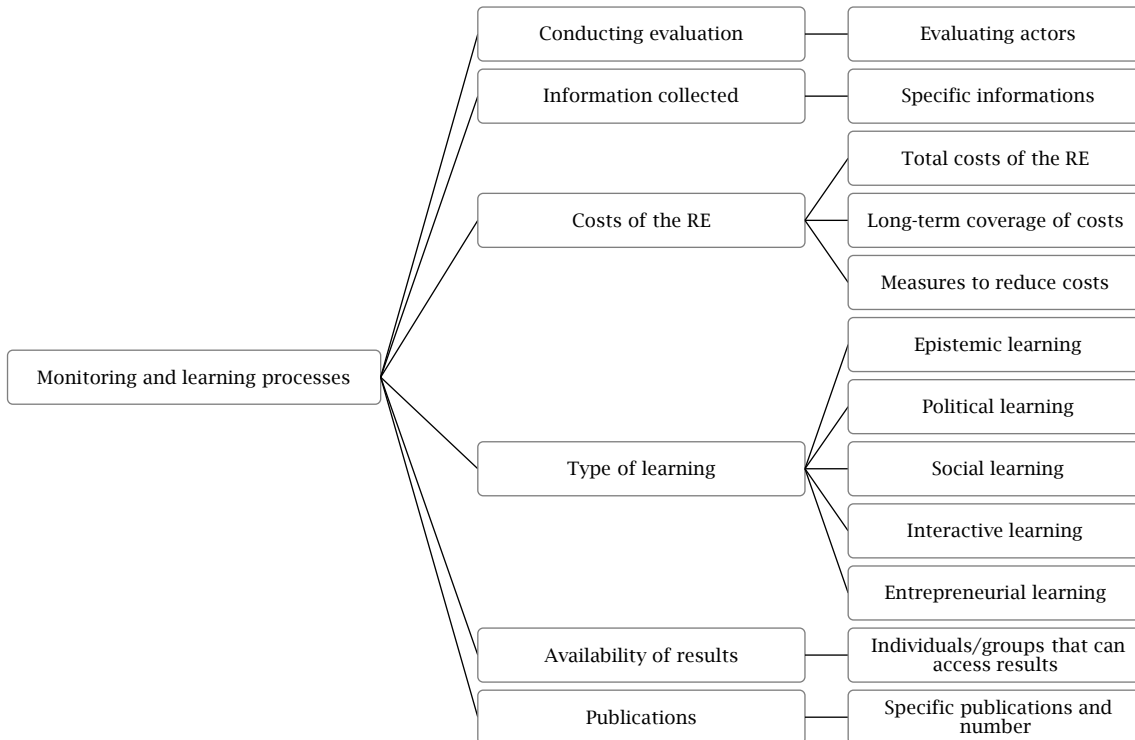


3.4. Monitoring and learning processes

The final part of our analytical framework covers the monitoring and learning processes of regulatory

experiments. This part specifies how the evaluation of the regulatory experiment takes place and the subsequent use of evaluation results. Figure 4 depicts each of the variables profiled in this part.

Figure 4. Variables for monitoring and learning processes



Note: RE = regulatory experiment.

The first two variables collect general information about the evaluation process. The open question “Who evaluates the regulatory experiment?” (“no one” being a possible answer here) characterises the variable *Conducting evaluation*. The open question “Which information is collected for evaluation purposes?” (again, “none” being a possible answer) characterises the second variable *Information collected*.

The next variables concern the costs of the regulatory experiment. The “total costs of the regulatory experiment” sums up the administrative costs in euro, whereas the characteristics “Is the long-term financing of the experiment covered?” and “Were specific measures undertaken to reduce costs?” (both yes or no are possible answers) illustrate the attention given to costs.

The variable *Type of learning* specifies whether and how insights from the regulatory experiment generate learning processes. The first two characteristics of this variable “epistemic” and “political” learning use the distinction of Ansell and Bartenberger (2016), whereby epistemic learning describes the accumulation of scientific knowledge whereas political learning is about changes in the preferences and goals of political actors.

To this typology, we add the characteristic “social learning” because regulatory experiments can also affect the preferences and goals of societal actors (Rocle & Salles, 2018). We also add “interactive learning”, which reflects the notion that regulatory experiments may affect actors’ behaviour regarding information acquisition, communication, and cooperation. Finally, we add “entrepreneurial learning” to collect information about how regulatory experiments might affect learning processes in firms, which in turn spark innovation.

The variable *Availability of results* examines through an open question “Who can access the results of the regulatory experiment?”. Finally, the variable *Publications* reveals whether the evaluation process of the regulatory experiment resulted in any publications. This completes our analytical framework to profile specific examples of regulatory experiments.

4. RESULTS – PRACTICAL ILLUSTRATION OF THE ANALYTICAL FRAMEWORK

To understand how to optimally use experiments in order to foster learning, researchers may want to move beyond single case studies and compare several experiments simultaneously. With our broad approach to experimentation, an empirical sample of cases of regulatory experimentation is likely to be very heterogeneous. Hence, there is a need to standardize them according to the relevant parameter, which is the purpose of the framework presented in Section 3 of this paper. The purpose of this section is to illustrate how this framework achieves this goal.

We examine the Finnish basic income experiment, regulatory sandboxes for financial services in the United Kingdom, and the German crafts deregulation in 2004 using qualitative content analysis (Mayring, 2010). The first case is one of a few real-world experiments closely following a deductive approach, the second falls into the sandbox-category whereas the latter is an

illustration of a natural experiment. As such, these experiments do constitute an (albeit small) heterogeneous sample (see Appendix for the three completed short profiles)⁵.

Our analysis of all three cases is based on publicly available documents (official websites, research publications, official documents)⁶. Following Mayring (2010), we examine these documents based on predefined rules, which means that we summarize and structure the material using a theory-based category system. The analytical framework developed in the previous section constitutes our category system. We first summarize the results of this categorization for each case separately before comparing the three cases.

4.1. The Finnish basic income experiment

The first part of the framework maps out features related to hypothesis testing. The Finnish basic income experiment case was a specific regulatory experiment. We classify it as a formal experiment that was conducted according to scientific standards. Indeed, its primary aim was to gather scientific information and testing a pilot project by investigating new concepts of social security on a small scale before possibly introducing them nationwide (Kangas, Jauhiainen, Simanainen, & Ylikännö, 2019). The knowledge potentially generated would foster a multitude of SDGs, namely “no poverty” (SDG 1), “good health and well-being” (SDG 3), “decent work and economic growth” (SDG 8), and “reduced inequalities” (SDG 10).

The second part of the framework that focuses on interactions between actors classifies the Finnish basic income experiment as “top-down” since the regulator has played a key role from the start. The Finnish government first made the decision to conduct the experiment. The Prime Minister’s Office assigned a preliminary study to a consortium of scientific facilities, which should evaluate different models of basic income before the start of the experiment. The Finnish social security institution (Kela) was also responsible for the implementation of the experiment (Kela, 2019). Hence, a variety of stakeholders were involved in the experiment. The target group as all finish people between 25 and 58 years of age that received labor market subsidy or basic unemployment allowance in November 2016 (Kela, 2019). Hence, a heterogeneous group of stakeholders participated in the experiment.

Particularly, interesting to note in the third part of the framework dedicated to causality is the fact that 2000 randomly selected individuals formed the treatment group. They received partial basic income of 560 euros per month for two years (2017-2018). The remaining individuals of the target group formed the control group (around 175000 individuals) (Kangas et al., 2017). Descriptive statistics comparing the composition of the treatment and control group suggest that the randomization procedure was successful. As a consequence, although the experiment covered the whole country and was not limited in space, the chances of internal

⁵ The sample in this section is only for illustrative purposes. Future empirical studies should base their sample selection on established sampling techniques, e.g., theoretical sampling (Glaser & Strauss, 1967).

⁶ We consider relying on publicly available documents as the most realistic approach if a large sample is to be obtained. Future empirical studies could also rely on further material, e.g., qualitative interviews.

valid evaluation results are high. External validity is also suspected to be present. Morton and Williams (2008) however, emphasize that one can only assess external validity by examining similar experiments which in this case is possible since basic income experiments have been done elsewhere such as in the US and Canada.

Finally, for the fourth part of the framework on monitoring and learning processes, the Finnish social security institution (Kela) in cooperation with other research institutes were from the on-set mandated with evaluating the experiment (Kangas et al., 2019). We found no information in evaluations regarding the implementation cost. The evaluators used register data for the first year of the experiment (2017) to analyse employment and income effects; and survey and interview data to measure the wellbeing effects of partial basic income. They find no significant difference in days in employment between the treatment group and the control group. However, a higher share of basic income receivers compared to the control group reported that they believe to find employment in the next twelve months. This indicates that there might be positive effects of basic income on employment in the longer run. A more thorough analysis planned for 2020 combining the register and survey data for the years 2017 and 2018 aims to control for unobserved characteristics (Kangas et al., 2019). Even so, we consider that several types of learning arose through these experiments, which resulted in a number of publicly available publications.

4.2. Regulatory sandboxes for financial services in the United Kingdom

With regards to hypothesis testing, although the authorities through the sandbox initiative aim to enhance the regulatory knowledge in new technological fields to create innovation through competition that benefits consumers (FCA, 2015, 2017), there was no explicit hypothesis to be tested formulated at the beginning of the experiment. This initiative “allows firms to test innovative products, services, and business models in a live market environment while ensuring that appropriate safeguards are in place” (FCA, 2017, p. 3). Although the initiative can lead to many different experiments, we consider here the general framework for testing several innovations. The experimental process is as such very informal since the authorities in each case have little influence over the choice of innovations to be tested and the design of the experimental process. In terms of its effects on sustainable development, given its success, this experiment promotes decent work, economic growth (SDG 8) and industry, innovation, and infrastructure (SDG 9).

Regarding the interactions between actors, it is worth highlighting that this is also a top-down experiment initiated by its main actor the Financial Conduct Authority (FCA). Firms applying for the regulatory sandbox form the second group of relevant actors. So far, the FCA has received 375 applications for regulatory exceptions divided into 5 cohorts. Among them, about 131 have been accepted (FCA, 2020).

Regarding causality, this experiment is not limited in space as it concerns the whole UK. It launched in 2016 and is also not limited in time.

It is, however, worth noting that the experiments by participating firms are of course limited in space and also in time (6 months). The target group is authorized firms, unauthorized firms that require authorization and technology businesses in the United Kingdom. It is hard to define a control group because the participating businesses test different innovative products. There is also no treatment randomization because firms need to apply for the experiment and hence, self-select into using the experimental framework. Participation is based on different eligibility criteria mainly related to innovation. Hence, the internal validity of effect evaluations is expected to be low as the limited test period of 6 months makes it difficult to cover long-term effects. The external validity is reduced by specific institutional factors in the United Kingdom (e.g., strong financial sector, specific national regulations).

Monitoring and evaluation of the experiment take place by the initiating authority the FCA itself as well as by participating firms. The costs to the public budget is expected to be low as the regulator provides the framework for testing and firms test themselves. For evaluation purposes, firms need to summarize the outcomes of their tests. The FCA (at least for the first two cohorts) monitors among other factors test completion, the share of firms receiving investments following tests, number of applicants, compliance with standard safeguards, and limitations of testing. The resulting information is available as internet information and FCA publications. Overall, the FCA draws positive conclusions after two years of testing (FCA, 2017). For instance, of the accepted firms in the first cohort, 75% completed the testing, of these 40% received investment during or after their tests and about 90% are working on a wider market launch after the testing. According to our profiling, the experiment resulted in political learning through knowledge about how to regulate tested innovations in the financial sector as well as interactions between firms, the regulator and customers, and entrepreneurial learning. The latter is also the initial purpose of the experiment.

4.3. The German crafts deregulation

The Trade and Crafts Code (*Handwerksordnung*) regulates the German craft sector. From 1953 to 2004, all German craft trades were subject to a licensing scheme. Only craftsmen with a Meister title (internationally recognized as tertiary education) could found a company. Since 2004, market entry in certain trades has been open, although craftsmen in these trades can still choose to acquire the Meister title. In the remaining trades, the licensing requirement remains fully or partially intact. Starting again first with the hypothesis-testing aspect, the objective of the regulator was to increase competition in the crafts sector by lowering entry barriers. As such, there was a clear policy goal and no ex-ante hypothesis to be tested. Nevertheless, as will be explained, we still consider the experimental process as formal since the reform (given certain assumptions) unintentionally constitutes a natural experiment. Goal attainment would support production, job creation and entrepreneurship and hence boost economic growth (SDG 8).

Regarding stakeholder interaction, this is a clear top-down regulatory experiment initiated and carried through by the German national regulator. Further relevant actors comprise the German crafts sector that makes up about 5 million professionals making up about 12% of the working population (German Federal Statistical Office, 2016) as well as the German Confederation of Skilled Crafts (ZDH).

This experiment was not limited in space as it concerned the whole of Germany. However, a control group was still present as the target group was first and foremost the self-employed in the fully or partially deregulated trades, and individuals working in the still fully and partially regulated crafts occupations were (at least directly) not affected. Studies also suggest that employees in these trades may also have been affected (e.g., Damelang, Haupt, & Abraham, 2018). While the experiment was initially not limited in time, the German government reversed the reform in 2020 for 12 of the 53 initially deregulated trades (perhaps again a natural experiment)⁷. Also, regarding randomization, there was no intention of randomizing treatment and controlling individuals when the reform was passed. However, in practice, this may have happened. It was the political intension that both occupations considered hazardousness and/or providing a significant contribution to vocational training in Germany should remain regulated. The minutes of the negotiations do however also provide evidence for interest group lobbying. Hence, the design of this experiment is likely to lead to (unintentionally) high internal validity of the evaluation results. Whether these are externally applicable to other cases of occupational licensing is debatable. Germany has a long tradition for occupational licensing in the crafts; and the regulation has effects on the country's vocational training system, which is very specific for Germany. However, given the number of licensing regulations, a number of studies on their effects from other countries and sectors exist for comparative purposes.

A number of researchers have taken advantage of this natural experiment to evaluate the effects of occupational licensing on market entry, market exit, the share of migrants, incomes, employment, and in-company vocational training. In sum, this literature yields the following insights: market entry has increased (Rostam-Afschar, 2014; Runst, Thomä, Haverkamp, & Müller, 2018; Koch & Nielen, 2017; Zwiener, 2017), market exit has (Runst et al., 2018) or has not (Rostam-Afschar, 2014) increased, incomes of self-employed craftsmen and craftsmen employees in deregulated trades have probably been negatively affected. Although the effects are very small (Lergetporer, Ruhose, & Simon, 2018; Damelang et al., 2018; Koch & Nielen, 2017; Fredriksen, 2018; Sonntag & Lutter, 2018), the probability of migrant self-employment has increased (Runst, 2018), effects on employment (Koch & Nielen, 2017; Zwiener, 2017) are undetermined and the number of Meister examinations (Koch & Nielen, 2017) and the level of basic vocational training provided by companies in deregulated trades have decreased (Runst &

Thomä, 2020). Hence, we conclude in our profile of the German crafts deregulation that this experiment has produced epistemic learning and political learning.

4.4. Comparative analysis

The profiling of these three cases suggests some insights for the design of regulatory experiments. The Finnish experiment is in many ways a textbook example of the deductive approach to experimentation. In accordance with the deductive way of approaching experiments, explicit hypotheses are tested. The aims at the onset of the experiment were to gather scientific information and to test a policy option. Also, importantly the experimenters formed treatment and control groups through randomization.

The German deregulation in the crafts sector shows that a regulatory experiment can follow a deductive logic without its regulator's or policy-maker's intention and without the ex-ante definition of hypotheses. The German national legislator in this case did not plan to generate treatment and control groups and certainly not by randomization. However, scientists have ex-ante showed that this turned out to be the case. Both of these experiments generated a high degree of epistemic learning.

In contrast, the United Kingdom financial sector regulatory sandbox neither started with an explicit hypothesis to be tested nor can any explicit control group be identified. This is likely to be a feature of many existing sandboxes. Hence, according to a deductive approach, one might conclude that this experiment leads to less knowledge-gain. However, an advantage of this experimental design is that by opening up existing regulatory frameworks, it inherently tests several paths that can advance the UN SDG goals related to innovation and economic growth. While testing several alternatives against one another in the two other cases considered in this report, in theory, would have been possible, it was not done in practice. In the Finnish case, despite the multiple designs proposed in a preliminary study, the government chose to only test the partial basic income model of 560 euros per month and a one-stage sampling procedure (Kela, 2019). We, therefore, conclude that regulatory experiments designed as sandboxes generate different types of knowledge-gain, in this case with more emphasis on entrepreneurial learning.

We can also use our framework to look for traces of the participatory and collaborative way of approaching experiments in our three cases. In the Finnish basic income experiment, several different actors took part in the planning, implementation, and evaluation processes and the learning processes arose via the public debate in addition to the formal evaluation of the experiment. In the German crafts deregulation, horizontal learning processes are not obvious. In particular, the participation of crafts consumers lacks. This is, however, not surprising, given the vertical learning in this case is also more of a by-product of the deregulation than the intention of the regulator. In the case of the United Kingdom financial sector sandbox, we conclude that interactive learning has taken place through

⁷ Pressrelease in German available here: <https://www.bmwi.de/Redaktion/DE/Pressemitteilungen/2019/20191009-altmaier-wiedereinfuehrung-der-meisterpflicht-starkes-signal-fuer-die-zukunft-des-handwerks.html>

interactions between firms, the regulator, and customers. However, we recognize that the survey of publicly available documents, which is our primary source of information to fill out the framework, is not sufficient to pick up all aspects of collaborative governance in experiments, which may often not be included in official documents. We, therefore, see a need for further research into horizontal learning processes in regulatory experiments using methods such as interviews.

In sum, we tentatively draw the following lessons from these three cases. Most cases of experiments in the real-world do not correspond to the formal experimental approach that we know from laboratories. We do not see this as a case against experimentation in the social sciences. Rather, it shows the rich diversity which this instrument embodies. We see no need to throw the baby out with the bathwater and relinquish an experimental approach fully just because ideal conditions are not met. But it of course means that one needs to take extra care when interpreting results.

To fulfil the deductive approach to regulatory experimentation, randomization is desirable and possible, as the Finnish basic income experiment shows. To fulfil the participatory approach to regulatory experimentation, it is useful to go beyond the target group and include all relevant actors when mapping out the stakeholders of a regulatory experiment, which for instance in the case of the German crafts deregulation would imply also considering its diverse effects on consumers.

Finally, these cases illustrate one key assumption in our framework and in this paper: the learning obtained from regulatory experimentation is diverse. The cases covered all generate at least one type of learning, however, the type of learning obtained varies according to the experiment considered. Epistemological learning follows a strong focus on the deductive side of regulatory experimentation whereas entrepreneurial learning will materialize in experiments classified as regulatory sandboxes. For reflexive governance, it is, however, not enough that learning arises in regulatory experiments, it must also be used for subsequent regulatory and policy-making decisions. Our framework does not cover this important aspect, which we leave for further research.

5. DISCUSSION

Despite the potential advantages of experimentation highlighted in this paper, the use of regulatory experiments faces several hurdles. To ensure that regulatory experiments lead to the desired knowledge gain, it is important to keep them in mind when designing such experiments.

As politicians care about re-election, it is unlikely that they will risk enforcing policies and regulations potentially leading to a large loss of votes. Public acceptance, therefore, poses an important challenge for regulatory experimentation. Greenstone (2009) notes that “people frequently have a visceral reaction against experiments that involve humans” (p. 121). To address this problem, he suggests developing a code of ethics for all experimentation. Furthermore, individuals may be able to overcome such inherent disapprobation if

they are offered a pecuniary benefit from it. Orcutt and Orcutt (1968) emphasise that incentive regulations lend themselves best for experimentation because by design every participant can make a net gain.

Greenstone (2009) also notes that “some consider randomized experiments unethical, because they relegate a significant number of people to the control group when there are non-experimental reasons to believe that the treatment will prove beneficial” (p. 17). In a similar vein, other authors have highlighted that experiments are not neutral and they affect various groups in the population differently, which consequently affects the political dynamics surrounding experimentation (Huitema et al., 2018)⁸. For this reason, political communication about social benefits plays an important role. Emphasising the advantages of reflexive governance could, therefore, enhance the public acceptance of regulatory experimentation. Orcutt and Orcutt (1968) argue that even experimentation that can be potentially dangerous for the treatment group can gain acceptance if the public regards the research objective as socially important. It could also be helpful to remind the public that regulators, such as the Federal Drug Administration in the US, already undertake experiments and that most would find it unacceptable that an untested drug should be available on the market.

Another important challenge of regulatory experimentation is that policy-makers and regulators must acknowledge that there is uncertainty tied to the outcome of a particular policy or regulatory measure. Politicians and regulators are likely unwilling to communicate uncertainty and probably prefer to show profound knowledge about future development towards voters. Instead, engaging in regulatory experimentation presupposes curiosity to investigate potential outcomes to lift this uncertainty.

Orcutt and Orcutt (1968) emphasise costs as an important barrier to experimentation. The authors highlight that costs are driven upwards by the fact that experiments must be of a certain scale (in terms of both time and space) to produce learning and because there must be some form of compensation to encourage participation. The authors propose several ways to reduce the costs of experimentation: 1) starting with a smaller number of participants and then increasing participation until obtaining the desired precision of results (sequential experimentation); 2) undertaking several experiments simultaneously and using the same control group, with joint use of experimental treatments by monitoring several observational variables of interest to different stakeholders; and 3) exploiting economies of scale by using the same infrastructure (specialists, field staff, and facilities) for executing different experiments.

Another challenge of regulatory experimentation is to generate valid learning, i.e., correctly linking cause and effect. This is often associated with methodological challenges since a regulatory experiment does not have controlled

⁸ Several empirical studies of experiments have highlighted such distributional effects (e.g., Castan Broto & Bulkeley, 2013; Doherty, Klima, & Hellmann, 2016). For instance, in their survey of urban climate change experiments, Castan Broto and Bulkeley (2013) find that environmental justice was a pronounced concern in 25 % of the cases.

laboratory conditions. Issues related to the causality of evaluated effects (internal validity) will probably arise, the extent of which primarily hinges on the quality of the control group. Another issue may be the universality of effects (external validity), which depends on treatment randomisation. In addition, other factors such as local institutions – which are by nature context-specific and rather constant over time – can affect the outcomes of experiments. Finally, learning must involve all stakeholders and evaluations, and publications must show the results of the experiment even if they do not show socially desired results.

Regulatory experimentation also faces legal challenges. Specific legal questions may arise with respect to concrete experiments – hence, these challenges will differ depending on the experimental setup and the respective areas of law. For the German context, BMWi (2019) provides some specific examples of legal obstacles such as the prohibition for doctors to treat patients exclusively via communication media. In the following, we consider more general legal requirements that are to be taken into account. These general legal aspects will be – to varying degrees – relevant in many legal systems.

Such general challenges prominently result from the principle of the rule of law. The rule of law can prohibit regulatory experimentation in densely regulated fields of law because deviations of the existing legal and regulatory framework might not be permitted. For example, Missling, Lange, Michaels, and Weise (2016) assert that German and European Energy Law do not permit temporal and geographical exceptions from the general rules. Every experiment with innovative rules in this field of law accordingly would require the adaptation of the respective laws by the legislator.

The rule of law also stipulates that significant decisions, i.e., those of substantial weight for the commonwealth, require parliamentary approval (BVerfGE 47, 46, 78f.). Essential questions regarding legal policy consequently have to be regulated in a formal law adopted by the parliament. Such essential questions mainly occur when a regulation encroaches in the exercise of a fundamental right (i.e., a subjective right, as guaranteed by the German Basic Law). The requirement of parliamentary law to authorize administrative actions thus can constitute an obstacle for flexible experimentation by the administration.

In cases of potential encroachments on fundamental rights, there might also exist a “prohibition of delegation” regarding the authorization of the administration to enact statutes or regulations (Hoffmann-Riem, 2005). The administration in such a case is not permitted to initiate regulatory experiments without additional parliamentary laws and respective procedures. Even if such a law exists, general formulations to authorize regulatory experiments might be regarded as insufficient to comply with the principle of legal certainty, which is an element of the rule of law. In the German context, general clauses that allow for some regulatory leeway are generally admissible. However, a law conferring powers to government and administration must be sufficiently defined and limited in content and purpose. Basic limitations for the authorities’ actions have to be defined in the law (BVerfGE 108, 52, 75f.). The required clarity and

determinedness of these limitations depend on the intensity of potential encroachments of fundamental rights.

This means on the one hand that laws that leave too much leeway for the administration to experiment, based on their own discretion, might in some cases not be permissible. On the other hand, the principle of legal certainty also implies that the law defines the objectives pursued with the regulation in a comprehensible way (Jarass & Pieroth, 2014). When the legislator adopts new laws to facilitate regulatory experimentation, the inclusion of a provision regarding the (policy) objective of the intended experiments might be considered.

Independently of implications for the requirement of parliamentary approval, legal experimentations have to consider if they infringe basic rights. For instance, legal experimentations must also consider the principle of equality, which prohibits arbitrary unequal treatment. It gets more stringent, the more (groups of) individuals are concerned, and it leaves more leeway for legislation when it targets the design of general living conditions (BVerfGE 88, 87, 96f.).

Accordingly, the principle is less demanding, if the unequal treatment of diverse groups of persons is unintended, but only factual circumstances are treated unequally. In such cases, reasonable consideration can justify a difference in treatment. If it is justifiable, the legislature may in principle also deviate from rules itself has enacted (Kahl, Hilpert, & Kahles, 2016). Overall, it is considered rather unlikely that regulatory experiments would fail because it is not possible to justify an unequal treatment (Missling et al., 2016). The requirement to balance the objective of the regulatory experiment and potentially affected rights however illustrates the need for a clear definition of the goals of a regulatory experiment.

6. CONCLUSION

The starting premise for this paper was that in the case of sustainable development, the outcomes of policies and regulations are even more uncertain than usual as greater changes are required. Indeed, sustainable development is a complex process with a high degree of uncertainty since it involves many intertwined dimensions, giving rise to trade-offs and unintended side effects. The ex-ante cost-benefit analysis thus becomes difficult to carry out as it rarely involves one policy or regulatory measure, but rather a sequence of measures which all dependent on side effects and behavioural responses.

Through trial and error, reflexive governance systems can improve policies and regulations over time and break path dependencies. Thus, learning is key to reflexive governance systems. Reflexive governance comprises both vertical and horizontal learning processes. This paper adds to the recent suggestions to embed experimentation as an integral part of governance structures to obtain the knowledge needed for sustainable development.

We have presented an encompassing overview of several concepts related to experimentation with policies and regulations in the social sciences literature and showed some similarities, as well as important differences. Building on this, our contribution is a novel approach to what we broadly

refer to as regulatory experimentation. Ideally, a regulatory experiment meets the four general conditions: 1) It is possible to use the policy to test clear-cut hypotheses regarding its effects. This can either be the explicit intention from the regulator's side or an implicit cause of the policy design. 2) In order to test this hypothesis, the experiment design addresses the fundamental problem of causal inference, i.e., in some way or another a control group that did not receive a certain treatment is present, and ideally, the subjects are randomly selected to the treatment and control group. 3) The alternative institutional setup involves cooperation between different societal actors, of which the regulator is an important part. 4) Finally, the experiment enables some form of monitoring and learning.

For all four general conditions, we formulated an analytical framework containing a total of 22 variables. Some of them are binary, and some with open questions, which can be used to characterise regulatory experiments more precisely. We summarised these variables in an analytical

framework to generally apply to all types of experiments. We showed how empirical work using this tool can provide insights that enable the improved design of regulatory experiments. Here the limitations of our paper become clear. The illustrative application of the analytical framework is only based on three cases of regulatory experiments. Future research should thus apply this framework to larger samples to be able to formulate more general conclusions. Additionally, we analysed the cases using publicly available documents. Future research should use additional data sources, e.g., qualitative interviews to obtain further detailed information on the respective experiments. The more experiments are sketched out in this way, the more knowledge researchers and practitioners will gain about the factors that make experiments more or less successful in inducing learning processes. This makes our approach the first suggestion for an encompassing categorisation of all design aspects of experiments that need to be considered when operationalising the concept as an integral part of reflexive governance structures.

REFERENCES

1. Abbott, K. W. (2017). Orchestrating experimentation in non-state environmental commitments. *Environmental Politics*, 26(4), 738-763. <https://doi.org/10.1080/09644016.2017.1319631>
2. Ansell, C. K., & Bartenberger, M. (2016). Varieties of experimentalism. *Ecological Economics*, 130, 64-73. <https://doi.org/10.1016/j.ecolecon.2016.05.016>
3. Banerjee, A. V., & Duflo, E. (2009). The experimental approach to development economics. *The Annual Review of Economics*, 1, 151-178. <https://doi.org/10.1146/annurev.economics.050708.143235>
4. Bauknecht, D., Heinemann, C., Stronzik, M., & Schmitt, S. (2015). Austesten von regulatorischen Innovationen – das Instrument der Regulatorischen Innovationszone. *Energiawirtschaftliche Tagesfragen*, 65(7), 61-64. Retrieved from <https://www.oeko.de/oekodoc/2371/2015-535-de.pdf>
5. Beck, U., Giddens, A., & Lash, S. (Eds.). (1994). *Reflexive modernization: Politics, tradition and aesthetics in the modern social order*. Cambridge, England: Polity Press.
6. Benz, A., & Dose, N. (Eds.). (2010). *Governance - Regieren in komplexen Regelsystemen: Eine Einführung*. Wiesbaden, Deutschland: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92325-3>
7. Bernstein, S., & Hoffmann, M. (2018). The politics of decarbonization and the catalytic impact of subnational climate experiments. *Policy Sciences*, 51(2), 189-211. <https://doi.org/10.1007/s11077-018-9314-8>
8. BMWi (Bundesministerium für Wirtschaft und Energie). (2018). *Reallabore - Testräume für Innovation und Regulierung*. Retrieved from <https://www.bmw.de/Redaktion/DE/Dossier/reallabore-testraeume-fuer-innovation-und-regulierung.html>
9. BMWi (Bundesministerium für Wirtschaft und Energie). (2019). *Freiräume für Innovationen - Das Handbuch für Reallabore* (Germany, Berlin). Retrieved from <https://www.bmw.de/Redaktion/DE/Publikationen/Digitale-Welt/handbuch-fuer-reallabore.html>
10. Bos, J. J., & Brown, R. R. (2012). Governance experimentation and factors of success in socio-technical transitions in the urban water sector. *Technological Forecasting and Social Change*, 79(7), 1340-1353. <https://doi.org/10.1016/j.techfore.2012.04.006>
11. Briglauer, W., Vogelsang, I., & Camarda, E. (2017). *Path dependencies versus efficiencies in regulation: Evidence from 'old' and 'new' broadband markets in the EU*. <https://doi.org/10.2139/ssrn.3050661>
12. Campbell, D. T. (1997). The experimenting society. In W. N. Dunn (Ed.), *The experimenting society: Essays in honor of Donald T. Campbell* (Policy Studies Review Annual, 11, 35-68). New Brunswick, Canada: Transaction Publishers.
13. Candel, J. J., & Biesbroek, R. (2016). Toward a processual understanding of policy integration. *Policy Sciences*, 49(3), 1-21. <https://doi.org/10.1007/s11077-016-9248-y>
14. Castán Broto, V., & Bulkeley, H. (2013). A survey of urban climate change experiments in 100 cities. *Minnesota Journal of Law, Science and Technology*, 9(2), 92-102. <https://doi.org/10.1016/j.gloenvcha.2012.07.005>
15. Crouch, C., Schröder, M., & Voelzkow, H. (2009). Regional and sectoral varieties of capitalism. *Economy and Society*, 38(4), 654-678. <https://doi.org/10.1080/03085140903190383>
16. Damelang, A., Haupt, A., & Abraham, M. (2018). Economic consequences of occupational deregulation: Natural experiment in the German crafts. *Acta Sociologica*, 61(1), 34-49. <https://doi.org/10.1177/0001699316688513>
17. Doherty, M., Klima, K., & Hellmann, J. J. (2016). Climate change in the urban environment: Advancing and achieving resiliency. *Environmental Science and Policy*, 66, 310-313. <https://doi.org/10.1016/j.envsci.2016.09.001>
18. Dryzek, J. (1987). *Rational ecology: Environment and political economy*. New York, NY: Basil Blackwell.
19. Erdmann, G. (2005). Innovation, time and sustainability. In M. Weber, & J. Hemmelskamp (Eds.), *Towards environmental innovation systems* (pp. 195-207). Heidelberg, Germany: Springer. https://doi.org/10.1007/3-540-27298-4_11

20. Evans, J., & Karvonen, A. (2014). "Give me a laboratory and I will lower your carbon footprint!" - Urban laboratories and the governance of low-carbon futures. *International Journal of Urban and Regional Research*, 38(2), 413-430. <https://doi.org/10.1111/1468-2427.12077>
21. Farrelly, M., & Brown, R. (2011). Rethinking urban water management: Experimentation as a way forward? *Global Environmental Change*, 21(2), 721-732. <https://doi.org/10.1016/j.gloenvcha.2011.01.007>
22. FCA (Financial Conduct Authority). (2015). *Regulatory sandbox*. Retrieved from <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf>
23. FCA (Financial Conduct Authority). (2017). *Regulatory sandbox lessons learned report*. Retrieved from <https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf>
24. FCA (Financial Conduct Authority). (2020). *Regulatory sandbox*. Retrieved from <https://www.fca.org.uk/firms/regulatory-sandbox>
25. Feindt, P. H., & Weiland, S. (2018). Reflexive governance: Exploring the concept and assessing its critical potential for sustainable development. Introduction to the special issue. *Journal of Environmental Policy and Planning*, 20(6), 661-674. <https://doi.org/10.1080/1523908X.2018.1532562>
26. Fitjar, R. D., & Huber, F. (2015). Global pipelines for innovations: Insights from the case of Norway. *Journal of Economic Geography*, 15(3), 561-583. <https://doi.org/10.1093/jeg/lbu017>
27. Fredriksen, K. (2018). *Does occupational licensing really impact incomes? - The German crafts case* (ifh Working Paper No. 11). Retrieved from <https://www.econstor.eu/bitstream/10419/191853/1/ifh-wp-11-2017-update.pdf>
28. Führ, M., & Bizer, K. (2007). REACH as a paradigm shift in chemical policy - responsive regulation and behavioural models. *Journal of Cleaner Production*, 15(4), 327-334. <https://doi.org/10.1016/j.jclepro.2005.10.003>
29. Führ, M., Dopfer, J., & Bizer, K. (2018). Simulation games in the regulatory impact assessment - Simulation of the EIA Amending Directive 2014/52/EU. *ELNI Review*, 2018(1), 17-24. Retrieved from https://www.elni.org/fileadmin/elni/dokumente/Archiv/2018/Heft_1/elni_2018-01_Fuehr.pdf
30. German Federal Government. (2016). *Deutsche Nachhaltigkeitsstrategie* (Neuaufgabe 2016, Kabinettsbeschluss). Retrieved from <https://www.bundesregierung.de/resource/blob/992814/730844/3d30c6c2875a9a08d364620ab7916af6/deutsche-nachhaltigkeitsstrategie-neuaufgabe-2016-download-bpa-data.pdf>
31. German Federal Statistical Office. (2016). *Statistisches Bundesamt*. Retrieved from <https://www.destatis.de/DE/Publikationen/Thematisch/UnternehmenHandwerk/Handwerkszaehlung/UnternehmenPersonenUmsatz.html>
32. Gisselquist, R., & Niño-Zarazúa, M. (2015). What can experiments tell us about how to improve government performance? *Journal of Globalization and Development*, 6(1), 1-45. <https://doi.org/10.1515/jgd-2014-0011>
33. Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory - Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.
34. Greenstone, M. (2009). Toward a culture of persistent regulatory experimentation and evaluation. In D. Moss, & J. Cisternino (Eds.), *New perspectives on regulation* (pp. 111-125). Cambridge, MA: The Tobin Project.
35. Grillitsch, M., & Nilsson, M. (2015). Innovation in peripheral regions: Do collaborations compensate for a lack of local knowledge spillovers? *The Annals of Regional Science*, 54, 299-321. <https://doi.org/10.1007/s00168-014-0655-8>
36. Grillitsch, M., Martin, R., & Sroolec, M. (2017). Knowledge base combinations and innovation performance in Swedish regions. *Economic Geography*, 93(5), 458-479. <https://doi.org/10.1080/00130095.2016.1154442>
37. Hoffmann-Riem, W. (2005). Gesetz und Gesetzesvorbehalt im Umbruch: Zur Qualitäts-Gewährleistung durch Normen. *Archiv des öffentlichen Rechts*, 130(1), 5-70. <https://doi.org/10.1628/000389105780423145>
38. Hoogma, R., Kemp, R., Schot, J., & Truffer, B. (2002). *Experimenting for sustainable transport: The approach of strategic niche management*. London, New York: Routledge.
39. Howe, K. (2004). A critique of experimentalism. *Qualitative Inquiry*, 10(1), 42-61. <https://doi.org/10.1177/1077800403259491>
40. Huitema, D., Jordan, A., Munaretto, S., & Hildén, M. (2018). Policy experiments: Core concepts, political dynamics and impacts. *Policy Sciences*, 51(2), 143-159. <https://doi.org/10.1007/s11077-018-9321-9>
41. Imbens, G. W., & Woolridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5-86. <https://doi.org/10.1257/jel.47.1.5>
42. ISGAN. (2019). *Innovative regulatory approaches with focus on experimental sandboxes* (ISGAN Annex 2 Smart Grid Case Studies). Retrieved from http://www.iea-iscan.org/wp-content/uploads/2019/05/ISGAN_Casebook-on-Regulatory-Sandbox-A2-1.pdf
43. Jarass, H. D., & Pieroth, B. (2014). *Grundgesetz für die Bundesrepublik Deutschland: GG* (Aufl. 13). München, Deutschland.
44. Kahl, H., Hilpert, J., & Kahles, M. (2016). *Experimentierklauseln im Energierecht: Zur rechtlichen Beurteilung von Sondervorschriften für Pilot - Und Demonstrationsvorhaben* (Stiftung Umweltenergierecht, Würzburg, Diskussionspapier No. 20). Retrieved from https://stiftung-umweltenergierecht.de/wp-content/uploads/2016/03/WueBerichte_20_Experimentierklauseln_im_Energierecht_2016-03-10.pdf
45. Kangas, O., Jauhiainen, S., Simanainen, M., & Ylikännö, M. (2019). *The basic income experiment 2017-2018 in Finland: Preliminary results* (Reports and Memorandums of the Ministry of Social Affairs and Health 2019:9). Retrieved from <https://julkaisut.valtioneuvosto.fi/handle/10024/161361>
46. Kangas, O., Simanainen, M., & Honkanen, P. (2017). Basic income in the Finnish context. *Intereconomics*, 52(2), 87-91. <https://doi.org/10.1007/s10272-017-0652-0>
47. Karvonen, A., & van Heur, B. (2014). Urban laboratories: Experiments in reworking cities. *International Journal of Urban and Regional Research*, 38(2), 379-392. <https://doi.org/10.1111/1468-2427.12075>
48. Kela. (2019). *Basic income experiment 2017-2018*. Retrieved from <https://www.kela.fi/web/en/basic-income-experiment-2017-2018>
49. Kerber, W. (2005). Applying evolutionary economics to public policy: The example of competitive federalism in the EU. In K. Dopfer (Ed.), *Economics, evolution and the state: The governance of complexity* (pp. 296-324). Cheltenham, England: Edward Elgar. <https://doi.org/10.4337/9781845428020.00021>

50. Kerber, W., & Eckardt, M. (2007). Policy learning in Europe: The open method of co-ordination and laboratory federalism. *Journal of European Public Policy*, 14(2), 227-247. <https://doi.org/10.1080/13501760601122480>
51. Kleihauer, S., Führ, M., & Schenten, J. (2019). Marktchancen für "nachhaltigere Chemie" durch die REACH-Verordnung - Am Beispiel globaler Lieferketten in der Textil- und Sportartikel-Industrie. *Sofia Studien zur Institutionenanalyse*, 19(1). Retrieved from https://www.sofia-darmstadt.de/fileadmin/Dokumente/Studien/2019/Netzversion_Schenten_SuSport_2019-04-18.pdf
52. Koch, A., & Nielen, S. (2017). Ökonomische Wirkungen der Handwerksnovelle 2004: Ergebnisse einer Kontrollgruppenanalyse. *Perspektiven der Wirtschaftspolitik*, 18(1), 72-85. <https://doi.org/10.1515/pwp-2017-0004>
53. Lergetporer, P., Ruhose, J., & Simon, L. (2018). *Entry barriers and the labor market outcomes of incumbent workers: Evidence from a deregulation reform in the German crafts sector* (Institute of Labor Economics Working Paper Series, IZA DP No. 11857). Retrieved from <https://www.iza.org/publications/dp/11857/entry-barriers-and-the-labor-market-outcomes-of-incumbent-workers-evidence-from-a-deregulation-reform-in-the-german-crafts-sector>
54. Li, P. (2017). Horizontal learning. In H. Bathelt, P. Cohendet, S. Henn, & L. Simon (Eds.), *The Elgar Companion to innovation and knowledge* (pp. 392-404). Northampton, England: Edward Elgar Publishing. <https://doi.org/10.4337/9781782548522.00033>
55. Li, P.-F. (2014). Horizontal versus vertical learning: Divergence and diversification of lead firms in the Hangji toothbrush cluster, China. *Regional Studies*, 48(7), 1227-1241. <https://doi.org/10.1080/00343404.2012.709610>
56. Liedtke, C., Baedeker, C., Hasselkuß, M., Rohn, H., & Grinewitschus, V. (2015). User-integrated innovation in sustainable livinglabs. An experimental infrastructure for researching and developing sustainable product service systems. *Journal of Cleaner Production*, 97, 106-116. <https://doi.org/10.1016/j.jclepro.2014.04.070>
57. Löher, J. A., & Schneck, S. (2018). *Potenziale der Reallaborforschung für die Wirtschaftspolitik* (Denkpapier No. 18-01, IfM Bonn). Retrieved from https://www.ifm-bonn.org/publikationen/denkpaepiere/publikationendetail/?tx_ifmstudies_publicationdetail%5Bpublication%5D=644&cHash=6554d5c413dbccbd9b05b03a60008547
58. Loorbach, D., & Rotmans, J. (2010). The practice of transition management. Examples and lessons from four distinct cases. *Futures*, 42(3), 237-246. <https://doi.org/10.1016/j.futures.2009.11.009>
59. Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25(3), 17-38. <https://doi.org/10.1257/jep.25.3.17>
60. Maaß, V. (2003). *Experimentierklauseln für die Verwaltung und ihre verfassungsrechtlichen Grenzen*. Berlin, Deutschland: Duncker & Humblot.
61. Malmberg, A. (1996). Industrial geography: Agglomeration and local milieu. *Progress in Human Geography*, 20(3), 392-403. <https://doi.org/10.1177/030913259602000307>
62. Mayntz, R. (2004). *Governance Theory als fortentwickelte Steuerungstheorie?* (MPiFG Working Paper Series, 04(1)). Retrieved from <https://ideas.repec.org/p/zbw/mpifgw/041.html>
63. Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey, & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie. VS Verlag für Sozialwissenschaften* (pp. 601-613). Weinheim, Deutschland: Beltz Verlag. https://doi.org/10.1007/978-3-531-92052-8_42
64. McDermott, R. (2002). Experimental methods in political science. *Annual Review of Political Science*, 5, 31-61. <https://doi.org/10.1146/annurev.polisci.5.091001.170657>
65. 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(2), 161-187. <https://doi.org/10.1007/s11077-017-9276-2>
66. Meppem, T., & Gill, R. (1998). Planning for sustainability as a learning concept. *Ecological Economics*, 26(2), 121-137. [https://doi.org/10.1016/S0921-8009\(97\)00117-1](https://doi.org/10.1016/S0921-8009(97)00117-1)
67. Missling, S., Lange, H., Michaels, S., & Weise, M. (2016). *Gutachterliche Stellungnahme zum Instrument einer Regulatorischen Innovationszone* (Report commissioned by the Ministry of the Environment, Climate Protection and the Energy Sector in the region of Baden-Württemberg). Retrieved from <https://um.baden-wuerttemberg.de/de/energie/versorgungssicherheit/smart-grids/gutachten-riz/>
68. Modell, S., Jacobs, K., & Wiesel, F. (2007). A process re(turn)? Path dependencies, institutions and performance management in Swedish central government. *Management Accounting Research*, 18(4), 453-475. <https://doi.org/10.1016/j.mar.2006.12.001>
69. Morton, R., & Williams, K. C. (2008). Experimentation in political science. In J. M. Box-Steffensmeier, J. M. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199286546.003.0014>
70. Munaretto, S., & Huitema, D. (2012). Adaptive comanagement in the Venice Lagoon? An analysis of current water and environmental management practices and prospects for change. *Ecology and Society*, 17(2), 1-16. <https://doi.org/10.5751/ES-04772-170219>
71. Nevens, F., Frantzeskaki, N., Gorissen, L., & Loorbach, D. (2013). Urban transition labs. Co-creating transformative action for sustainable cities. *Journal of Cleaner Production*, 50, 111-122. <https://doi.org/10.1016/j.jclepro.2012.12.001>
72. Ofgem. (2018). *What is a regulatory sandbox?* Retrieved from <https://www.ofgem.gov.uk/publications-and-updates/what-regulatory-sandbox>
73. Orcutt, G. H., & Orcutt, A. G. (1968). Incentive and disincentive experimentation for income maintenance policy purposes. *The American Economic Review*, 58(4), 754-772. Retrieved from <https://www.jstor.org/stable/1815531>
74. Park, R. E. (1929). The city as social laboratory. In T. V. Smith, & L. D. White (Eds.), *Chicago: An experiment in social science research*. Chicago, IL: University of Chicago Press.
75. Parodi, O. (2019). Wider eine Engführung des Reallabor-Konzepts. *Ökologisches Wirtschaften*, 34(2), 8-9. <http://dx.doi.org/10.14512/OEW340208>
76. Pielle, R. A. (2007). *The honest broker: Making sense of science in policy and politics*. Cambridge, England: Cambridge University Press. <https://doi.org/10.1017/CBO9780511818110>

77. Radaelli, C. M. (2009). Measuring policy learning: Regulatory impact assessment in Europe. *Journal of European Public Policy*, 16(8), 1145-1164. <https://doi.org/10.1080/13501760903332647>
78. Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin III, F. S., Lambin, E. F., ... Foley, J. A. (2009). A safe operating space for humanity. *Nature*, 461, 472-475. <https://doi.org/10.1038/461472a>
79. Rocle, N., & Salles, D. (2018). Pioneers but not guinea pigs': Experimenting with climate change adaptation in French coastal areas. *Policy Sciences*, 51(2), 231-247. <https://doi.org/10.1007/s11077-017-9279-z>
80. Rostam-Afschar, D. (2014). Regulatory effects of the amendment to the HwO in 2004 in German craftsmanship. *Empirical Economics*, 47, 1067-1101. <https://doi.org/10.1007/s00181-013-0773-7>
81. Runst, P. (2018). Does the deregulation of occupational licensing affect the labor market participation of migrants in Germany? *European Journal of Law and Economics*, 45(3), 555-589. <https://doi.org/10.1007/s10657-018-9583-x>
82. Runst, P., & Thomä, J. (2020). Does occupational deregulation affect in-company vocational training? *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie)*, 240(1), 51-88. <https://doi.org/10.1515/jbnst-2018-0059>
83. Runst, P., Thomä, J., Haverkamp, K., & Müller, K. (2018). A replication of 'Entry regulation and entrepreneurship: A natural experiment in German craftsmanship. *Empirical Economics*, 56(6), 2225-2252. <https://doi.org/10.1007/s00181-018-1457-0>
84. Schäpke, N., Stelzer, F., Bergmann, M., Singer-Brodowski, M., Wanner, M., Caniglia, G., & Lang, D. J. (2017). *Reallabore im Kontext transformativer Forschung: Ansatzpunkte zur Konzeption und Einbettung in den internationalen Forschungsstand* (IETSR Discussion Papers in Transdisciplinary Sustainability Research). Retrieved from <https://www.econstor.eu/handle/10419/168596>
85. Schneidewind, U., & Singer-Brodowski, M. (2013). *Transformative Wissenschaft. Klimawandel im deutschen Wissenschafts-und Hochschulsystem*. Marburg, Deutschland: Metropolis Verlag.
86. Schwarting, G. (2003). *Von der Experimentierklausel zur Standardöffnung - Ein neuer Weg der Vorbereitung von Rechtsvorschriften?* Retrieved from <https://www.uni-speyer.de/files/de/Studium/Lehrende/Schwarting/aufsatz8.pdf>
87. Simon, D., Korniek, R. L., Straßheim, H., Jarzebski, S., Hengstermann, C., Hausmann, J., & Streiter, F. (2018). *Reallabore - Eine neue Form der experimentellen Politikberatung* (Stiftung Mercator GmbH and Wissenschaftszentrum Berlin für Sozialforschung GmbH). Retrieved from https://www.stiftung-mercator.de/media/downloads/3_Publikationen/2018/Maerz/Reallabore_Online.pdf
88. Sonntag, N., & Lutter, M. (2018). Wer profitiert vom Meisterzwang? Die Reform der Handwerksordnung als natürliches Experiment zur Prüfung der Theorie beruflicher Schließung. *SozW*, 69(3), 213-251. <https://doi.org/10.5771/0038-6073-2018-3-213>
89. Tosun, J., & Lang, A. (2017). Policy integration: Mapping the different concepts. *Policy Studies*, 38(6), 553-570. <https://doi.org/10.1080/01442872.2017.1339239>
90. van den Bosch, S. J. M. (2010). *Transition experiments: Exploring societal changes towards sustainability* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/1765/20714>
91. Vang, J., & Chaminade, C. (2007). Cultural clusters, global-local linkages and spillovers: Theoretical and empirical insights from an exploratory study of Toronto's film cluster. *Industry and Innovation*, 14(4), 401-420. <https://doi.org/10.1080/13662710701523942>
92. Voss, J.-P., Kemp, R., & Bauknecht, D. (2006). Reflexive governance: A view on an emerging path. In J.-P. Voss, D. Bauknecht, & R. Kemp (Eds.), *Reflexive governance for sustainable development* (pp. 419-437). Cheltenham, England: Edward Elgar. <https://doi.org/10.4337/9781847200266>
93. Voytenko, Y., McCormick, K., Evans, J., & Schliwa, G. (2016). Urban living labs for sustainability and low carbon cities in Europe: Towards a research agenda. *Journal of Cleaner Production*, 123, 45-54. <https://doi.org/10.1016/j.jclepro.2015.08.053>
94. Waas, T., Hugé, J., Verbruggen, A., & Wright, T. (2011). Sustainable development: A bird's eye view. *Sustainability*, 3(10), 1637-1661. <https://doi.org/10.3390/su3101637>
95. WBGU. (2016). *Humanity on the move: Unlocking the transformative power of cities* (Flagship Report 2016). Retrieved from <https://www.wbgu.de/en/publications/publication/humanity-on-the-move-unlocking-the-transformative-power-of-cities>
96. WCED. (1987). *Our common future* (Report of the UN World Commission on Environment and Development). Retrieved from https://sswm.info/sites/default/files/reference_attachments/UN%20WCED%201987%20Brundtland%20Report.pdf
97. Weber, M. (1968). *Economy and society: An outline of interpretative sociology*. New York, NY: Bedminister Press.
98. Westley, F., Antadze, N., Riddell, D. J., Robinson, K., & Geobey, S. (2014). Five configurations for scaling up social innovation. *The Journal of Applied Behavioral Science*, 50(3), 234-260. <https://doi.org/10.1177/0021886314532945>
99. Zwiener, H. S. (2017). *Essays on the German labor market* (Doctoral dissertation, Humboldt University of Berlin). Retrieved from <https://edoc.hu-berlin.de/bitstream/handle/18452/18439/zwiener.pdf?sequence=1>

APPENDIX

Table A.1. Short profile of the Finnish basic experiment

Category	Variable	Characteristic	Basic income Finland	
Testing explicit hypotheses	Aim	Gathering scientific information	X	
		Testing policy options		
		Testing a pilot project	X	
		Promoting innovations		
	Object of the RE	Does the example describe a specific RE or is it a framework for several experiments?	Specific RE because a specific policy option is tested	
		What is the object of the RE?	Partial basic income	
	SDG orientation	Which SDGs does the RE concern?	SDG 1, SDG 3, SDG 8, SDG 10	
Controllability of RE-process		Formal (scientific standards)	X	
		Informal (unintended implementation of the RE and limited possibility of control)		
Test of theories/hypotheses	Is the experiment used to test a theory or hypothesis?	Yes. E.g., the effect of basic income on employment.		
Interaction between different actors	Impulse for the RE	Bottom-up		
		Top-down	X	
	Participating actors at the beginning of the RE	Homogeneous		
		Heterogeneous	X	
	Change in the composition of actors	Dynamic		
Static		X		
Unknown				
Causality	Geographical scope	Is the RE deliberately limited in space?	No	
		What is the geographical scope of the RE?	Finland	
	Time frame	Is the RE limited in time? If yes, for how long?	Yes, the experiment is limited for two years (2017 & 2018)	
	Target group	Which groups are affected by the RE?	Individuals between 25 and 58 that received labor market subsidy or basic unemployment allowance.	
	Control group	Presence of a control group? If yes, how is the control group defined? (different time, space or individuals)	Yes, explicitly (different individuals).	
	Randomization	Are units of observations selected randomly?	Yes	
	External validity of results		Features of the RE that make transferability of results difficult (e.g. institutional, geographical or temporal factors)	Institutional specificities of Finland and its social security system.
			Is the experiment designed in a way that allows/simplifies transferability results?	Yes, at least within Finland (through randomization).
		Is it possible to replicate the experiment? Are there already replications of the experiment in different contexts?	No	
		Are there other projects/measures/etc., which test similar hypotheses?	Similar experiments in the US and Canada.	
Monitoring and learning processes	Conducting evaluation	Who evaluates the RE?	Kela in cooperation with different scientific facilities	
	Information collected	Which information is collected?	Information about employment, income, well-being, attitudes towards basic income and experiences with bureaucracy	
	Public costs of the RE	Total costs of the RE	Unknown	
		Long-term coverage of costs	Yes, for the time frame of the experiment.	
		Measures to reduce costs	No	
	Type of learning		Epistemic learning (scientific knowledge)	Yes. Effects of the basic income model on employment, well-being, etc.
			Political learning (knowledge that affects the preferences and goals of political actors)	Yes. A pilot project was carried out.
			Social learning (knowledge that affects the preferences and goals of societal actors)	Yes. Attitudes towards basic income.
			Interactive learning (changes in the behavior of actors regarding information, communication, and cooperation)	Yes
			Entrepreneurial learning (changes in learning processes in firms that enable innovations)	No
Availability of results	Who can access the results of evaluations?	Public		
Publications	Are there any publications related to the RE?	Yes > 5 (still increasing)		

Table A.2. Short profile of regulatory sandboxes for financial services in the United Kingdom

Category	Variable	Characteristic	Regulatory sandboxes FCA UK
Testing explicit hypotheses	Aim	Gathering scientific information	
		Testing policy options	
		Testing a pilot project	
		Promoting innovations	X
	Object of the RE	Does the example describe a specific RE or is it a framework for several experiments?	Framework for several experiments
		What is the object of the RE?	Innovations in the financial sector
	SDG orientation	Which SDGs does the RE concern?	SDG 8, SDG 9
Controllability of RE-process	Formal (scientific standards)		
	Informal (unintended implementation of the RE and limited possibility of control)	X (formal implementation of the framework but no randomization procedures or specific control group)	
Test of theories/hypotheses	Is the experiment used to test a theory or hypothesis?	No	
Interaction between different actors	Impulse for the RE	Bottom-up	
		Top-down	X
	Participating actors at the beginning of the RE	Homogeneous	
		Heterogeneous	X
	Change in the composition of actors	Dynamic	X (new firms in every cohort)
Static			
Unknown			
Causality	Geographical scope	Is the RE deliberately limited in space?	No
		What is the geographical scope of the RE?	UK
	Time frame	Is the RE limited in time? If yes, for how long?	No
	Target group	Which groups are affected by the RE?	Authorized firms, unauthorized firms that require authorization and technology businesses in the UK
	Control group	Presence of a control group? If yes, how is the control group defined? (different time, space or individuals)	No
	Randomization	Are units of observations selected randomly?	No. Firms need to apply in order to use the experimental framework. Participation is based on different eligibility criteria mainly related to innovation.
	External validity of results	Features of the RE that make transferability of results difficult (e.g., institutional, geographical or temporal factors)	Institutional factors of the UK (e.g., strong financial sector, specific regulations), temporal factors (due to limiting the test period to 6 month difficult to cover long-term effects)
		Is the experiment designed in a way that allows/simplifies transferability results?	No
		Is it possible to replicate the experiment? Are there already replications of the experiment in different contexts?	No
		Are there other projects/measures/etc., which test similar hypotheses?	Yes. E.g., a regulatory sandbox in Denmark.
Monitoring and learning processes	Conducting evaluation	Who evaluates the RE?	FCA, firms
	Information collected	Which information is collected?	Firms need to summarize the outcomes of their tests. FCA (at least for the first two cohorts): test completion, the share of firms receiving investments following test, number of applicants, compliance with standard safeguards, limitations of testing, etc.
	Public costs of the RE	Total costs of the RE	Unknown. (As the regulator provides the framework for testing and firms test themselves, we expect costs for the regulator to be rather low).
		Long-term coverage of costs	Unknown
		Measures to reduce costs	No
	Type of learning	Epistemic learning (scientific knowledge)	No
		Political learning (knowledge that affects the preferences and goals of political actors)	Yes. Learning about how to regulate tested innovations in the financial sector.
		Social learning (knowledge that affects the preferences and goals of societal actors)	No
		Interactive learning (changes in the behavior of actors regarding information, communication, and cooperation)	Yes. Interactions between firms, the regulator, and the customers.
		Entrepreneurial learning (changes in learning processes in firms that enable innovations)	Yes. The main purpose of the regulatory sandboxes.
	Availability of results	Who can access the results of evaluations?	Internet information and some publications of the FCA.
Publications	Are there any publications related to the RE?	Yes > 5	

Table A.3. Short profile of the German crafts deregulation

Category	Variable	Characteristic	Crafts reform Germany	
Testing explicit hypotheses	Aim	Gathering scientific information		
		Testing policy options		
		Testing a pilot project		
		Promoting innovations	X	
	Object of the RE	Does the example describe a specific RE or is it a framework for several experiments?		Specific RE
		What is the object of the RE?		Licensing requirement for self-employment in the German crafts
	SDG-orientation	Which SDGs does the RE concern?		SDG 8.1: "Sustain per capita economic growth. SDG 8.3: "Support productive activities, decent job creation, entrepreneurship [...] encourage the formalization and growth of micro-, small-, and medium-sized enterprises
	Controllability of RE-process	Formal (scientific standards)		X (given certain assumptions it is a natural experiment)
Informal (unintended implementation of the RE and limited possibility of control)				
Test of theories/hypotheses	Is the experiment used to test a theory or hypothesis?		No	
Interaction between different actors	Impulse for the RE	Bottom-up		
		Top-down	X	
	Participating actors at the beginning of the RE	Homogeneous		X
		Heterogeneous		
	Change in the composition of actors	Dynamic		
		Static		X
Unknown				
Causality	Geographical scope	Is the RE deliberately limited in space?	No	
		What is the geographical scope of the RE?	Germany	
	Time frame	Is the RE limited in time? If yes, for how long?	No. But a reversal now seems likely	
	Target group	Which groups are affected by the RE?	Self-employed craftsmen, possibly also craftsmen employees and consumers of crafts services	
	Control group	Presence of a control group? If yes, how is the control group defined? (different time, space or individuals)	Yes, implicitly. Different individuals. Still fully- and partially regulated crafts occupations	
	Randomization	Are units of observations selected randomly?	In theory no, but can in practice have happened. It was the political intension that both occupations considered hazardousness and/or providing a significant contribution vocational training in Germany should remain regulated. The minutes of the negotiations do however also provide evidence for interest group lobbying	
	External validity of results	Features of the RE that make transferability of results difficult (e.g. institutional, geographical or temporal factors)		Germany has a long tradition for occupational licensing in the crafts. The German-specific vocational training system is dependent on the number of companies with a Meister title
			Is the experiment designed in a way that allows/simplifies transferability results?	Yes. Removing occupational licensing schemes can easily be done by a decision from the relevant regulatory body
		Is it possible to replicate the experiment? Are there already replications of the experiment in different contexts?		No replications exist
		Are there other projects/measures/etc., which test similar hypotheses?		Yes. National occupational licensing schemes have been steadily on the rise since the Second World War, in both Europe and the US. About 30% of the American workforce is employed in professions subject to such regulations. A number of studies evaluate the economic effects of licensing practices, in particular in the US

Table A.3. Short profile of the German crafts deregulation (Continued)

<i>Category</i>	<i>Variable</i>	<i>Characteristic</i>	<i>Crafts reform Germany</i>
Monitoring and learning processes	<i>Conducting evaluation</i>	Who evaluates the RE?	No one has been officially granted the task. But several scientists have exploited the German crafts case to assess the economic effects of occupational licensing
	<i>Information collected</i>	Which information is collected?	Market entry, market exit, share of migrants, incomes, employment, in-company vocational training
	<i>Public costs of the RE</i>	Total costs of the RE	0
		Long-term coverage of costs	No
		Measures to reduce costs	No
	<i>Type of learning</i>	Epistemic learning (scientific knowledge)	Yes
		Political learning (knowledge that affects preferences and goals of political actors)	Yes. One of the leading parties (CDU) now favors a reversal of the 2004 reform.
		Social learning (knowledge that affects preferences and goals of societal actors)	No
		Interactive learning (changes in the behavior of actors regarding information, communication and cooperation)	No
		Entrepreneurial learning (changes in learning processes in firms that enable innovations)	No
	<i>Availability of results</i>	Who can access the results of evaluations?	Everyone. Results are publicly available
<i>Publications</i>	Are there any publications related to the RE?	Yes > 10	