Governance of Experimentation: a framework for evaluating transition experiments for urban flood risk management

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Laura Johanna van Buggenum

Student number: 4139631

To be defended in public on August 20, 2018

Graduation committee

First Supervisor   : Dr. M.L.C. de Bruijne, Organisation & Governance
Second Supervisor   : Prof.Dr.Ir. I.R. van de Poel, Ethics/Philosophy of Technology
External Supervisor : Dr.Ir. N. Frantzeskaki, Dutch Research Institute for Transitions

Preface

Almost 11 months ago I started laying the foundations of what would later become this Master's Thesis report in Stockholm, Sweden. It has been a challenging and sometimes tedious process, but also a very inspiring and rewarding one. The defending of this master thesis marks the end of my time at the TU Delft and, although I look forward to see what the future will hold, it is a bittersweet goodbye.

There are a lot of people without whom it would not have been possible to complete this work. I would first and foremost like to thank my first supervisor Mark de Bruijne for his excellent guidance and willingness to read and correct a substantial number of versions this report. Second, I would like to thank Niki Frantzeskaki and Ibo van de Poel for their invaluable advice and support.

I owe special thanks to Pat for all his help in Antwerp and for showing me around the neighborhood. Similarly, I would like to thank Georgi and Jana for all their help in Burgas. I furthermore owe a major thanks to Tsanislava, who selflessly offered to help a stranger by translating information to and from Bulgarian.

Most importantly, I would like to thank my friends and family, who were always there for me for moral support, practical assistance and some much-needed coffee breaks. Without them, I would not be where I am today.

Last, but not least, I would like to thank you, the reader, for taking the time to read this report. I hope you will find it interesting!

Laura van Buggenum
Delft, July 2018

Executive Summary

Flood risks are expected to increase in the coming years as a result of ongoing climate change. Especially cities, with their high population density and impervious surfaces, are sensitive to these changes. The current water system, consisting of hard-engineered solutions, is unable to adequately respond to these challenges. Transformational change of the system is therefore desired. Transforming such a complex system, however, is challenging, with highly normative and constantly changing requirements to account for. Small-scale transition experiments can offer a solution by allowing for the implementation of innovative solutions without creating a lock-in. The governance of such experiments is thought to be highly influential on the outcomes. However, this topic is still sorely under researched.

This master thesis therefore sets out to address this gap in the literature by designing an evaluation method for transition experiments. The main research question asked in this thesis is:

How can the governance of transition experiments be evaluated in relation to the experiment's success, focusing on flood risk management?

This research departs from a tentative evaluative scheme based on empirical research. This evaluative scheme is designed as a logic model, structuring input, activities, output and outcomes of experiments. In addition to the scheme's features, the relationship between these features is a second element that is assessed by such a logic model. By aligning this empirical scheme with literature on related topics, the scheme can be adapted to fit the research focus. The criteria found in the literature are combined and operationalized based on example operationalization found in the literature. The final operationalized framework makes use of 31 indicators to analyze and evaluate transition experiments. Examples of such indicators include stakeholder awareness of flood risks, mechanisms for collaboration, built capacities and resource efficiency. The high variety of the indicators increases the comprehensiveness of the evaluative scheme, which allows for a more thorough evaluation of the transition experiments.

In order to validate this framework, it is applied to two real-world transition experiments. Case selection takes place based on the 'diverse-case method' that aims to achieve variance in order to increase the representativeness of the cases selected. Both semi-structured interviews and document analysis are used for data collection. Systematic coding and thematic analysis are then performed to analyze the data. By comparing the case findings to the other case as well as the evaluative scheme, patterns with a high degree of external validity can be identified.

Firstly, the importance of learning-by-doing for achieving social change is emphasized in both cases. Much of the transition experiments' outcomes focus on this type of change, which consists of, for example, built capacities, sense of responsibility and shared values and perspectives. Most of this social change could only be observed in stakeholders that were actively involved in the design and execution of the experiment. The social change observed in stakeholders that were passively involved in the experiment was much smaller. Secondly, for the experiment outcomes to "spillover" in the wider neighborhood, there appears to be a relation between observing peers contributing to transformational change and the willingness to become more proactive in realizing this change. Empowering stakeholders can therefore not only have a positive impact on the experiment participants, but also on stakeholders that were not directly involved in the experiment. Thirdly, in both cases, stakeholders reported that the application of good governance principles, such as transparency, responsiveness and inclusion, helped build trust and commitment. These principles therefore appear to contribute to successful processes of collaboration and co-creation. Based on the research findings, some statements about the

validity of the evaluative scheme could be made as well. Overall, one can state that the method of evaluation adapted in this research has been successful in identifying governance mechanisms and their impact on the transition experiments. The logic model design of the framework turned out to be useful in structuring the analysis and identifying patterns and relations. The features themselves could indeed almost all be observed to play a role within the experiments. However, the size of the evaluative scheme is a major point of concern. It allows for broad evaluation, but limits the opportunities for in-depth evaluation. Given the time and effort needed to apply the evaluative scheme, one can question if the evaluation efforts are proportional to the, often small-scale, experiments to be evaluated. Furthermore, it seems doubtful that the framework will allow for cross-case comparison between a large number of transition experiments as applying the framework to even just a single case takes a significant amount of time and effort.

The first set of recommendations is therefore related to the evaluative scheme. Further research into the applicability of the evaluative scheme is desired. The continued application of the scheme to real-world cases can help refine the framework's features and identify features that should be added or excluded from the scheme. Next, some interesting avenues for research into the topic of the governance of transition experiments are identified. The relationship between hands-on experience and good governance and the experiment outcomes needs to be further investigated. The impact of input features on the governance and outcomes of the experiment is a second interesting avenue for future research. In terms of policy recommendations, it should be emphasized that more research into the topic is needed. However, some tentative recommendations can be made. Firstly, there appears to be a relation between the involvement and responsibilities of the experiment participants and the degree of social change that can be achieved, both for the experiment participants as well as their less-involved peers. If social change is a goal of the experiment, it can therefore be recommended to design the experiment in such a way that the groups one wishes to empower gain hands-on experience with the design and execution of the experiment. Second, ensuring a transparent, inclusive and responsive process of experimentation is thought to contribute to stakeholder commitment and satisfaction. Including these good governance principles in the design and execution of the experiment can therefore be recommended.

Contents

List of Figures

List of Tables

List of Appendix Figures

List of Appendix Tables

List of Acronyms

EEA	European Environment Agency
FRM	Flood Risk Management
MLP	Multi-layer Perspective
NBS	Nature-based Solutions
TM	Transition Management

Governance of Experimentation
A framework for evaluating transition experiments for urban flood risk management

Research Report

1. Introduction

"For all the challenges we face, the growing threat of climate change could define the contours of this century more dramatically than any other."
- *Barack Obama, 2015*

1.1. The need for sustainable transitions

The quote above was part of a speech delivered by then president of the United States Barack Obama at the 2015 United Nations Climate Change Conference (White House Office of the Press Secretary, 2015). Yet despite these words of warning, the emission of greenhouse gases, thought to be the main driver of climate change, continued to increase over the past years and is projected to increase further in the years to come (Le Quéré et al., 2017). As a result of this continued emission of greenhouse gases, commonly accepted climate models sketch a future of extreme heat, drought, storms and rainfall as well as rising sea levels (NASA, 2018). This poses a clear threat for cities specifically, as the high built-up density of cities magnifies the impact of climate change by, for example, retaining heat in the building materials and hindering surface water absorption (Gill, Handley, Ennos, & Pauleit, 2007; Heusinkveld, Steeneveld, van Hove, Jacobs, & Holtslag, 2014). In addition to facing the threat of climate change, cities are dealing with increased urbanization. This is especially evident in Europe, where well over 80% of the population is projected to live in cities by 2050 (European Commission, 2018). This uncoordinated urbanization further amplifies the negative impact of climate change by increasing the built-up density, putting strain on infrastructures, such as the sewage system, and increasing the number of casualties in case of a disaster (Doorn, 2016; Hegger et al., 2016; Kabisch et al., 2016).

As a response to this threat of climate disasters, the European Commission adopted a climate change adaptation strategy in 2013 in which promoting local 'climate-proofing' plays a key role (European Commission, 2013b). In its accompanying impact assessment, a rather bleak picture of the state of European climate change adaptation is sketched. At the time of the assessment, only half of the nation states had adopted a climate change adaptation strategy, of which only two had developed detailed methods and indicators and of which none had considered externalities, such as social issues (European Commission, 2013a). A report published by the European Environment Agency (EEA) in 2016 found that many European cities had carefully started to design and implement climate change adaptation strategies, but an even larger number had not yet done so (European Environment Agency, 2016). In the same report, the EEA emphasizes the need to go beyond 'Band-Aid' solutions and instead focus on transforming entire urban systems. Transforming such a system, however, is highly challenging, with many uncertainties, complexities and economic, technological and institutional barriers to be dealt with (Loorbach & Shiroyama, 2016). A growing number of scholars and policymakers therefore argue that the process of transformation is an incremental one of trial and error, rather than a one-time adjustment (e.g. European Environment Agency, 2016; Geels, 2002; Rotmans, Kemp, & Van Asselt, 2001; Voß, Smith, & Grin, 2009).

The importance of small-scale transition experiments is often emphasized in this process of transformation as these experiments contribute to generating valuable insights in the impact of a wide range of technical and social innovations without creating a lock-in (Schot & Geels, 2008). Learning from these experiments and the consequent broadening and upscaling of these experiments are thought to be important drivers in facilitating desirable change (Schanze, 2017; van Buuren, Ellen, & Warner, 2016; Weiland, Bleicher, Polzin, Rauschmayer, & Rode, 2017). Stakeholder participation is seen as a key element for gaining a thorough understanding of the system and its plurality of values and interests (Doorn, 2016; Du, Zhang, Wu, Moura, & Kar, 2017). Employing a collaborative learning-by-doing approach allows stakeholders to deepen

their communal understanding of the system and take incremental steps towards increasing the transforming of the system (S. Reed et al., 2015; van Popering-Verkerk & van Buuren, 2017). Additionally, participation in such experiments helps build trust, networks and guidelines for collective action that will increase the stakeholders' capacity to respond to future disturbances to the system (Béné et al., 2017; Lebel, Anderies, Campbell, & Folke, 2006).

1.2. Transitioning towards resilient systems

However, before such transition experiments can be designed, an agreement should first be reached on what 'climate-proof' systems entail. This requires answering questions such as: which climate risks can be expected? How much risk is acceptable? Which risk distribution is acceptable? And at which cost (Johnson, Penning-Rowsell, & Parker, 2007)? Given the high level of uncertainty related to the impact of climate change, answers to these questions are likely to change over time (Ahern, 2011). Additionally, many of the questions require a normative answer that will likely differ greatly per city or even per neighborhood, depending on a wide range of variables (Krieger, 2013). The concept of resilience provides a perspective on how to deal with these complex questions. It acknowledges that the dynamics and uncertainty of a system make it impossible for a system to be designed to be fail-safe and, instead, propagates the design of systems that are safe-to-fail (Ahern, 2011; van Buuren et al., 2016). In order to do so, there is not only a need for a technical infrastructure that is better equipped to deal with change and uncertainty, but also for a social and societal infrastructure, consisting of, for example, horizontal networks and empowered citizens, that has the capacity to respond to gradual changes as well as sudden shocks to the system (Reed et al., 2015; Tyler & Moench, 2012). The goal of 'climate-proofing' a city, in short, is not to create a system in equilibrium that is able to withstand climate disaster, but one that is inherently adaptable and able to anticipate and respond to different shocks to the system (Béné et al., 2017).

One of the clearest examples of a need for transformative change is the urban water system. Cities are traditionally built near water bodies, such as oceans and rivers, as these locations provided a number of advantages, including access to fresh water and maritime transport opportunities (Walsh, 2017). This makes cities very vulnerable to flooding caused by rising sea levels or rivers that can no longer facilitate increased discharge caused by heavy rainfall either locally or upstream (Hirabayashi et al., 2013). The high density of impervious surfaces such as streets and buildings in cities also increases surface water runoff during periods of heavy rainfall, which too can lead to flooding (Mentens, Raes, & Hermy, 2006). The occurrence of floods can lead to economic loss, health risks due to, for example, bacteria and parasites as well as emotional stress and, in the worst case, the loss of lives (UN Department of Economic and Social Affairs, 2004).

However, traditional flood risk management approaches often frame flood risk as a technical problem (Fortier, 2010). As a result of this dominant paradigm, solutions for reducing flood risks tend to consist of centralized, hard-engineered solutions such as storm drains, dams and dikes (Leskens, Boomgaard, Zuijlen, & Hollanders, 2013; van Popering-Verkerk & van Buuren, 2017). A characteristic of these types of solutions is that they are made up of rigid physical structures that are expensive to produce and install (Gralepois et al., 2016; Van Wesenbeeck et al., 2014). Consequently, these solutions have a high degree of path-dependency: once in place, they are very difficult to adapt or replace (Werbeloff & Brown, 2011). For example, the sewage systems traditionally used to discharge surface water would have to be designed in such a way that they can accommodate peak discharge. As peak discharge does not occur often, the system is not utilized at full capacity for the majority of time. This makes the system expensive, as well as much more difficult to maintain and operate (Wu & Wu, 2013). Additionally, due to the high uncertainty related to future climate change and urbanization, it is unsure how high this capacity should be: too high would be even more

inefficient, whereas too low would expose the city to serious flood risks, as well as additional costs if the system would have to be adapted later on (Van Wesenbeeck et al., 2014). These dynamics highlight the need for a more resilient flood risk management system that is able to accommodate this uncertainty (Alexander, Priest, & Mees, 2016).

1.3. Innovations in urban flood risk management

One innovative solution that has the potential of increasing the resilience of the water management system is incorporating nature-based Solutions (NBS) in the existing system (European Commission, 2014; United Nations Environment Programme, 2018). Nature-based solutions consist of soft-engineered measures aimed at restoring or adapting natural ecosystems within cities. This enhances the ecosystem's capacity to absorb shocks and adapt naturally to changes (Keesstra et al., 2018; Nesshöver et al., 2017). For flood risk mitigation, a number of measures is available. Permeable surfaces allow the soil to absorb rainwater and reduce surface water runoff (European Commission, 2015). Vegetation, in turn can absorb and retain groundwater, increasing the soil's capacity to absorb excess rainwater (European Commission, 2014, 2015). Floodplains and natural waterbodies can also be used to temporarily retain water in case of heavy rainfall, high sea levels and increased river discharge (European Commission, 2014, 2015; van Slobbe et al., 2013). Literature on NBS indicates that, when compared to traditional measures, NBS measures are an effective, low-cost and low-maintenance way to reduce flood risks (Saleh & Weinstein, 2016; Van Der Nat, Vellinga, Leemans, & Van Slobbe, 2016; Van Wesenbeeck et al., 2014; Vuik, Jonkman, Borsje, & Suzuki, 2016). Additionally, these types of solutions can offer supplementary benefits, such as improved air and groundwater quality, heat mitigation and increased human wellbeing (Lafortezza, Chen, van den Bosch, & Randrup, 2018; Maes & Jacobs, 2017; Xing, Jones, & Donnison, 2017). Low investment and maintenance costs as well as increased efficiency due to secondary benefits make nature-based solutions low-regret measures that can yield benefits even if they contribute to overcapacity in the water management system (European Climate Adaptation Platform, 2018; European Commission, 2015). When used in combination with traditional, hard-engineered solutions, nature-based solutions can offer the flexibility these traditional measures lack, but is badly needed to increase the resilience of the system (European Commission, 2015; van Slobbe et al., 2013).

Despite all of this theoretical potential, cities have not yet started to implement nature-based solutions on a large scale (Nesshöver et al., 2017; Schanze, 2017). Firstly, the lack of implementation in itself is a barrier for further implementation, as flood risk management actors consider the impact of nature-based solutions on reducing flood risks too unproven to be seriously considered (van Herk, Zevenbergen, Ashley, & Rijke, 2011). Secondly, implementation is limited by a lack of awareness of its opportunities or even negative perceptions of NBS, for example caused by an association with insects or dirt (Kabisch et al., 2016). Lastly, given the high building density in cities, many nature-based solutions will have to be installed on or adjacent to private property, such as roofs and façades, which requires novel forms of cooperation with private actors such as citizens, companies and organizations (European Commission, 2015). The involvement of private actors is therefore an important aspect of implementing nature-based solutions, but it is unclear how to engage and activate them. Transition experiments as described in paragraph 1.1 could be a starting point for larger-scale implementation of NBS, combining innovative water management solutions with perception management and new forms of stakeholder cooperation.

1.4. Designing successful transition experiments

However, as also already highlighted in paragraph 1.1, these transition experiments are by no means simple undertakings. Their goals go beyond testing an innovation in order to facilitate

different types of learning that can, in turn, contribute to a larger transition towards a resilient urban system (Kivimaa, Hildén, Huitema, Jordan, & Newig, 2017). The presence of these different desired outcomes signals the presence of different sub-governance arrangements, aimed at, for example, stakeholder inclusion or solution design (Alexander et al., 2016). These governance arrangements are defined by the United Nations Development Programme as: "[…] the mechanisms, processes and institutions through which citizens and groups articulate their interests, exercise their legal rights, meet their obligations and mediate their differences" (United Nations Economic and Social Council, 2006, p.3). The governance of these transition experiments is thought to play a key role in the success of these experiments in facilitating a sustainable transition (Caniglia et al., 2017; Hildén, Jordan, & Huitema, 2017).

Despite this potential, however, the actual governance of such transition experiments is still sorely under-researched (Kivimaa et al., 2017; Weiland et al., 2017). This lack of understanding poses a barrier for the design of successful transition experiments and thereby hinders the transition towards resilient urban systems. This can therefore be interpreted as a major gap in the literature on sustainability transitions. Although some work has been published on different aspects of the design of transition experiments, these studies often focus on a narrow part of individual experiments and there is limited coordination between the different studies (Hildén et al., 2017). In order to address these issues, Luederitz et al. (2017) designed an evaluation scheme for the comprehensive evaluation of transition experiments that could be a valuable tool in generating and coordinating knowledge on the design and governance of transition experiments. This scheme, however, has not been applied to any real-world cases yet and its use should still be proven. The aim of this research is therefore to generate knowledge on the validity and applicability of this evaluation method by comparing it to related fields of literature and applying it to real-world cases. In addition to providing insights on the evaluation method, findings from the case studies could provide insights in the design of transition experiments that can serve as a starting point for further research (Yin, 2012a). This should contribute to the research aim of addressing the lack of knowledge on the design and governance of transition experiments.

1.5. Research Approach

It can be concluded from the previous paragraph that the research has two goals that overlap. The main goal is to validate the evaluation scheme as designed by Luederitz et al. (2017). Embedded within this goal is a sub goal of identifying governance practices that contribute to the success of transition experiments. The degree to which this sub goal can be achieved will influence the assessment of the validity of the evaluative scheme.

The research can be separated into three different phases. The first phase can be described as theoretical validation or the design phase: based on the literature on the topic, is the evaluative scheme fit-for-purpose? A literature review will be performed focusing on a number of different aspects of the evaluation of these types of transition experiments. This should lead to insights on the applicability of the framework and could lead to adjustments to the framework. Embedded in this first phase is the operationalization of the, adjusted, framework, to prepare it for application. This operationalization will be based on the findings of the literature review.

The second phase will focus on the empirical validation of the scheme by applying it to real-world experiments. Case studies are the most appropriate research method for doing so (Zainal, 2007). In order to allow for theoretical replication of the findings and an even broader understanding of the governance and the applicability of the framework, a cross-case comparative analysis will be performed (Yin, 2012).

Based on the findings of the evaluation of the experiments, the framework itself can now be evaluated. By revisiting the aspects of the framework discussed in phase one, a structured assessment of the applicability of the different aspects of the framework can take place. This should lead to either the validation of the evaluative scheme or proposals for the improvement of the scheme for future applications.

1.6. Research Questions

Based on the aim and goal of the research, the following research question has been formulated:

> How can the governance of transition experiments be evaluated in relation to the experiment's success, focusing on flood risk management?

As this research question is exploratory and qualitative in nature, no formal hypothesis to be proven or disproven can be formulated. Instead, the relation between a specific evaluation method and the quality of the research outcomes are to be evaluated. In order to do so, a number of sub-questions need to be answered before the main research question can be answered:

> 1. How can the evaluation of transition experiments be structured?
> 2. What are the characteristics of transition experiments to be included in the evaluation?
> 3. How can the characteristics be operationalized to allow for evaluation?
> 4. How does the governance of transition experiments contribute to their success?

It should be noted that, in order to increase the readability of the sub-questions, the term 'transition experiments' has been used as a collective term which includes this research's focus on resilient flood risk management through nature-based solutions.

1.7. Research Paper Outline

Introduction	Chapter
Research introduction	1
Theoretical validation and evaluation design	
Sub question 1: theoretical validation of evaluation method	2
Sub question 2: theoretical validation of evaluative scheme's features	3
Sub question 3: design of evaluation criteria	4
Sub question 3: design of evaluation indicators and measurement scales	5
Empirical evaluation	
Case study design	6
Sub question 4: individual case analysis	7
Sub question 4: cross-case analysis and framework validation	8
Conclusion	
Main question: research conclusion and discussion of results	9

Table 1.1: Research paper outline

2. Evaluating Transition Experiments

Transition experiments are the object of research in this report. Understanding how these transition experiments are defined and embedded in theory will provide a clearer insight into the goals and dynamics of the experiments. In addition, an appropriate logic model for the evaluation of transition experiments will be identified. Next, the first conceptualization of this logic model for evaluating transition experiments by Luederitz et al. (2017) will be discussed and, using this logic model, the success of transition experiments will be defined. The chapter will end with an assessment of the logic model and the identification of challenges that need to be addressed before the framework can be applied.

2.1. Theoretical frame transition experiments

In order to gain a thorough understanding of transition experiments, it is important to understand the broader theoretical context in which literature on transition experiments is embedded. Two concepts related to socio-technical transitions are often used to make sense of transition experiments: the Multi-level Perspective (MLP) and Transition Management (TM) (e.g.; Berkhout *et al.*, 2010; Van De Meene, Brown and Farrelly, 2011; Bos and Brown, 2012; Caniglia *et al.*, 2017; Kivimaa *et al.*, 2017; Weiland *et al.*, 2017). The former builds on the understanding that a transition towards a more sustainable system takes place in three different but interacting levels (Geels, 2002). The socio-technical regime is the level consisting of stabilized systems governed by settled rules. The socio-technical landscape consists of more general cultural and political paradigms and interrelates to the socio-technical regime. The technological niche is where innovation takes place (Geels, 2002, 2011). As the socio-technical regimes are stable, adapting or replacing such a regime in case of system failure is highly challenging (Berkhout et al., 2010). Multi-level Perspective argues that the process of deepening, broadening and upscaling of experiments taking place in the niche level can 'break through' existing barriers and disrupt the socio-technical regime, forcing a regime change (Geels, 2002; Kivimaa et al., 2017). Transition Management builds on this multi-level perspective and investigates processes of deepening, broadening and scaling up that lead to these so-called regime changes. Transition Management emphasizes the manageability of these processes and stresses the need for decision makers to link long-term visions to these short-term experiments to ensure the experiments enhance each other and function as stepping stones towards this more sustainable socio-technical regime, instead of serving as mere Potemkin villages (Schot & Geels, 2008). Governments are thought to play an important role in formulating these long-term visions, mobilizing actors and guiding these transitions (Rotmans et al., 2001). Nevertheless, it should be noted that grassroots experiments can still be linked to long-term visions, as their existence is often a, subconscious, result of developments in the regime and landscape layers (Forrest & Wiek, 2014). Additionally, TM is not focused on finding a solution for a particular system failure, but rather on an explorative design process, experimenting and learning from different designs which all seek to disrupt existing regimes and result in transitions (Rotmans & Loorbach, 2008). In addition to learning about technologies and their functioning, like in MLP technological regimes, TM is also concerned with the social aspects of transitions, encouraging actors to collaborate to facilitate social learning, the building of networks and the creation of joint sustainability visions (Schot & Geels, 2008).

2.2. Definition of transition experiments

Integrating the insights from both concepts, one can conclude that the goal of transition experiments is to generate knowledge about technical and social innovation in order to promote a sustainable transition (Weiland et al., 2017). A very general definition formulated by Van den Bosch and Rotmans (2008) defines these types of experiments as "innovation projects with a

societal challenge as a starting point for learning aimed at contributing to a transition" (Van Den Bosch and Rotmans, 2008, p. 13). A more widely used definition of transition experiments is the definition by Berkhout et al. (2010), who define these experiments as: "planned initiatives that embody a highly novel socio-technical configuration likely to lead to substantial (environmental) sustainability gains." (Berkhout et al., 2010, p.262). Although this definition does narrow down the focus of transition experiments, it fails to include the element of learning that is present in TM literature. Neither of the definitions includes the aspect of multi-stakeholder inclusion. However, stakeholder participation and cross-organizational participation are described as an important characteristic of these experiments as they contribute to many of the goals defined in TM literature, such as social learning and network creation (Kivimaa et al., 2017; Luederitz et al., 2017; Voytenko, Mccormick, Evans, & Schliwa, 2016). Multi-stakeholder collaboration is therefore deemed a key element of transition experiments and should be included in the definition. The definition by Van den Bosch and Rotmans (2008) will therefore be adapted to reflect the experiments' focus on sustainability as well as the inclusion of stakeholders.

> *"Sustainability transition experiments are innovation projects with a societal challenge as a starting point for learning, aimed at contributing to a sustainability transition by employing a multi-stakeholder approach."*
>
> Adapted from Van den Bosch and Rotmans (2008, p. 13)

2.3. Evaluation of transition experiments

Transition experiments are often complex, consisting of many different sub-goals and governance features. An instrument that is considered to be useful in structuring these complex sets of often interdependent components is the logic model framework (Cooksy, Gill, & Kelly, 2001; Julian, 1997; McLaughlin & Jordan, 2015; Savaya & Waysman, 2005). The logic model framework is commonly applied to programs, defined by Newcomer, Hatry and Wholey (2015, p.7) as: "[…] a set of resources and activities directed toward one or more common goals, typically under the direction of a single manager or management team.". Although the first part of the definition of programs matches the definition of transition experiments, the second part is in direct contrast with the multi-stakeholder approach of transition experiments. Placing an emphasis on "*typically",* however, one can argue that this definition does not explicitly exclude distributed, networked control of the project. As a result, it can be argued that transition experiments can be interpreted as a type of program and can therefore be evaluated by using the logic model framework (Forrest & Wiek, 2014; Luederitz et al., 2017).

The basic premise of the logic model is the understanding that an experiment reaches desired outcomes by performing activities and making use of resources, while operating within a specific context (McLaughlin & Jordan, 2015). The framework consists of four main categories: inputs, activities, outputs and outcomes (Cooksy et al., 2001; McLaughlin & Jordan, 2015; Savaya & Waysman, 2005). *Inputs*, by some authors also described as resources, consist of components that enable the experimental activities such as financial and human resources (McLaughlin & Jordan, 2015; Savaya & Waysman, 2005). Similarly, absence of these inputs can be seen as barriers that limit the impact of the activities. Interpreting inputs in a broader context, more 'fuzzy' concepts like trust and community awareness can also be included in this category (Luederitz et al., 2017). *Activities* encompass the processes, actions and events that take place in the experiment and are necessary to produce the required outputs (Forrest & Wiek, 2014; Julian, 1997; Luederitz et al., 2017; McLaughlin & Jordan, 2015). *Outputs* is the term

used to describe the direct results of the activities, which can include a wide range of features. Examples of outputs are implemented technologies, number of people reached in the experiment and generated knowledge (Luederitz et al., 2017; McLaughlin & Jordan, 2015; Savaya & Waysman, 2005). Lastly, *outcomes* refer to the changes to program or experimental targets as a result of the experimental activities and outputs (McLaughlin & Jordan, 2015). Common criteria for this category are, for example, resource efficiency and social equity (Forrest & Wiek, 2014; Gibson, 2006). All of these outcome features need to be accomplished to a sufficient extent for an experiment to truly contribute towards a sustainability transition (Luederitz et al., 2017).

Inputs *What was invested?*		Activities *How was it done?*		Outputs *What was generated?*		Outcomes *What was accomplished?*

Figure 2.1: logic model of evaluation

The major strength of the logic model is the consideration of the links between these four categories (Julian, 1997). Unlike black-box evaluation, where the impact of the intervention on the outcomes is usually researched, or method-driven evaluation, where a specific method is used to guide the evaluation, the logic model makes use of so-called theory-driven evaluation (Chen, 2012). Theories on the relations between inputs, activities, outputs and outcomes, be it from literature or experiment-specific documentation and stakeholder interviews, forms the basis of the logic model (Chen, 2012). This means that these models assist in making assumptions about the relations between the different categories more explicit, allowing for a more thorough understanding of the workings of the experiment (Cooksy et al., 2001). Although theory is usually developed ex ante to guide the design of an experiment, it can also be developed ex post to identify and explain observed relationships (McLaughlin & Jordan, 2015).

2.4. Evaluative Scheme for Sustainability Transition Experiments

A generic logic model framework for evaluating these sustainability transition experiments has been designed by Luederitz et al. (2017). The framework is based on an earlier framework for evaluation that focuses on Urban Sustainability Transition Labs (Wiek, Kay, & Forrest, 2017). By analyzing 61 individual case studies detailing experiments that explicitly focused on sustainability as well as long-term transition goals, Luederitz et al. (2017, p.63) were able to design a logic model framework, capturing the features and relationships common in sustainability experiments. Two important adjustments to a typical basic logic model were made. First, the sequence of the categories has been changed to start with the outputs and outcomes and trace back from there which inputs and activities led to these outcomes (Forrest & Wiek, 2014). Second, the categories of the evaluative framework should be interpreted as parallel and interdependent as experimentation takes place in iterative cycles rather than one linear process (Luederitz et al., 2017).

Inputs	Activities	Outputs	Outcomes
Awareness	Sequence of actions	Built capacities	Socio-ecological integrity
Commitment	Sound methodology	Actionable knowledge	Livelihood sufficiency &
Expertise	Collaboration	Accountability	opportunity
Trust	Reflexivity & learning	Structural changes	Intra- & intergenerational equity
Support (incl. funding)	Transparency	Changes in physical structure	Resource maintenance & efficiency

		Changes in social structure Transferability Scalability Accounting for unintended consequences	Socio-ecological stewardship and democratic governance Precaution & Adaption

Table 2.1: Logic model of sustainability transition experiments by Luederitz et al. (2017)

The goal of the framework is to structure the analysis and comparison of a large number of cases. Given this goal, one can interpret the framework as being positivist: it is a highly-structured research method fit for the analysis of large samples. The framework elements can be measured both quantitatively and qualitatively, which allows for some variation in the operationalization of the framework and the related analysis of the transition experiments, but overall, the evaluation is highly structured.

Figure 2.2: Research philosophy

Now that a basic logic model for the evaluation of transition experiments has been conceptualized, it can be used to guide the definition of success. When searching for literature on the success of sustainability or transition experiments, only a small number of papers explicitly dealing with this topic can be found[1]. An even smaller number of papers actually sets out to define and operationalize success of these types of experiments. Most of the definitions are based on the understanding that an experiment is successful when it facilitates a transition towards a more sustainable system (Collins, Boyd, & Curzon, 2017; Forrest & Wiek, 2014; Grabs, Langen, Maschkowski, & Schäpke, 2016). Facilitating a transition towards a more sustainable system is understood as satisfying sustainability criteria (Forrest & Wiek, 2015), empowering social actors to take responsibility for dealing with sustainability problems (Feola & Nunes, 2014; Forrest & Wiek, 2014; Grabs et al., 2016) and meeting short-term experiment targets (van den Heiligenberg, Heimeriks, Hekkert, & van Oort, 2017). Gibson (2006) identified a number of established sustainability criteria that can be used to measure the success of sustainability transition experiments that also incorporates the empowerment of social actors (Forrest & Wiek, 2014, 2015). The features grouped in the outcomes category of the evaluative scheme by Luederitz et al. (2017) are based on these sustainability criteria. These features will therefore be used to assess the success of the experiment in the long term. The success of experiments in the short term, on the other hand, can be measured by looking at the outputs realized by the experiments. An initial set of output features that are considered instrumental in aiding the satisfaction of the sustainability criteria are included in the evaluative scheme for sustainability transition experiments and will serve as a starting point to measure short-term success.

---

[1] For example, a search in Scopus using the operator (Success AND "Transition Experiment" OR "Sustainability Experiment") yields only 6 results.

2.5. Assessment of the evaluative scheme

Although this empirically supported evaluative scheme provides a good starting point for the evaluation of transition experiments, it would be bad practice to apply it without critically assessing it first. In the paper introducing the evaluative scheme, the authors acknowledge the tentative nature of the scheme and invite researchers and practitioners to reflexively apply the framework and propose adjustments if needed (Luederitz et al., 2017, p. 72). A preliminary assessment of the scheme does indeed highlight some weaknesses and challenges that should be addressed. Firstly, it should be noted that the scheme is highly generic and operationalization of the criteria is required to practically apply such a generic framework to the specific nature of the experiment (Savaya & Waysman, 2005). However, the short-term criteria nor the long-term criteria included in the scheme are based on objective, quantifiable definitions of success. Instead, they are based on generalized insights from both sustainability transition literature and real-world experiments (Gibson, 2006; Luederitz et al., 2017). Operationalizing the criteria of success therefore heavily depends on the focus of the experiments and is at risk of being influenced by the perspective of the researcher or experiment participants (Collins et al., 2017; Forrest & Wiek, 2014). To accommodate the former while minimizing the impact of the latter, a review of literature on urban resilience, guided by a clear search protocol, will therefore be performed to give direction to the operationalization of the short- and long-term criteria of success in light of the experiments' focus on resilience.

Second, the statements made in section 1.4 about the lack of research on the governance of transition experiments also indicate another possible weakness of the evaluative scheme. If literature on the governance of these experiments is limited, then one can argue that an evaluative scheme based on a literature review is at risk of reflecting these shortcomings in the literature. This could lead to the exclusion of relevant governance mechanisms and could impact the richness and usefulness of the evaluation outcomes. A second literature review will therefore be performed to look for and fill in any governance gaps that might be present in the evaluative scheme. As the literature of transition experiments will provide little guidance on relevant elements to include in the evaluation, literature on other, more established governance types will be used as a reference. This should allow for the operationalization and, if needed, expansion of the existing evaluative scheme. Specifically, a literature review will be performed focusing first on flood risk management, relating to the experiments' focus on the water system. In brief, this type of governance is concerned with mitigating flood risks through collaboration, with a strong emphasis on fairness and inclusion (Alexander et al., 2016; Ansell & Gash, 2008). In relation to transition experiments, however, it is possible that flood risk management does not offer enough insights on how to facilitate transformative change, as its main focus is on risk mitigation, not transitions. In addition to flood risk management, literature on reflexive governance will therefore be included in the literature review. Reflexive governance is a governance method commonly associated with the transformative learning that transition experiments are thought to facilitate (Schot & Geels, 2008; Simmons, Giraldo, Truong, & Palmer, 2018; Sol, van der Wal, Beers, & Wals, 2017; Weiland et al., 2017). Combining the insights from flood risk management, dealing with the experiments' focus on flood risks in the public-private sphere, and reflexive governance, dealing with the experiments' focus on facilitating a transition, should lead to a more comprehensive analysis of governance mechanisms that could be evaluated using the evaluative scheme.

3. Conceptualizing the governance of success

In the previous chapter, the strategy to review literature on urban resilience, flood risk management and reflexive governance has been introduced. Such a review will be performed in order to assess and determine the desired focus of the framework for this research's goal of evaluating transition experiments with a focus on the urban water system. In this chapter, the results of this literature review will be described. First, the method of the review will be briefly introduced, followed by an in-depth exploration of the three concepts.

3.1. Method

The online abstract and citation database Scopus has been used to search for relevant literature on the three concepts. In order to ensure some degree of scientific rigor only peer-reviewed articles where selected initially. Exceptions were made in cases where the peer reviewed literature was limited or insufficient. In such cases, the search criteria were expanded to also include non-peer reviewed literature, such as conference proceedings, reports and books. The credentials of the authors, publishers or target audience were assessed to weed out any literature that might not meet academic standards. Only literature written in English was included in the review, which could lead to biased findings. However, the impact of limiting any review to literature published in the English language is still hotly debated and no definite conclusion regarding the impact of language on literature reviews has been drawn (Shea et al., 2007). Additionally, one could argue that literature on concepts like governance and urban resilience are by definition biased, as such concepts are mainly developed and applied in developed countries (Haque, Shyaka, & Mudacumura, 2017). Limiting the review to literature published in English is therefore not thought to have a significant bias when compared to all literature available on the concepts, but the likely lack of insights from developing countries is something that should be considered when interpreting the findings.

A number of strategies have been employed to find the most relevant documents. In most cases, a combination of most cited papers and most recently published papers, containing the state-of-the-art, has been included in the review. In some cases, "snowballing" has been applied if a paper that did not appear in the search results was mentioned often in the selected papers. The snowballing method was also applied to identify relevant sources for providing a theoretical frame. The full search process is explained and illustrated in Appendix A.

Each of the concepts will be analyzed in four separate sections. The first section will sketch a frame of theories and approaches that the concepts are commonly grounded in. These can be used to make sense of the concepts and their conceptualization. The second section will provide a definition of the concepts that, based on the literature as well as the research focus, is considered to be the most appropriate for this research. The third section will detail the conceptualization of the concept according to the literature reviewed. It will provide an overview of the different dimensions of the concept that will be clustered according to the logic model framework described in chapter 2. In addition to clustering the criteria, each of them will be briefly described and different aspects, or potential indicators, of the dimension will be mentioned. The last section will briefly highlight the desired short- and long-term outcomes that will determine the success of the experiment that could be concluded from the literature.

3.2. Urban Resilience

3.2.1. Theoretical frame

The foundation of most of the literature on urban resilience lies in the concept of socio-ecological systems (Ahern, 2011; Bhamra, Ab, & Burnard, 2011; Folke et al., 2010; Lebel et al., 2006; Meerow, Newell, & Stults, 2016; Pickett, Cadenasso, & Grove, 2004; Tyler & Moench, 2012). Summarizing the characteristics of these systems, as identified by Redman, Grove and Kuby (2004), one can define socio-ecological systems as complex systems of

interacting biophysical and social elements that are subject to constant change (Redman, Grove, & Kuby, 2004, p.163). Different system elements and layers are connected through feedback loops, which emphasizes the need for a holistic approach when analyzing and designing these systems (Halliday & Glaser, 2011). In practice, this means that research on elements of socio-ecological systems takes into consideration the wider context and patterns of social and biophysical interaction relating to that element (Halliday & Glaser, 2011). Next to biophysical and social elements, the analysis of socio-ecological systems can also include manmade technologies (Béné et al., 2017). It is this socio-ecological and sociotechnical system thinking that heavily influences the way urban resilience is studied and conceptualized (Lebel et al., 2006). Given the complex interrelations between society, nature and flood risk mitigation technologies identified in the introduction chapter, this systems approach provides a highly appropriate backdrop for conceptualizing resilience in the context of sustainability transition experiments.

3.2.2. Definition of Urban Resilience
Although the concept of resilience has first been introduced in the 1970's, it was not until the mid 2000s that the scientific interest in the urban resilience started to increase exponentially (Béné et al., 2017). As a result of this relatively young field of research, there is not yet a commonly accepted definition of resilience (Béné et al., 2017). Literature reviews performed by Folke (2006) and more recently by Béné et al. (2017) classify the most common definitions into three different types. The first, the engineering interpretation, defines resilience as the capacity of a system to withstand or absorb shocks while remaining within critical operational thresholds (Ahern, 2011). The second interpretation, the ecosystem interpretation, expands the engineering interpretation by also recognizing a system's ability to anticipate, accommodate or recover from a shock (Béné et al., 2017; Holling, 1973). The last interpretation is the so-called socio-ecological interpretation. It represents a broader interpretation than the previous two interpretations by including the system's capability to learn, adapt or transform as the result of a disturbance (Bhamra et al., 2011; Folke, 2006; Meerow et al., 2016; Walker, Holling, Carpenter, & Kinzig, 2004). It focuses on the technological aspect of the system as well as the ecological and social aspects. As previously described, designing flood risk defense solutions that are infallible is near to impossible, which is why the engineering interpretation is considered to be too narrow for this research. Additionally, a defining characteristic of socio-ecological systems like the water system is that they are subject to constant change. Failure to learn from previous failures and adapt to this change, increases the vulnerability of the system (Tyler & Moench, 2012). The socio-ecological definition is therefore considered to be the most appropriate for this research.

> *"Urban resilience refers to the ability of an urban system, and all its constituent socio-ecological and socio-technical networks across temporal and spatial scales, to maintain or rapidly return to desired functions in the face of a disturbance, to adapt to change, and to quickly transform systems that limit current or future adaptive capacity."*
>
> *(Meerow et al,, 2016, p. 45)*

3.2.3. Conceptualization of Urban Resilience
Nearly all papers reviewed, with the exception of Walker, Holling, Carpenter and Kinzig (2004), provided a conceptualization of resilience. These features of urban resilience have been

clustered according to the logic model framework. The features of resilient system are summarized in table 3.1.

Inputs	
Access to resources	(Olsson, Folke, & Berkes, 2004; Tyler & Moench, 2012)
Leadership	(Folke, 2006; Folke et al., 2010; Lebel et al., 2006; Olsson et al., 2004)
Trust	(Carpenter, Walker, Anderies, & Abel, 2001; Folke, 2006; Folke et al., 2002; Lebel et al., 2006; Olsson et al., 2004)
Activities	
Collaborative learning	(Carpenter et al., 2001; Folke, 2006; Folke et al., 2002, 2010; Lebel et al., 2006; Olsson et al., 2004; Pickett et al., 2004; Tyler & Moench, 2012)
Collaborative sense-making	(Meerow et al., 2016; Olsson et al., 2004; Pickett et al., 2004; Tyler & Moench, 2012)
Multi-actor decision-making	(Olsson et al., 2004; Pickett et al., 2004; Tyler & Moench, 2012)
Monitoring	(Folke et al., 2002)
Outputs	
Diversity	(Ahern, 2011; Bhamra et al., 2011; Folke et al., 2010; Tyler & Moench, 2012)
Resource Efficiency	(Bhamra et al., 2011; Tyler & Moench, 2012)
Multi-scale networks	(Ahern, 2011; Bhamra et al., 2011; Folke et al., 2002; Meerow et al., 2016; Tyler & Moench, 2012)
Multifunctionality	(Ahern, 2011; Pickett et al., 2004; Tyler & Moench, 2012)
Redundancy and Modularity	(Ahern, 2011; Tyler & Moench, 2012)
Technological Innovation	(Folke, 2006; Folke et al., 2010; Tyler & Moench, 2012)
Institutional Innovation	(Folke, 2006; Folke et al., 2010; Tyler & Moench, 2012)
Capacity to self-organize	(Carpenter et al., 2001; Lebel et al., 2006; Olsson et al., 2004)
Outcomes	
Equity	(Ahern, 2011; Folke et al., 2010; Meerow et al., 2016)
Adaptive planning and design	(Ahern, 2011; Bhamra et al., 2011; Carpenter et al., 2001; Folke, 2006; Folke et al., 2002, 2010; Lebel et al., 2006; Meerow et al., 2016; Tyler & Moench, 2012)
Facilitates sustainable transition	(Folke, 2006; Folke et al., 2010; Meerow et al., 2016; Pickett et al., 2004)

Table 3.1: Features of Urban Resilience

3.2.4. Inputs
The input category contains the features that are thought to be important enablers for achieving resilience. Stakeholders' *access to resources*, such as knowledge, financial funding and materials supports or inhibits collaboration, innovation and change. Stakeholders need access to reliable and credible information about risks, vulnerabilities and options in order to make decisions (Tyler & Moench, 2012). Additionally, funding is needed to allow stakeholders to take action (Olsson et al., 2004).

Despite the collaborative nature of building resilience, *leadership* is considered to be vital to initiate an experiment, guide interactions and resolve conflicts (Folke et al., 2010; Olsson et al., 2004). Additionally, leaders can give direction to the process and inspire and motivate stakeholders when the transition process hits a rough patch (Lebel et al., 2006).

Trust is vital in ensuring and open collaborative stakeholder process. It will legitimize the learning as well as stimulate stakeholder engagement (Olsson et al., 2004; Tyler & Moench, 2012; B. Walker et al., 2004). Trust aids mobilization and self-organizing capacity and is built up through repeated positive interaction (Lebel et al., 2006).

3.2.5. Activities
The activities category contains features that are thought to contribute to resilience as well.

Different types of *collaborative learning* are required to facilitate a transition towards a more resilient system. Learning about innovations is useful for identifying tools to facilitate a transition (Tyler & Moench, 2012). Broad learning about the system itself increases the understanding of the system, its weaknesses and opportunities for resilient transformation (Carpenter et al., 2001; Olsson et al., 2004). Social learning helps form a joint perspective on the problem and the need for change and generates a social memory that increases the communities capacity to respond to changes (Folke, 2006; Lebel et al., 2006; Olsson et al., 2004; Pickett et al., 2004).

Collaborative sense-making can be interpreted as a type of learning that focuses on combining or prioritizing actors' underlying assumptions, norms and values (Olsson et al., 2004; Pickett et al., 2004). As the understanding of what contributes to resilience and how to best achieve it is highly normative and deeply embedded in stakeholders' values, reaching an agreement on these requirements is required before changes to the system can be made (Meerow et al., 2016; Olsson et al., 2004; Tyler & Moench, 2012).

Multi-actor decision-making indicates that processes of decision-making need to include all relevant actors. Humans are understood to be part of the system and including their perspectives in the decisions made will increase the effectiveness of the system interventions (Pickett et al., 2004). The robustness of the solution will also be increased by including stakeholders with a wide variety of skills and knowledge (Olsson et al., 2004). Lastly, the legitimacy of the decisions will also be improved by including all affected stakeholder groups in the decision-making process (Tyler & Moench, 2012).

Last, *monitoring* allows for learning during a process of change and can generate early warnings when failure or disturbances threaten to occur (Folke et al., 2002).

3.2.6. Outputs

In the output category, features are included that are described as important characteristics of resilient systems. Bio-, social, physical and economic *diversity* all contribute to a wider variety of forms and behaviors (Bhamra et al., 2011). Spatial diversity reduces the likeliness that all assets of the system are affected by a shock at once (Tyler & Moench, 2012). Additionally, functional diversity allows for a more diverse response to disturbances, increasing the ability of the system to adapt and resume its functions after the occurrence of shocks (Ahern, 2011).

Resource efficiency is relevant as system that is able to perform its functions while consuming a limited amount of resources is at a lower risk of failure due to resource scarcity (Bhamra et al., 2011).

Connecting different *multi-scale networks*, like storm drains, streams and rivers will enhance the functioning of each of those networks and will ensure functionality after network disturbances (Ahern, 2011). The other way around, disconnections between or within networks increases the vulnerability of a system (Folke et al., 2002). Examples of different networks that exist in the urban system are governance networks, material flow networks and ecological and technical infrastructures (Meerow et al., 2016).

Multifunctional solutions that can perform a number of functions are spatially efficient and can make the most of the limited space available in cities (Ahern, 2011). This means that more solutions can be installed that can substitute for each other when one of them fails (Tyler & Moench, 2012). Multifunctional solutions also offer additional benefits, which can make them more economically efficient than solutions that only perform one function (Pickett et al., 2004).

Distributed systems with *redundant and modular* functions reduce the path-dependency associated with focusing on just one system or solution (Tyler & Moench, 2012). Buffer capacities and back-up functions increase the system's capacity to withstand shocks (Ahern, 2011).

Technological Innovation, consisting of new products or new ways to use existing products, can support an increase in the aforementioned characteristics and thereby increase the resilience of the system as a whole (Folke et al., 2010).

Institutional innovation which can consist of, for example, innovative governance tools, networks or frameworks can support all of the processes summarized above to contribute to an increase in social resilience (Folke et al., 2010; Tyler & Moench, 2012).

Self-organization through the formation of formal and informal networks and actor coalitions will increase the stakeholders' capacity to respond to future disturbances to the system (Olsson et al., 2004). A bottom-up transition approach will allow for a more swift and flexible process of adaptation as it does not need to rely on outside steering and support (Lebel et al., 2006; Olsson et al., 2004).

3.2.7. Outputs

Lastly, the output category contains the more long-term goals that are not necessarily a direct output of the experiment itself, but a required consequence of the outputs. Ensuring e*qual* distribution of costs and benefits will reduce the number of weak, or sensitive, parts of the system (Meerow et al., 2016).

Adaptive Planning and Design recognizes that maintaining system resilience is an ongoing process of responding, experimenting and learning-by-doing (Ahern, 2011; Bhamra et al., 2011). It encompasses an stakeholder understanding that a system is never in equilibrium and actions beyond the experiment are required (Carpenter et al., 2001; Meerow et al., 2016; Tyler & Moench, 2012).

Last, the experiment should *facilitate a sustainable transition:* As the external and internal context of a system are constantly changing, so should the system itself. A resilient system is therefore designed in such a way that it encourages stakeholders to continue to take responsibility for increasing the resilience of the system (Folke et al., 2002; Meerow et al., 2016; Pickett et al., 2004).

3.2.8. Success in Urban Resilience

In short, for experiments aiming to increase the resilience of the system, the long-term success can be measured by asking if the experiment facilitates an ongoing transition process and ensures equal opportunities for different stakeholders. Outputs that contribute to this long-term success and can be used to measure the success on the short-term are related to the changes in physical structure realized by the experiment, including if these changes increase the diversity and multifunctionality of system elements, evidence-supported knowledge on technological and social innovations as well as the empowerment of actors to self-organize. The activities mainly focus on collaborative processes, as these are thought to contribute to the aforementioned empowerment of the relevant stakeholders. Input features such as trust and access to resources are thought to be relevant enablers for the process of collaboration.

3.3. Flood Risk Management
3.3.1. Theoretical frame

The concept of governance is dominant in the literature on Flood Risk Management (FRM). Most of the papers referred to different types and definitions of governance, such as "good governance" (Alexander et al., 2016), "water governance" (Evers, Jonoski, Almoradie, & Lange, 2016) and "environmental governance" (Benson, Lorenzoni, & Cook, 2016). However, the type of governance most often used as a lens through which to view FRM is collaborative governance (Evers et al., 2016; Newig, Kochskämper, Challies, & Jager, 2016; Thaler & Levin-Keitel, 2016; van Herk et al., 2011). This type of governance is initiated by governments and aims to bring multiple stakeholders together to engage in collaborative processes dealing with

public policies or programs (Ansell & Gash, 2008). The emphasis is on joint decision-making, as collaborative governance is not about stakeholder consultation, but about reaching a stakeholder consensus (Ansell & Gash, 2008). The focus on the governance process that leads to flood risk mitigation rather than flood risk mitigation itself can be explained by the high social uncertainty surrounding flood risk appraisal and mitigation (Doorn, 2016). Different value conflicts arise when managing flood risks and context-dependent resolution of these values will determine which outcomes are considered desirable or successful (Ravesteijn & Kroesen, 2015). As a result, most of the literature reviewed did not focus on flood risk or technical solutions mitigating this risk, but rather on inputs and processes that allow stakeholders to define and explore flood risks and desirable flood risk management. This governance focus is clearly evident in both the definition of flood risk management as formulated in the next paragraph as well as the conceptualization of flood risk management and can provide targeted propositions regarding the governance of transition experiments.

3.3.2. Definition of Flood Risk Management

As a result of the governance-focused approach described in paragraph 3.3.1., the definition of FRM is rather fuzzy. Many papers did not even introduce FRM and instead immediately focused on the conceptualization of this type of governance (Geaves & Penning-Rowsell, 2016; Newig et al., 2016; Thaler & Levin-Keitel, 2016; Wilby & Keenan, 2012). Those that did introduce FRM often did so in comparison to previous water management practices. Heintz, Hagemeier-Klose and Wagner (2012) describe current FRM practices as holistic, rather than flood defense-based. Van Herk et al. (2011) describe FRM as "a transition from 'fighting against water' to 'living with water'" (p. 545). Many other authors underline this transition from defense to management (Butler & Pidgeon, 2011; Evers et al., 2016; Hartmann & Driessen, 2017; Hartmann & Spit, 2016; Krieger, 2013; Porter & Demeritt, 2012; Ward, Pauw, van Buuren, & Marfai, 2013). This new form of flood risk management consists of an integrated approach that combines structural, including hard and soft-engineered solutions, and non-structural solutions, such as spatial planning, awareness and evacuation plans, to manage flood risk (Butler & Pidgeon, 2011; Hartmann & Driessen, 2017; Hartmann & Spit, 2016; Krieger, 2013; van Herk et al., 2011; Van Wesenbeeck et al., 2014; Ward et al., 2013; Wilby & Keenan, 2012). FRM Takes into consideration the many layers of the water management system, including the social, ecological and economic aspects of FRM (Evers et al., 2016; Johnson et al., 2007; Ward et al., 2013). Integrating these different aspects of FRM, the following definition is proposed:

> *"Flood Risk Management is a governance approach aiming to reduce the likelihood and impact of flooding by implementing structural and non-structural solutions, while taking into consideration the social, ecological and economic aspects related to flooding."*
>
> Adapted from Van Wesenbeeck et al. (2014, p. 4)

3.3.3. Conceptualization of Flood Risk Management

All of the papers reviewed identified features of FRM that are thought to contribute to the successful flood risk management. Much like the features of urban resilience, the features of FRM could be categorized in the four categories of the logic model framework based on the authors' descriptions of their workings and goals. The categories and their corresponding features are displayed in Table 3.2.

Inputs	
Awareness	(Butler & Pidgeon, 2011; Evers et al., 2016; Hartmann & Spit, 2016; Heintz, Hagemeier-Klose, & Wagner, 2012; Thaler & Levin-Keitel, 2016; Wilby & Keenan, 2012)
Trust	(Benson et al., 2016; Evers et al., 2016; Kuklicke & Demeritt, 2016; Thaler & Levin-Keitel, 2016)
Access to resources	(Alexander et al., 2016; Thaler & Levin-Keitel, 2016; Wilby & Keenan, 2012)
Access to information	(Benson et al., 2016; Evers et al., 2016; Porter & Demeritt, 2012; Wilby & Keenan, 2012)
Local autonomy	(Butler & Pidgeon, 2011; Newig et al., 2016; Thaler & Levin-Keitel, 2016)
Institutional alignment	(Alexander et al., 2016; Krieger, 2013; Porter & Demeritt, 2012; Thaler & Levin-Keitel, 2016; Wilby & Keenan, 2012)
Activities	
Fairness	(Alexander et al., 2016; Johnson et al., 2007)
Transparency	(Alexander et al., 2016; Evers et al., 2016; Geaves & Penning-Rowsell, 2016; Krieger, 2013; Ward et al., 2013)
Multi-actor decision-making	(Alexander et al., 2016; Benson et al., 2016; Butler & Pidgeon, 2011; Evers et al., 2016; Geaves & Penning-Rowsell, 2016; Hartmann & Spit, 2016; Heintz et al., 2012; Johnson et al., 2007; Thaler & Levin-Keitel, 2016; van Herk et al., 2011; Ward et al., 2013)
Collaborative sense-making	(Benson et al., 2016; Evers et al., 2016; Hartmann & Driessen, 2017; Krieger, 2013; Thaler & Levin-Keitel, 2016; Ward et al., 2013)
Collaborative learning	(Benson et al., 2016; Evers et al., 2016; Newig et al., 2016; van Herk et al., 2011)
Legitimacy	(Alexander et al., 2016; Hartmann & Driessen, 2017; Hartmann & Spit, 2016; Thaler & Levin-Keitel, 2016; Ward et al., 2013)
Outputs	
Reduced flood risk	(Butler & Pidgeon, 2011; Evers et al., 2016; Hartmann & Driessen, 2017; Hartmann & Spit, 2016; Krieger, 2013; Porter & Demeritt, 2012; Ward et al., 2013)
Stakeholder empowerment	(Alexander et al., 2016; Butler & Pidgeon, 2011; Hartmann & Spit, 2016; Thaler & Levin-Keitel, 2016; Wilby & Keenan, 2012)
Horizontal networks	(Benson et al., 2016; Newig et al., 2016)
Outcomes	
Equity	(Alexander et al., 2016; Johnson et al., 2007)
Adaptive planning and design	(Heintz et al., 2012; Newig et al., 2016)

Table 3.2: Features of Flood Risk Management

3.3.4. Inputs

The first feature in the input category is *awareness*. Awareness of FRM and its importance to society increases stakeholders' willingness to participate in the design process (Evers et al., 2016; Heintz et al., 2012; Thaler & Levin-Keitel, 2016). Awareness of the need for change will also increase the acceptance of solutions or interventions (Geaves & Penning-Rowsell, 2016; Hartmann & Spit, 2016; Johnson et al., 2007). Furthermore, it will contribute to the willingness to change behaviors in order to contribute to the common goal of flood safety (Evers et al., 2016; Ward et al., 2013).

Trust is a key factor in ensuring an open and constructive design process (Evers et al., 2016; Thaler & Levin-Keitel, 2016). A high level of trust between stakeholders helps reach widely accepted agreements and builds networks for future collaboration (Benson et al., 2016). Next to trust between project stakeholders, trust in the reliability of the available information and the relevant authorities is vital in ensuring stakeholder cooperation (Kuklicke & Demeritt, 2016).

Access to resources such as funding, knowledge and time are important tools for the design and implementation of FRM (Alexander et al., 2016; Thaler & Levin-Keitel, 2016). Access to resources also ensures that monitoring and, if needed, adaptation of the design can take place after concluding the design process (Wilby & Keenan, 2012).

Access to information: although seemingly part of the access to resources feature, encompasses the stakeholders' capacity to understand and work with the information available. This requires adapting available, often scientific, information to reflect local politics, needs and requirements as well as present the information in such a way that it is accessible for stakeholders (Benson et al., 2016; Porter & Demeritt, 2012; Wilby & Keenan, 2012). A lack of access to information or access to information that fails to disclose the possible blind spots is a major barrier for FRM (Evers et al., 2016).

Local autonomy reflects the complex nature of water management that favors local, tailor-made management over top-down, centralized forms of management (Thaler & Levin-Keitel, 2016). As national governments are often unaware of local norms and values, too high a degree of top-down interference is likely to create conflicts with local actors (Butler & Pidgeon, 2011). Additionally, stakeholders might be less inclined to join a collaborative design process if they believe that they will have limited power and participation is not interesting for them (Heintz et al., 2012).

Last, *institutional alignment* is needed to facilitate local autonomy and innovation that is uninhibited by conflicting laws and regulations (Alexander et al., 2016; Krieger, 2013; Wilby & Keenan, 2012). Conflicting visions at different institutional levels can create conflicts that can slow down or even stop the design process (Porter & Demeritt, 2012). Information exchange and cooperation between different institutional levels, on the other hand, can improve the design process (Alexander et al., 2016; Wilby & Keenan, 2012). Lastly, alignment between different institutional layers is needed to secure access to resources, such as funding (Thaler & Levin-Keitel, 2016).

3.3.5. Activities

The first feature in the activities category is *transparency*. A transparent process increases stakeholders' acceptance of the processes and outcomes and provides stakeholders and citizens with the opportunity to assess the quality of these processes and outcomes (Alexander et al., 2016; Evers et al., 2016; Krieger, 2013). Transparency in terms of clear responsibilities will also increase the efficiency of the design and implementation of solutions by emphasizing accountability (Ward et al., 2013).

Next, *multi-actor decision-making* gives all relevant stakeholders the opportunity to influence the outcomes of the design process. This helps to build trust and confidence and increases the acceptability and quality of the outcomes (Alexander et al., 2016; Geaves & Penning-Rowsell, 2016; Ward et al., 2013). Factors such as the quality of the decision-making process, the range of stakeholders involved and the degree to which they can influence the decision-making process are all relevant factors to facilitate multi-actor decision-making (Alexander et al., 2016).

Collaborative sense-making is concerned with actors' underlying assumptions, norms and values. Accounting for these values in the design of FRM will make the design more responsible (Ward et al., 2013). Collaborative sense-making can guide the decision-making process and is a way to deal with the lack of an objective truth by identifying shared conceptions and interests (Benson et al., 2016; Evers et al., 2016; Hartmann & Driessen, 2017; Thaler & Levin-Keitel, 2016). It can also help identify and appoint different roles and responsibilities of the stakeholders (Krieger, 2013).

Collaborative learning will increase the stakeholders' understanding of the system, problems and solutions and will lead to more effective and robust designs (Benson et al., 2016; Evers et al., 2016; Newig et al., 2016). Learning can also be about the planning and decision-making process, which will help make these processes more efficient (Newig et al., 2016). Learning can take place internally, within a project, but learning can also take place by sharing knowledge with stakeholders from other projects (Newig et al., 2016).

Three types of *legitimacy* are relevant for FRM. Input legitimacy, including authorization, representation and accountability, is relevant for committing stakeholders to the process. Throughput legitimacy, consisting of inclusive and open processes, is needed to justify the process and outcomes (Johnson et al., 2007). Output legitimacy, summarized as the results for citizens and stakeholders, is considered to be the result of input and throughput legitimacy (Alexander et al., 2016; Hartmann & Driessen, 2017; Hartmann & Spit, 2016).

3.3.6. Outputs

The most obvious output of FRM is a *reduced flood risk.* Flood risk reduction is understood as either or both the reduced likelihood of flooding and the impact of flooding (Hartmann & Driessen, 2017; Porter & Demeritt, 2012).

Next, *stakeholder empowerment* encourages self-organization and increases the local stakeholders' abilities and willingness to deal with future FRM challenges (Butler & Pidgeon, 2011; Hartmann & Spit, 2016; Thaler & Levin-Keitel, 2016; Wilby & Keenan, 2012). Stakeholder empowerment can also take place on an individual level when stakeholders are empowered to take individual measures to protect themselves against flood risks (Alexander et al., 2016).

Last, *horizontal networks* between different local stakeholder coalitions enable cross-pollination between different FRM initiatives and assist in capturing learning (Benson et al., 2016; Newig et al., 2016).

3.3.7. Outcomes

As for the outcomes, *equity* encompasses the distribution of risks, costs and benefits in such a way that it does not contribute to inequality between different stakeholder groups (Alexander et al., 2016; Johnson et al., 2007).

Adaptive planning and design, on the other hand, emphasizes the need for experimentation to facilitate this collaborative learning and allows for the incremental design, implementation and adaptation of FRM (Newig et al., 2016). It encompasses the understanding that risk management is an ongoing process that needs to take place continuously (Heintz et al., 2012).

3.3.8 Success in Flood Risk Management

Unlike the conceptualization of urban resilience, FRM literature does not focus on the technological aspect of flood risk management. Rather, the focus seems to be on "Good Governance" of the decision-making processes. As a result, there is a strong emphasis on inputs and activities and less of a focus on outputs and outcomes. The underlying assumption appears to be that a high-quality decision-making process will automatically lead to desirable outcomes, which reduces the need to make these outcomes explicit. Regardless of this emphasis on inputs and activities, the long-term success of flood risk management activities can be measured by assessing if these measures ensure equity as well as ongoing processes of adaptation and innovation. On the short term, the most tangible measure of success is the reduction of flood risk, which is the main goal of FRM. Stakeholder empowerment and the formation of horizontal networks will provide stakeholders with the tools to ensure the adaptive planning and design, which is why they are used as indicators for the short-term success of FRM. Aspects of good governance, such as legitimacy, trust and transparency will help ensure a fair and collaborative decision-making process that should lead to the implementation of desirable solutions as well as lay the foundation for further collaboration.

3.4. Reflexive Governance
3.4.1. Theoretical Frame
The concepts most often used in the reviewed literature to make sense of reflexive governance are so-called wicked problems (Rotmans & Loorbach, 2008; Termeer, Dewulf, Karlsson-Vinkhuyzen, Vink, & van Vliet, 2016; Voss & Kemp, 2005a) and transition management (Schäpke, Omann, Wittmayer, van Steenbergen, & Mock, 2017; Van Den Bosch & Rotmans, 2008; Voss & Bornemann, 2011). The former is used to describe problems that are difficult to solve due to their complex nature and contradicting and often changing requirements (Rittel & Webber, 1973). The latter is a form of governance aimed at facilitating sustainability transitions (Rotmans et al., 2001). The relationship between wicked problems and reflexive governance becomes very clear in the conceptualization of this type of governance, as it tends to prioritize the processes of problem structuring and collaborative sense-making over the design of solutions. Reflexive governance is also embedded in transition management, which is evident in the attention for outcomes that support ongoing processes, and thus gradual transition, over fixed stopping times after which the reflective governance activity comes to an end (Voss & Bornemann, 2011). Its focus on problem structuring and facilitating ongoing transitions links Reflexive Governance to all three aspects of the object of this research: experiments, resilience and flood risk management. TM is highly relevant for both sustainability transition experiments and resilient systems. Problem-structuring and collaborative sense-making, on the other hand, are key features of flood risk management. Reflexive Governance therefore seems to be an appropriate approach to guide resilient, nature-based flood risk management experiments.

3.4.2. Definition of Reflexive Governance
The literature available on the topic of reflexive governance is fairly limited, with a Scopus search on the concept yielding just slightly over 400 results in February 2018. As a result of this still largely unexplored mode of governance, there is not yet a common definition of reflexive governance. Instead, most of the papers reviewed used broad and vague descriptions to define reflexive governance, if at all. Instead, two frames of reflexive governance could be identified. The first is concerned with so-called double loop learning (Scott, 2010). According to this definition, reflexive governance comprises mechanisms and processes that allow stakeholders to not just learn about the results of actions but also about the knowledge, values and assumptions that shape these actions and goals (Buizer, Elands, & Vierikko, 2016; Termeer et al., 2016; Van Den Bosch & Rotmans, 2008; Voss & Bornemann, 2011). The second frame is focused on so-called triple loop learning about the governance process itself (Dryzek & Pickering, 2017; Voss & Kemp, 2005a, 2006). Voss and Kemp (2005a, p.3) describe this type of reflexive governance as follows: "[Reflexive governance] comprises practices of governing that are concerned with the preconditions for their own working – one of them being the embedding of steering activities within societal development itself, the very same process that is to be governed.". Although implicit in this definition, triple loop learning is also concerned with first and second order learning about the solutions and the assumptions (Pahl-Wostl, Palmer, & Richards, 2013). An adaptation of the definition by Voss and Kemp (2005) is proposed to accommodate both frames as identified within the literature and make all learning loops explicit.

> *"Reflexive governance comprises practices of governing that encourage actors to learn about innovations, evaluate and revise their underlying assumptions, values and frames and assess the process of governance itself."*
>
> Adapted from Voss and Kemp (2005a, p.3)

3.4.3. Conceptualization of Reflexive Governance

Features of reflexive governance were identified in all of the papers, although the focus of these features differed based on the definition of reflexive governance used in the papers. For the features of reflexive governance, two themes similar to those of flood risk management could be found: context and design process. The features categorized into one of these themes are displayed in table 3.3 and discussed below.

Inputs	
Institutional alignment	(Voss & Kemp, 2006)
Social capital	(Schäpke et al., 2017)
Expertise	(Dryzek & Pickering, 2017)
Activities	
Monitoring and evaluating	(Bellamy, 2016; Buizer et al., 2016; Hendriks & Grin, 2007; Rotmans & Loorbach, 2008; Termeer et al., 2016)
Adaptive planning and design	(Bellamy, 2016; Dryzek & Pickering, 2017; Rotmans & Loorbach, 2008; Termeer et al., 2016; Voss & Bornemann, 2011; Voss & Kemp, 2005b, 2006)
Multi-actor decision-making	(Bellamy, 2016; Dryzek & Pickering, 2017; Hendriks & Grin, 2007; Rotmans & Loorbach, 2008; Voss & Bornemann, 2011; Voss & Kemp, 2006)
Collaborative learning	(Rotmans & Loorbach, 2008; Schäpke et al., 2017; Voss & Bornemann, 2011; Voss & Kemp, 2005b, 2006)
Collaborative sense-making	(Buizer et al., 2016; Hendriks & Grin, 2007; Termeer et al., 2016; Voss & Kemp, 2005b, 2006)
Reflexive foresight	(Bellamy, 2016; Voss & Kemp, 2005b, 2006)
Outputs	
Stakeholder empowerment	(Schäpke et al., 2017)
Outcomes	
Adaptive planning and design	(Bellamy, 2016; Dryzek & Pickering, 2017; Rotmans & Loorbach, 2008; Termeer et al., 2016; Voss & Bornemann, 2011; Voss & Kemp, 2005b, 2006)

Table 3.3: Features of Reflexive Governance

3.4.4. Inputs

The first category is again the input category, which contains tools that can aid the design process. *Institutional alignment* is highly relevant, as reflexive governance acknowledges that sustainable transitions are not guided by one single actor, but through interactions between different actors and institutions. Even when not directly involved in a project of experiment, these can influence the outcomes through, for example, lawmaking or funding (Voss & Kemp, 2006). Misalignment of these institutions could create a barrier for the transition.

    Social capital can be structural, like networks, or content-related, like shared values and trust. Social capital can exist within a group or between groups and facilitates smooth collaboration during the experiment as well as continued (community) collaboration after the experiment ends (Schäpke et al., 2017).

    Although *expertise* alone is considered to be too narrow to solve collective problems such as flood risk management, it can provide valuable insights and direction to a decision-making process when used in combination with lay deliberation (Dryzek & Pickering, 2017).

3.4.5. Activities

The activities category contains most of the features of reflexive governance. *Monitoring and evaluating* is relevant, as the process of experimentation is never straightforward. Rather, new or unintended tensions occur often, which need to be identified and dealt with through constant monitoring and evaluation (Buizer et al., 2016). Both the process itself, in terms of, for example,

fairness, equity and efficiency, as well as the progress and anticipated outcomes need to be observed and evaluated (Bellamy, 2016; Hendriks & Grin, 2007; Rotmans & Loorbach, 2008). This will facilitate learning as well as timely steering the process in the desired direction (Termeer et al., 2016).

The steering aspect of monitoring and evaluating takes place in the *adaptive planning and design feature*. This will give the actors a tool to deal with uncertainty, changes and windows of opportunity (Termeer et al., 2016; Voss & Kemp, 2005b). Additionally, facilitating and open-ended and adaptive design process will help generate more robust and legitimate solutions by incorporating diverse perspectives on the problem and solution as well as allow for the integration of new perspectives (Bellamy, 2016; Dryzek & Pickering, 2017; Rotmans & Loorbach, 2008).

The focus within the conceptualization of *multi-actor decision-making* is on the presence of politics. Reflexive governance acknowledges that actors' competing frames can lead to power struggles (Hendriks & Grin, 2007; Termeer et al., 2016; Voss & Bornemann, 2011). In order for the actors to reach a consensus, or a meta-consensus where at least the process of decision-making is widely accepted by the stakeholders, these politics will have to be dealt with within the decision-making process (Dryzek & Pickering, 2017; Voss & Bornemann, 2011; Voss & Kemp, 2005a).

Collaborative learning can help generate practical knowledge about the system and system interventions, both social and technical, as well as the governance design of the project or experiment (Bellamy, 2016; Rotmans & Loorbach, 2008; Schäpke et al., 2017; Voss & Bornemann, 2011). Additionally, social learning can take place to increase the awareness of stakeholders and citizens of the problem which will help commit them to the process of change (Schäpke et al., 2017; Voss & Kemp, 2006).

Collaborative sense-making is another form of learning about actors underlying values and assumptions (Buizer et al., 2016; Hendriks & Grin, 2007; Schäpke et al., 2017). This type of learning is concerned with reframing the problem by critically assessing frames as well as connecting them and convincing others to join a frame (Termeer et al., 2016; Voss & Kemp, 2006). This should lead to the design of a more widely accepted solution.

Reflexive foresight encompasses the activity of joint anticipation of different future scenarios. As many solutions will know some degree of path dependency, designing them in such a way that they perform well under different circumstances will make them more robust (Bellamy, 2016; Voss & Kemp, 2006). Reflexive foresight also entails anticipating the occurrence of externalities or parts of the system that can be indirectly affected by the solution (Voss & Kemp, 2005b).

3.4.6. Outputs
Only one feature could be identified in the outputs category. *Stakeholder empowerment* enables stakeholders to react to problems collectively as well as individually and commits them to the process of change. Indicators of stakeholder empowerments are increased power in the decision-making process and increased power over the resources as well as newly gained capacities, such as skills and knowledge (Schäpke et al., 2017).

3.4.7. Outcomes
Lastly, the only outcome feature is *adaptive planning and design*. This acknowledges that reflexive governance is an ongoing process that needs to go on beyond the project or experiment to facilitate a long-term transition towards a more desirable system (Dryzek & Pickering, 2017; Voss & Bornemann, 2011).

3.4.8. Success in Reflexive Governance
As was to be expected from literature on reflexive governance, most of the features identified can be categorized in the Activities category. What sets apart reflexive governance from urban resilience and FRM is the categorization of 'Adaptive planning and design' in both outcomes as well as activities, due to its double role in both the short- and the long-term. The characteristic that reflexive governance shares with FRM, however, is the assumption that a well-designed, reflexive governance process will automatically lead to desirable outcomes. The outcomes of such a governance process are therefore highly implicit. The only long-term outcome identified is the feature that signals if the experiment facilitates an ongoing transition. The only short-term outcome deals with the stakeholder empowerment that is thought to be the driving factor behind this ongoing transition. A number of activities, including monitoring and evaluating and employing reflexive foresight, are thought to lead to desirable outcomes that are unspecified in the literature. The input features of institutional alignment, social capital and expertise are thought to be tools for facilitating the activities.

3.5. Assessment of the evaluative scheme
Based on the findings from the literature review, some preliminary statements about the evaluative scheme by Luederitz et al. (2017) can already be made. Firstly, one can conclude the use of such a logic model framework, separating input, throughput and output is not an alien concept for governance evaluation and different logic models have, in fact, already been applied to real-world cases (e.g. Alexander et al., 2016; Forrest & Wiek, 2014; Kivimaa et al., 2017; Mees, Driessen, & Runhaar, 2014). From a theoretical perspective, this too increases the likelihood that the evaluation scheme can indeed be applied to evaluate the governance of transition experiments.

It should be noted, however, that the logic models used in these types of research were significantly smaller than the one to be used in this research, often focusing on a specific aspect of the case. This does raise some concerns about the feasibility of performing such a comprehensive evaluation. In addition to this previous statement about the already quite extensive nature of the evaluation scheme, more features can be added to the scheme based on the literature review. Although their addition to the scheme should still be supported by empirical data, it does raise questions if aiming for comprehensive evaluation is truly feasible with such a complex research focus or just leads to uncontrolled 'evaluation creep'.

It can lastly be noted that many of the scheme's features are indeed echoed in the literature on the three different concepts, which enhances its representativeness. However, that the interpretations of the features can vary per concept, which indicates that many possible indicators could be designed to assess and compare the different features. As the logic model for evaluation is already of unprecedented size, it seems unwise to include each of these possible indicators in the evaluative scheme. A next step is therefore to combine and refine the different interpretations of the features and to distill the most relevant indicators that can be used for the evaluation of transition experiments.

4. Evaluative Scheme Transition Experiments

In chapters 2 and 3, the four concepts of transition experiments, urban resilience, flood risk management and reflexive governance have been conceptualized and organized based on a logic model framework for evaluation. In this chapter, the four frameworks will be combined into one comprehensive framework. First, the relevant criteria of each of the concepts, as identified in the literature, will be compared and, where possible, combined. In chapter 5, indicators for each of the defined criteria will be designed. This should lead to a framework adapted to fit the research focus and operationalized to be applied in this research.

4.1. Conceptualization of nature-based FRM experiments for urban resilience

The first step in designing a framework for the evaluation of nature-based flood risk management experiments is combining the logic model frameworks from all four concepts. As described in chapter 2, the evaluative scheme for appraising sustainability experiments by Luederitz et al. (2017) will be used as the basis framework for evaluation. The three frameworks created in chapter 3 will be compared to this framework to see if different criteria should be combined, added or removed.

Literature on program and sustainability evaluations, however, provided little support on how to reduce the number of criteria. Suggestions made were often highly generic, such as selecting criteria that are representative of the goals of the object of evaluation (Newcomer et al., 2015) or most representative of the goals of the evaluation (Abrams, Borrini-Feyerabend, Gardner, & Heylings, 2003). Some more practical requirements could also be found, such as safeguarding ease of use, for example by selecting criteria that can be measured based on readily available information (Reed, Fraser, & Dougill, 2006), or ensuring all categories of the logic model framework are represented (Posavac, 2016). Expanding the literature search to other scientific fields dealing with criteria did not yield any silver bullet solutions either, with, for example, Design Science Research stating that criteria included in the framework should be sensitive to the environment in which the framework is being applied (Gregor & Hevner, 2013) or literature on Q Methodology stating that coverage and balance are important when selecting what to include or exclude in the research (Watts & Stenner, 2012). Some practical advice was provided by Watts & Stenner (2012), however, in their advice to broaden the criteria so they cover a wider range of criteria within one statement. Overall, however, authors seem to acknowledge the lack of guidelines for selecting criteria and a number of them recognize the researcher's discretion in selecting appropriate criteria (Abrams et al., 2003; Newcomer et al., 2015; Watts & Stenner, 2012).

A number of tools for criteria selection has therefore been applied to reduce the number of criteria to include in the case studies. An initial selection excluding all criteria that were overlapping, immeasurable or not relevant given the focus of this research could be made. Next, the more abstract goals of each of the features as identified in the previous chapters were used to search for convergence and divergence between the different criteria. Where the criteria converged, new criteria capturing the essence of the existing criteria were defined or adapted. As a stopping rule, it was decided that each of the features should have one, or at maximum two, criteria. This way, all criteria identified as relevant within the four concepts could be included in the case studies while ensuring the feasibility of the evaluations. The drawback of such a relatively unstructured approach is that it is susceptible to researcher bias, especially in the last few steps of criteria selection. By continuing to consider the research goal and focus as well as staying close to the interpretations of the different criteria as defined in the literature and continuous reflection of the selection process, the impact of researcher bias should have been reduced. A second drawback of this approach is, of course, that the broad range of criteria included in the research limits the depth in which each of these features can be evaluated. This

is something that will need to be considered when interpreting the outcomes of this research and when answering the research questions.

In the next paragraphs, this process of selection, convergence and divergence will be briefly explained per criterion. All statements about the transition experiment features are based on the paper by Luederitz et al. (2017). Statements about the resilience, FRM and reflexive governance concepts have already been discussed in chapter 3, in which one can find the references to the corresponding authors. The criteria identified for each of the features will be summarized in tables 4.1, 4.2, 4.3 and 4.4.

4.2. Combining the input features

Awareness

Both the evaluative schemes for transition experiments and flood risk management identified awareness as an important input feature. Though the framing of awareness as the stakeholders' consciousness of the need for sustainable change is similar in both concepts, the goals this awareness is thought to serve differs. For transition experiments, awareness is seen as a tool to activate a sense of urgency that will keep the experiment moving forward (Luederitz et al., 2017, p. 70). In FRM, however, this awareness is described as the intrinsic motivation to commit to the process and contribute towards change (see 3.3.4). In the former, it relates to all experiment participants, in the latter to the stakeholders that are not involved in the organization of the experiment. The common ground in both interpretations is that awareness is a characteristic that will commit stakeholders to the experiment and will drive them to continue to work towards an outcome. The criterion for awareness can therefore be defined as the existence of an awareness of local flood risks and the need for flood risk management for all possible stakeholders.

Commitment

The feature of commitment could only be explicitly identified in the transition experiment scheme and therefore does not need to be adapted. Commitment can be defined as the willingness of stakeholders to participate in the experiment (Luederitz et al., 2017, p. 70). This commitment can depend on many different factors and interests, such as direct benefits or an intrinsic motivation to help. Regardless of the motive, a high level of commitment at the beginning of the process is thought to ensure that all relevant stakeholders remain actively involved throughout the process of experimentation and increase the feeling of accountability. The criterion for commitment can therefore be defined as the level of stakeholder involvement throughout the process of experimentation.

Expertise

Expertise has been identified as an important feature in the literature on reflexive governance as well as transition experiments. The understanding of expertise and the goal of this feature are similar in the frameworks on transition experiments and reflexive governance, which indicates that there is no need for further convergence or divergence. In both cases, expertise is defined as the skills and (professional) experiences of the stakeholders involved in the experiment that can be helpful for the conducting of the experiment and can contribute towards reaching desired outcomes (see 3.4.4). For expertise, the criterion can therefore be defined as the inclusion of relevant stakeholders with the skill or expertise to design or conduct the experiment.

Trust

Trust has been identified as a relevant input feature in transition experiments, urban resilience and FRM. Much like expertise, the definitions of trust within the three concepts were largely

similar. Trust is understood as the stakeholders' willingness to work together, which includes their attitudes towards other stakeholders and the degree to which they are willing to rely on these other stakeholders (Luederitz et al., 2017, p. 70). Trust it is thought to facilitate a process of open collaboration, which aids collaborative exchange (see 3.3.4). According to literature on transition experiments and urban resilience, trust in the other stakeholders at the beginning of the experiment also contributes to stakeholder confidence in the process of experimentation and commits stakeholders to the experiment (see 3.2.4). Additionally, FRM emphasizes that a high level of trust throughout the experiment increases the chance that the stakeholders will form networks that will remain active after the experiment has ended (see 3.3.4). However, as trust is primarily classified as an input feature, this latter understanding of the role of trust will be excluded from this feature. The criterion for trust can therefore be defined as the level of trust between the stakeholders at the beginning of the experiment.

Support
Within the different concepts of transition experiments, urban resilience and FRM, there were quite some differences in terms of the description and goals of support. All three concepts, however, included resources, such as funding, information and tools, in their description of support. The transition experiment framework additionally identifies assistance from external parties in the design and execution of the experiment as relevant support (Luederitz et al., 2017, p. 70), whereas the frameworks for urban resilience and FRM indicate that support in terms of helping the stakeholders understand the available information and utilize the available tools is a relevant part of support (see 3.2.4). In terms of goals, all three concepts described that support served as a tool to reach the desired direct outcomes. FRM identified that resources can also serve to ensure some degree of continuance of the experiment after the experiment has officially ended, which means that resources also play a role in facilitating an ongoing transition process (see 3.3.4). As neither of these concepts is mutually exclusive, the original definition and goal as described in the transition experiment framework can be expanded by including assistance in operationalizing the available resources as well as expanding the timeframe beyond the experiment itself. The criterion for support can therefore be defined as the access of stakeholders to resources that are instrumental in reaching the experiment outcomes.

Institutional alignment
The feature of institutional alignment was not included in the transition experiment framework, but was included in the frameworks for both FRM and reflexive governance. As water management is primarily concerned with the public space, implementation of physical artefacts designed in the experiment will have to take place in compliance with the local laws and regulations (see 3.3.4). If these do not allow for the implementation or heavily favor specific solutions, truly independent experimentation cannot take place (see 3.4.4). As the institutional environment is an important barrier or enabler of experimentation in the case of water management, it should be included in the evaluation framework. As it does not relate to any of the existing features, it is therefore argued that institutional alignment should be included as a separate feature in the framework. The criterion for institutional alignment can therefore be defined as an intuitional environment that is supportive of the experiment.

Leadership
The second addition to the transition experiment framework is leadership. Leadership is included in the urban resilience framework and identifies the need for an entity that serves as the leader of the experiment (see 3.2.4). Roles the leader could fulfill include experiment initiation and providing inspiration. The goal of the inclusion of a leader in the experiments is to create and maintain momentum to keep the experiment moving forward. This is also why

leadership was not added to the expertise feature, as a leader, in this context, is more concerned with inspiration and motivation and an expert with more functional matters, such as knowledge and process management. The criterion for leadership is therefore set as the presence of a leader or leadership entity throughout the experiment.

Social capital
Lastly, social capital is added to the input features. Social capital is recognized within the literature on reflexive governance as an important input feature (see 3.4.4). It can consist of existing stakeholder networks and shared values. The more social capital exists between the stakeholders, the easier the process of collaboration is likely to be. Additionally, it should also increase the chances that the networks will remain active after the experiment has ended. The criterion in this case is the presence of social capital in the local environment.

Feature	Main criteria
Awareness	The experiment includes stakeholders that are aware of flood risks and their own role in change
Commitment	The experiment includes stakeholders that are committed to conducting the experiment
Expertise	The experiment includes stakeholders that have the skills and expertise to conduct the experiment
Trust	The stakeholders trust each other
Support	The stakeholders have access to support in the form of funding, information, tools and training
Institutional Alignment	The institutional environment is supportive of experimentation
Leadership	The experiment includes an inspiring leader
Social Capital	The experiment includes pre-existing social capital

Table 4.1: Input criteria

4.3. Combining the activities features
Sequence of actions
The sequence of actions feature was only mentioned in the transition experiment framework. It is concerned with the sequencing and timing of the different activities within the experiment (Luederitz et al., 2017, p. 69). In doing so, one can not only make sure that threats, opportunities and requirements are thought of well in advance, but one can also create and make use of windows of opportunity to advance the experiment. The criterion can be defined as the use of a meaningful sequence of actions for conducting the experiment.

Methodology
Methodology could also only be found in the transition experiment framework. It focuses on the use of methods within the experiment. Methods to be applied in the experiment can, for example, be comprised of methods for problem analysis, solution design, monitoring and evaluating (Luederitz et al., 2017, p. 69). The goal is to ensure the experiment reaches the desired outcome of facilitating transformational change and not just focuses on data collection and analysis or meeting a smaller, short-term goal. The criterion can be defined as the use of a sound methodology for conducting the experiment.

Collaboration
Collaboration was identified as a key feature in all four concepts. In the transition experiment framework, collaboration is defined as stakeholder inclusion, mechanisms of collaboration and the modes of interaction (Luederitz et al., 2017, p. 69). In the literature on urban resilience and

FRM, the focus of collaboration is more on the *legitimacy* of the collaboration process. That means that it is focused not so much on the how, but on the what: all relevant stakeholders should be included in the process of collaboration and have equal opportunity to influence the decision-making process if desired (see 3.2.5 and 3.3.5). Within the literature on reflexive governance, collaboration is understood as a process of politics, which means that, in order to reach a meta-consensus, the process of collaboration and decision-making needs to be agreed upon by all stakeholders (see 3.4.5). For all concepts, however, the main goal was described as the reaching of robust and widely accepted outcomes. Collaboration, according to urban resilience and FRM literature, should also lead to throughput and output legitimacy. By actively involving stakeholders in the collaboration process stakeholders should also become more empowered to continue to address sustainability issues after the experiment has ended. As collaboration consists of two distinctive features, two criteria have been formulated: all relevant stakeholders need to be able to participate in the collaboration process and collaboration mechanisms should be present as well as accepted by the stakeholders.

Reflexivity and learning
Reflexivity and learning was also included in all four concepts. In the transition experiment framework, learning was focused on the process of experimentation itself: by evaluating the process of experimentation throughout the experiment, processes can be timely adjusted or designed in order to produce the desired outcomes (Luederitz et al., 2017, p. 69). Urban resilience literature, on the other hand, describes the desired learning within an experiment as learning about the system, system externalities and tools to influence the system as well as learning about the stakeholders' values and perspectives (see 3.2.5). The ultimate goal is not only to design a solution for a specific system failure, but also to include the, converged, stakeholder values in this design. FRM and Reflexive governance represent a combination of both descriptions and goals (see 3.3.5 and 3.4.5). One could therefore conclude that the experiment should foster learning about the problem, normative solutions and tools to implement these solutions learning during the process of experimentation as well as timely steering based on these types of learning.

Transparency
Transparency was included in both the transition experiment framework and the FRM framework. In both cases, transparency is understood as the open reporting on intentions, processes and roles. It is thought to increase the accountability of the experiment participants as well as increase the understanding and acceptance of choices made throughout the experimentation process (Luederitz et al., 2017, p. 70). In FRM literature, transparency is also understood as an element of *throughput legitimacy* and is described as a good governance practice (see 3.3.5). The criterion for transparency is that open sharing of all relevant information has to take place throughout the process of experimentation.

Feature	Main criteria
Sequence of actions	The experiment is conducted according to a meaningful sequence of actions
Methodology	The experiment is conducted according to a sound methodology
Collaboration	All relevant stakeholders are included in the experiment Collaboration mechanisms that are accepted by the stakeholders are included in the experiments
Reflexivity and learning	The experiment fosters learning The experiment is responsive to lessons learned

| Transparency | The experiment ensures open sharing of all relevant information throughout the experiment |

Table 4.2: Activities criteria

4.4. Combining the output features

Built capacities

All four concepts acknowledge the importance of built capacities as the outcome of a process of experimentation or collaboration. All four concepts define built capacities as skills that allow stakeholders to act sustainably and respond to changes in the system, both individually and collectively (see 3.2.6, 3.3.6 and 3.4.6). Summarized, it is about the empowerment of stakeholders to address future sustainability challenges and continue to contribute to a transition towards a more resilient system. The criterion can therefore be defined as follows: the experiment should empower stakeholders to design and implement resilient solutions.

Actionable knowledge

Actionable knowledge was only included in the transition experiment framework. Nevertheless, it can be seen as an important feature, as it is one of the direct outcomes of the learning processes in the activities category. Evidence supported knowledge about the design of transition experiments, the problem and possible solutions and the related goals and values. The combination of this knowledge provides input for different transition pathways that can be used to guide the transition towards a more resilient system (Luederitz et al., 2017, p. 64). The criterion of actionable knowledge is therefore defined as: the experiment should produce evidence-based transition pathways.

Accountability

As already introduced in the transparency paragraph, accountability is part of the *legitimacy* requirement in the FRM literature (see 3.3.6). In both the FRM and the transition experiment frameworks, accountability is described as the stakeholders claiming ownership over the experiment: they feel responsible for the outcomes of the experiment and are committed to positive change (Luederitz et al., 2017, p. 66). Accountability is needed to ensure both the quality of the direct outcomes of the experiment and to commit stakeholders to contributing towards a transition beyond the experiment. The criterion of accountability is therefore the degree to which the experiment succeeds in building a sense of problem ownership in the stakeholders.

Changes in physical structures

Changes in physical structures were described in all concepts but reflexive governance. For transition experiments, the definition was rather broad, by mainly stating that the experiment should adjust existing technologies or add new technologies (Luederitz et al., 2017, p. 66). Urban resilience and FRM literature defined a number of additional requirements, including that the adjustment or addition needs to reduce flood risk, increase system diversity and be multifunctional. This should lead to a more resilient system as well as transformational change (see 3.2.6 and 3.3.6). For the water management system, nature-based solutions meet all of these requirements (see 1.3). The criterion for changes in physical structures is therefore that a nature-based solution needs to be implemented.

Changes in social structures

Like for changes in physical structures, only reflexive governance did not identify changes in social structures are desired outcomes of an experiment. Such changes include new horizontal networks and linkages between networks, values and behavior, second order learning, and

governance processes, third order learning (Luederitz et al., 2017, p. 66) (see 3.2.6 and 3.3.6). This differs from the actionable knowledge feature as it does not focus on the knowledge, but on the adoption of the knowledge: the changes in perception of the stakeholders. Here too, the goal is to create both a more resilient social system as well as transformational change in the practices of how water management challenges are approached. As the formation of networks is not so much a process of learning as it is of collaboration, it cannot be merged with the two types of learning involved in changing social structures. Therefore, two criteria are defined for change in social structures: new networks should emerge and changed perspectives should be observable.

Transferability
Transferability is only identified as a relevant feature in the transition experiments framework. It encompasses generalized knowledge that can be applied to experiments or projects in different contexts as well (Luederitz et al., 2017, p. 66). This feature holds a somewhat unique position in this framework, as generating this transferable knowledge through a comparative study is the goal of this case study research. Within this feature, however, it suffices that the stakeholders have considered which lessons can be transferred to other contexts and have validated these lessons

Scalability
Scalability is also only identified as a relevant feature in the transition experiments framework. It is understood as the lessons learned that are independent of the size of the experiment. In other words: the lessons can be applied regardless of the scale of the experiment or project. Luederitz et al. (2017, p. 67) warn, however, that the actual application of these lessons is beyond the scale of the experiment. For the experiment itself, it suffices that the stakeholders have considered which lessons can possibly be used to scale up the experiment. The criterion for scalability is therefore that the experiment generates knowledge about the possibility to scale up the experiment and validates this knowledge.

Accounting for unintended consequences associated with uptake
Lastly, accounting for unintended consequences associated with uptake was also only identified in the transition experiments framework, yet generating this type of knowledge was also mentioned in the reflexive governance activities category (see 3.4.5). As experiments take place within complex systems, outputs may lead to externalities within the current system or under future scenarios (Luederitz et al., 2017, p. 67). By carefully considering and anticipating these possible externalities, for example by adapting the design to mitigate these effects, the impact of these unintended consequences can be minimized. The criterion for this feature can be defined as follows: the experiment should account for externalities caused by implementation and uptake of the experiment.

Feature	Main criteria
Built capacities	The experiment empowers stakeholders to design and implement resilient solutions
Actionable knowledge	The experiment produces evidence-based transition pathways
Accountability	The experiment increases the stakeholders' sense of problem ownership
Physical changes	The experiment leads to the implementation of a nature-based solution
Social changes	The experiment encourages the formation of new, horizontal networks

	The experiment changes stakeholder attitudes towards sustainable problems
Transferability	The experiment indicates how lessons learned can be transferred
Scalability	The experiment indicates how the experiment can be up scaled
Consequences	The experiment accounts for externalities related to implementation and uptake of the experiment

Table 4.3: Output criteria

4.5. Combining the outcome features

Socio-ecological integrity

Socio-ecological integrity is included in the transition experiments framework as part of Gibson's (2006) sustainability criteria. It recognizes the interdependencies between physical, social and ecological systems and aims to harmonize these interactions. The goal of this harmonization is to make use of the natural capacity of ecosystems to regenerate and have these effects spill over into the other systems as well. Much like for the changes in the physical system, these characteristics are inherent to nature-based solutions. An expressed intention to continue to implement similar or different nature-based solutions is therefore selected as a criterion for this feature.

Livelihood sufficiency and opportunity

Livelihood sufficiency and opportunity is only included in the transition experiments framework. It encompasses the access of citizens to commodities that are needed to live a decent life, including, for example, resources that are needed for sustaining life or performing economic activities (Luederitz et al., 2017, p. 67). The former, livelihood sufficiency, is not relevant for nature-based solutions, as these types of solutions to not give or limit access to vital resources. Livelihood opportunities, however, are relevant. One could, for example, imagine that closing roads or removing parking spaces could have a significant impact on the businesses in the area if they depend on accessibility by car. The criterion for livelihood sufficiency and opportunity is formulated as: the experiment's outcomes do not negatively impact livelihood opportunity.

Intra- and intergenerational equity

This feature was included in the frameworks of transition experiments, urban resilience and FRM. In all three cases, however, it had a slightly different meaning. For urban resilience, the focus is on a balanced spread between risks, costs and benefits in order to reduce weak spots in the system (see 3.2.7). FRM recognizes a similar definition of equity, but the goal is to ensure fairness and outcome legitimacy (see 3.3.7). The transition experiments framework adds to this that all stakeholders should get equal access to decision-making processes and the opportunities for future generations to live sustainable lives should not be jeopardized (Luederitz et al., 2017, p. 68). As the latter is not particularly relevant for nature-based flood risk management, the focus of this feature will be on intergenerational equity. It adds to outputs such as built capacities and changes in social and physical structures by assessing if these outcomes truly include all stakeholders or just privileged citizens. The criterion for equity can therefore be formulated as follows: the experiment's outputs should benefit all relevant stakeholder groups.

Resource maintenance and efficiency

Resource maintenance and efficiency is, again, only included in the transition experiments framework. It is defined as reducing the input that is needed for one unit of output, by, for example, recycling or more efficient production technologies (Luederitz et al., 2017, p. 68). The focus of this feature is very much on physical resources, such as fuels and materials, and

not on resources such as money and skills. For nature-based solutions, one could, for example, consider the carbon footprint of the transport or installation of the materials used. The criterion for resource maintenance and efficiency is therefore defined as: the outputs of the experiment included materials that require minimal resources for implementation.

Socio-ecological stewardship and democratic governance
This feature was included in all four concepts and deals with both stakeholder empowerment to continue to take care of system resilience as well as the inclusion of all relevant stakeholders in these processes. It can be seen as the outcome of the changes in the social structures, as it is the result of the perception changes and network formations within this feature (see 3.2.7, 3.3.7 and 3.4.7). The goal of this feature is to facilitate an ongoing transition by empowering actors to operationalize their changed perspectives and to become more involved in traditional top-down decision making (Luederitz et al., 2017, p. 68). As the former is already captured by the socio-ecological integrity feature, the criterion for socio-ecological stewardship and governance is: the experiment's outputs should lead to an increased involvement of stakeholders in top-down decision making.

Precaution and adaption
This feature, which was also included in all four frameworks, is the outcome of the accounting for unintended consequences feature. It not involves the identification and implementation of risk-averse measures for scaling up (parts of) the experiment (Luederitz et al., 2017, p. 68). This should increase the resilience of the measure as well as the resilience of the system as a whole (see 3.2.7). The criterion for precaution and adaption are defined as follows: the experiment's outputs should lead to mitigated flood risks and the accounting for uncertainties

Feature	Main criteria
Socio-ecological integrity	The experiment encourages stakeholders to continue to implement nature-based solutions
Livelihood sufficiency and opportunity	The experiment's outputs do not negatively impact livelihood opportunity
Intra- and intergenerational equality	The experiment's outputs create benefits for all relevant stakeholder groups
Resource maintenance and efficiency	The experiment's outputs include materials that require minimal resources for implementation
Socio-ecological stewardship and democratic governance	The experiment's outputs lead to an increased involvement of stakeholders in top-down decision making
Precaution and adaption	The experiment's outputs contribute to minimizing flood risks

Table 4.4: Outcome criteria

5. Operationalization of the Evaluative Scheme

Before the criteria defined in chapter 4 can be applied they need to be further operationalized. In this chapter, the operationalization of the criteria will be explained. First, a causal relationship diagram will be designed to gain a better understanding of the possible relations between the different criteria.

5.1. Causal relationships within the evaluative scheme

As mentioned before, the interaction between the high number of criteria included and the limited time available for doing the case study research means that the criteria cannot be analyzed in-depth. Rather, the most representative indicators for each criterion will have to be selected to ensure the feasibility of the research. Abrams et al. (2003, p.41) provide some guiding requirements for selecting the right indicators, including perceived importance by the researcher, influence of the indicators on the object of analysis and sense of urgency. In order to gain a better understanding of the criteria, the refined logic model framework has been adapted to include relationships between the features, either already identified in the literature or intuitively defined. The result has been visualized in figure 5.1. These relationships made it possible to get a general idea of different lines of inquiry, clusters and key features. This helped prioritize which (aspects of) the features needed to be investigated in more detail. An example is collaboration, which appears to be playing a key role in the activities category and will therefore be researched more extensively. Additionally, the relationships helped define which indicators would be more or less appropriate in which category. Actions inspired by the experiment, for example, are described as an indicator in the outputs category as well as the outcomes category by Luederitz et al. (2017). However, it would make little sense to assess the same indicator twice. This inspired the decision to only assess the stakeholders' changed perspectives in the outputs category and follow up with the actions inspired by these new perspectives in the outcomes category.
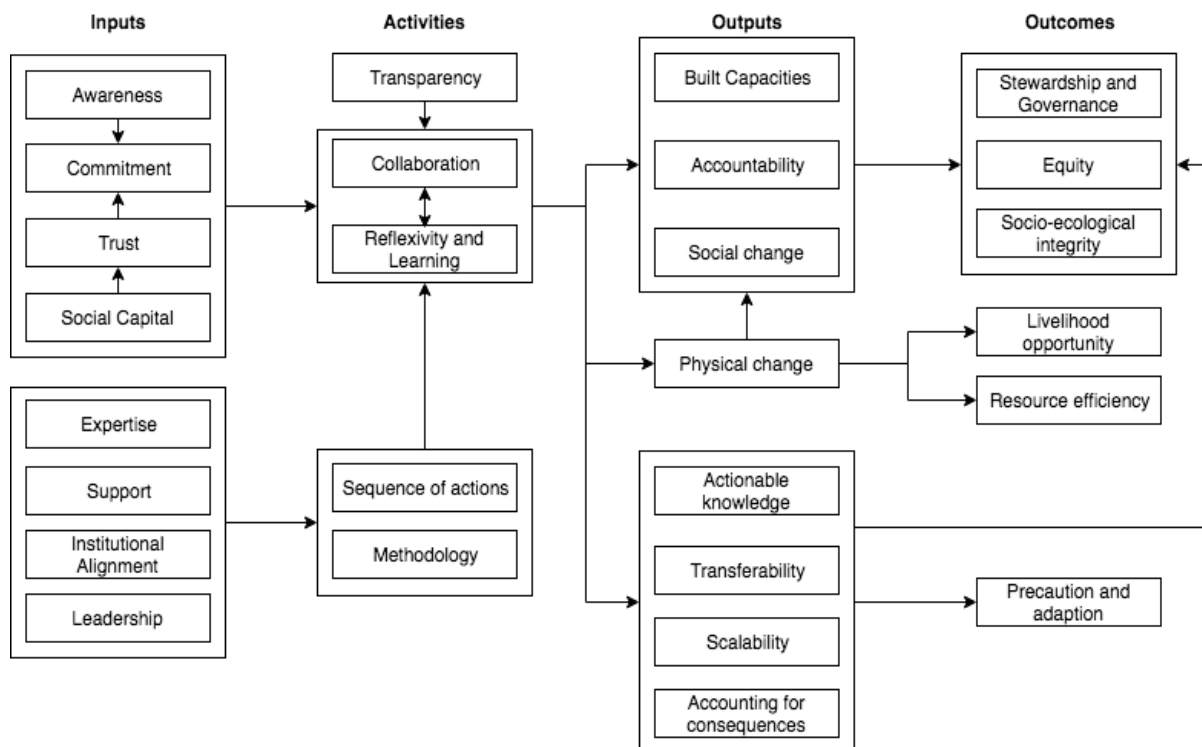

Figure 5.1: Causal relationship diagram

Some preliminary observations about the scheme's features can be made based on the relations displayed in figure 5.1. For the input and activities features, the distinction could be made between two groups of features. The first relate to the experiment participants and those affected by the experiment. Awareness of the issue, trust and social capital are all thought to contribute to the commitment of stakeholders to the experiment. Each of these features is considered to enable to process of collaboration and learning. Transparency is a good governance practice that aids this process. The second group of features relate to the boundary conditions of the experiment. Expertise, support, institutional alignment and leadership all have the capacity to enable or hinder the experiment, but are not necessarily related to the experiment participants. Similarly, the sequence of actions and methodology are not activities of the experiment, but boundary conditions that guide the activities. It is therefore that only the actual activities are thought to contribute to the different output features. Built capacities, accountability and social change are all changes that occur within the group of experiment participants or wider stakeholder group. These changes are thought to contribute to outcomes that continue the transition, such as socio-ecological integrity and democratic governance. A second output is physical change, which can either create a barrier for or enable livelihood opportunity and resource efficiency. The last output group focuses not so much on local change, but rather on transition potential. Actionable knowledge, transferability, scalability and accounting for consequences are all experiment characteristics that can contribute to transforming the system. These do not only lead to precaution and system adaption, but can also support the outcomes dealing with speeding up the transition.

After this general direction for defining indicators has been identified, appropriate indicators have been defined. Most of the indicators used to assess the features were inspired by or adapted from the indicators for evaluating governance by Abrams et al. (2003) and the examples described by Luederitz et al. (2017). Their design was also tested against the characteristics for good indicators as defined by Abrams et al. (2003, p. 40). The *significance* of the indicators was protected by ensuring the indicators provided a measurement of the criteria that have been identified as most relevant. The indicators have been designed in such a way that they are *measurable* as well as *sensitive* to changes in proportion to the changes of the item being measured. This means that, where deemed possible, measurements on a ratio scale are used. However, one can argue that, for many criteria, the number of observations will only provide a limited insight into the nature of the criterion (Larrue, Hegger, & Trémorin, 2013). One such example is the number of pre-existing networks included in the experiment: only counting the number of networks will provide limited insights in their added value. Due to the high variety of possible networks included in the experiment, however, comparison of the individual networks is also likely to be challenging. In cases like the aforementioned example, categorical scales that can categorize and combine different elements are added to the ratio scale. This way, some more insight can be gained into the added value of the elements observed while still allowing for comparison within and between experiments. In some cases, such detailed combined quantitative and qualitative indicators were deemed *impractical* or too *complicated* to measure. In those cases, qualitative indicators have been used instead. In order to accommodate these measurement challenges while remaining some degree of detail, ordinal scales will be designed with *precisely* defined scales to ensure validity. As the scope of the research does not allow for highly detailed measurements, 3-point scales will be designed. These scales should also allow for some more detailed *comparison* of the impact of different features within and between cases than a binary scale, but are generic enough to accommodate the broad research focus.

5.2. Operationalization of input criteria

The operationalization of each of the criteria for the input features will be briefly explained below. In order to increase the readability of the text, the main body of text will just contain the arguments for the choices made. Table 5.1 displays the indicator scales that have been designed for each of the features.

Awareness

The indicator for awareness is the stakeholder's awareness of the need for flood risk management, which is thought to help citizens commit to the experiment (Luederitz et al., 2017, p. 70). Additionally, one can argue that awareness of the stakeholders' own in flood risk management can be an enabling factor as well, as it will reduce the barrier for stakeholders to join the experiment. As the civil servants involved in such experiments are generally responsible for climate change mitigation within their municipal department, only the level of awareness of the non-professional stakeholders will be measured. As awareness cannot be measured within the scope of this research on a ratio scale, an ordinal scale has been designed, with the lowest level of awareness being the majority of the stakeholders being not aware of flood risk management and the highest level being the majority of the stakeholders being aware of the need for flood risk management and their own role therein. It can be theorized that the latter will lead to a more successful experiment, due to the feature's positive relation with the activities.

Commitment

The indicator for commitment is the frequency of participation of each of the experiment participants. Highly committed stakeholders are thought to be an enabling factor for collaboration and learning. Ideally, this would be measured on a ratio scale (Luederitz et al., 2017, p. 70). However, due to the informal nature of such experiments, such specific data is unlikely to be available. An ordinal scale is designed instead. Without any stakeholders participating, the experiment could not have taken place. The lowest point is therefore defined as: a majority of stakeholders participated only once or twice. The highest point, on the other hand, is defined as: a majority of stakeholders participated in all events.

Trust

The indicator for trust is the stakeholder's willingness to rely on other participants' judgement and capacities (Luederitz et al., 2017, p. 70). As trust is grouped in the input category, only the level of trust between the participants before the start of the experiment is included in this operationalization. As trust cannot be measured on a ratio scale, an ordinal scale has been designed. The lowest level of trust has been defined as: stakeholders are not willing to rely on other participants' judgement and capacities, scaling up to a willingness to rely on others' judgement and capacities under certain conditions and to a high willingness to rely on each other.

Social Capital

Social capital can consist of shared values and shared knowledge, often combined in different networks. The presence of neighborhood networks in the experiment is therefore chosen as the indicator for social capital. In order to compare the different networks that were included in the experiment as well as the type of knowledge and values they represent, the presence of social capital is measured by counting the number of networks included in the experiment and the category they represent. The four main categories typically active within a neighborhood are: advocacy groups, such as a group lobbying for a new park, nonprofit organizations, such as a

youth association, for-profit organizations, such as a housing association, and governmental organizations, such as for example public consultation groups.

Expertise

The indicator for expertise is the number and type of experts included in the experiment (Luederitz et al., 2017, p. 70). As experts included in the experiment could represent a wide range of different input, such as ethics, process management and scientific knowledge, including each of these disciplines to the list of categories would wield an unmanageable list. The categories list has therefore been reduced to three categories that represent the role their expertise plays in the experiment. These categories are: experts providing passive input, such as information, experts providing active input, such as process design and experts providing practical input, such as guidance of a brainstorm session.

Support

The four main categories of support, according to Abrams et al. (2013), are financial support, such as funding, information, such as a risk assessment, tools, such as a communication platform, and assistance, for example through coaching or training opportunities. The occurrence of these type of support will be measured on a ratio scale where possible, for example the amount of funding or the number or hours of training sessions.

Institutional Alignment

The indicator for institutional alignment is the presence of institutional barriers hindering the design of implementation of the experiment. In particular, local laws and regulations are considered to be possible barriers for experimentation. Due to the often adaptive and open nature of transition experiments, it is possible that no documentation exists of the number of times an institutional barrier was encountered during the design and implementation process. For example, when considering possible designs, a number of them might have been excluded due to conflicting laws and regulations. It is unlikely there will be documentation of each of these instances. Instead of a ratio scale, an ordinal scale has therefore been designed. The lowest scale representing elements of an experiment that could not be executed or implemented due to institutional barriers and the highest stale representing no issues with institutional barriers at all.

Leadership

In order to measure leadership, it is important to not just look at the presence of a leadership entity, but also at the satisfaction of stakeholders with the effectiveness of this leadership entity (Abrams et al., 2003). An ordinal scale has therefore been designed measuring both aspects. The least effective form of leadership is considered to be: no leadership entity could be identified. If a leadership entity could be defined, the degree of stakeholder satisfaction with this entities effectiveness determines if leadership is classified as medium or high.

Feature	Operationalization
Awareness	**Ordinal Scale:** Low: Stakeholders are not aware of flood risks, nor of their own role in flood risk management Medium: Stakeholders are aware of flood risks, but not of their own role in flood risk management High: Stakeholders are aware of flood risks and of their own role in flood risk management
Commitment	**Ordinal Scale:** Low: A majority stakeholders participated only once or twice Medium: A number of stakeholders participated only once or twice, and a number of stakeholders participated in most events

	High: A majority of stakeholders participated in all events
Trust	**Ordinal Scale:** Low: Stakeholders are not willing to rely on other participants' judgement and capacities Medium: Stakeholders are willing to rely on other participants' judgement and capacities, but under specific conditions High: Stakeholders are willing to rely on other participants' judgment and capacities
Social Capital	**Ratio Scale, categories:** Advocacy group (related or non-related) Nonprofit organization For-profit organization Governmental organization
Expertise	**Ratio Scale, categories:** Passive Active Practical
Support	**Ratio Scale, categories:** Financial Information Tools Assistance
Ins. Alignment	**Ordinal Scale:** Low: stakeholders experienced institutional barriers and could not execute parts of the experiment as a result Medium: stakeholders experienced institutional barriers, but could execute the experiment High: stakeholders did not experience any institutional barriers
Leadership	**Ordinal Scale:** Low: no leadership entity could be identified Medium: a leadership entity could be identified, but stakeholders do not consider him to be fully effective High: a leadership entity could be identified and stakeholders consider him to be fully effective.

Table 5.1: Operationalization of input features

5.3. Operationalization of the activities criteria

Transparency

Transparency relates to the stakeholder satisfaction with the access to information. This satisfaction relies on 5 pillars: timeliness, relevancy, comprehensiveness, accuracy and reliability (Abrams et al., 2003). The number of pillars that the stakeholders do not have any complaints about determines the degree of stakeholder satisfaction of the access to information. The scales of the indicators for this feature and the other features in the activities category are displayed in table 5.2.

Collaboration

For collaboration, legitimacy, in terms of stakeholder inclusion and satisfaction with the process of collaboration, is the main point of focus (Alexander et al., 2016). A larger number of indicators has been designed to measure and evaluate legitimacy (Abrams et al., 2003). These include the number of decisions made through collaboration, stakeholder representation and stakeholder satisfaction with the process of collaboration. In addition, the type of mechanisms for collaboration are also included in the list of indicators in order to gain a more qualitative understanding of the collaboration process (Luederitz et al., 2017, p. 69). For the mechanisms of collaboration, the same categories will be used as for input: passive, information sharing, active, making designing the process or solutions and practical, implementing the solutions or guiding the process. In terms of the number of big decisions made, the type of decision can relate to the process of collaboration or the output of the collaboration. For stakeholder representation, the inclusion of less privileged stakeholder groups is especially interesting.

Ideally, one would gain insights into all involved stakeholders and their socio-economic background. However, due to the non-committal nature of such large-scale public-private experiments, it seems likely that documentation does not exist. An ordinal scale has therefore been designed, ranging from a homogeneous stakeholder group to a heterogeneous stakeholder group that includes less privileged stakeholder groups. Last, stakeholder satisfaction is also measured using an ordinal scale. Their willingness to repeat the process of collaboration is used as an indicator, with not being willing to repeat the process as the lowest scale and a willingness to exactly repeat the process as the highest scale.

Reflexivity and learning
As the causal diagram indicated that the actual learning will be measured in the output category, the indicators for the reflexivity and learning feature aims to measure the intention to facilitate learning. This intention can be evaluated through the presence of learning goals as well as the design of learning mechanisms (Luederitz et al., 2017, p. 69). An ordinal scale has been designed to measure this intention, ranging from no learning goals and mechanisms have been designed to both learning goals and mechanisms have been designed.

Sequence of actions
The criterion for the sequence of actions states that the experiment should be conducted according to a meaningful sequence of actions. Luederitz et al. (2017, p. 69) describe a meaningful sequence of actions as the careful planning of actions and their interaction with the timeline of the experiment. In order to measure the sequence of actions, one can look at the way each of the subsequent actions relates to the previous actions. If all of the actions relate to one or more previous actions, the sequence can be called meaningful. An ordinal scale has been designed to assess this criterion.

Methodology
The criterion for methodology states that the experiment needs to be conducted according to a sound methodology. A common indicator for this feature is the presence of structured procedures for generating output (Luederitz et al., 2017, p. 69). The adequacy of the methods will be measured in the output category. In order to measure the degree to which the methodology was considered by the experiment designers, a scale similar to the one used to measure the sequence of actions has been designed. If no procedures have been designed beforehand, methodology is classified as low, whereas if all procedures have been designed beforehand, methodology is classified as high.

Feature	Operationalization
Transparency	**Ordinal Scale:** Low: stakeholders are satisfied with 0 or 1 of the characteristics Medium: stakeholders are satisfied with 2 or 3 of the characteristics High: stakeholders are satisfied with 4 of 5 of the characteristics
Collaboration	*Mechanisms for collaboration* **Ratio Scale:** Passive Active Practical *Number of big decisions made through collaboration* **Ratio Scale:** Process Output *Stakeholder representation*

	Ordinal Scale:
	Low: stakeholder group is homogeneous
	Medium: stakeholder group is heterogeneous, but no involvement of less privileged groups
	High: stakeholder group is heterogeneous and less privileged groups are involved
	Stakeholder satisfaction with collaboration
	Ordinal Scale:
	Low: stakeholders would not repeat the process of collaboration
	Medium: stakeholders would repeat the process of collaboration with adjustments
	High: stakeholders would repeat the process of collaboration
Reflexivity	**Ordinal Scale:**
	Low: no learning goals and learning mechanisms have been designed
	Medium: learning goals or learning mechanisms have been designed
	High: both learning goals and learning mechanisms have been designed
Sequence	**Ordinal Scale:**
	Low: there is no relation between subsequent actions
	Medium: there is a clear relation between subsequent actions, with some exceptions
	High: there is a clear relation between subsequent actions
Methodology	**Ordinal Scale:**
	Low: methods for generating output were not structured
	Medium: methods for generating output were structured, with some exceptions
	High: methods for generating output were structured

Table 5.2: Operationalization of activities features

5.4. Operationalization of the output criteria

Built capacities

Stakeholders can gain or enhance a wide range of skills by participating in the experiment, which should empower them to continue working on the transition towards a more resilient system. Due to the large degree of variation in different skills possible, a ratio scale is deemed unfeasible to compare the different experiments. An ordinal scale has been designed instead. As transition experiments to increase flood risk resilience are traditionally public-private experiments, both private, most likely citizens, and public, most likely civil servants, parties can build capacities. The ordinal scale is therefore designed in such a way that both parties gaining or enhancing skills is the highest outcome measurable. The scales for all of the features in the output category are displayed in table 5.3.

Accountability

Accountability is interpreted in this research as a sense of problem ownership that encourages experiment participants to continue to increase flood risk resilience. Much like built capacities, it is a driver for transformational change. Luederitz et al. (2017, p. 66) describe this sense of problem ownership as dual: it encompasses both a feeling of responsibility as well as a commitment to implementing change. An ordinal scale has therefore been designed to measure accountability, ranging from no sense of responsibility to a sense of responsibility and a willingness to take on a more proactive role in flood risk management.

Social change

Luederitz et al. (2017, p. 66) describe indicators for social change as "new or altered activities, practices, routines, as well as social relationships and partnerships". As it would be rather time-consuming to address all features, the choice has been made to look at new partnerships which can encompass all other indicators. The number of new partnerships is measured on a ratio scale and the activities performed by the network are categorized as a continuation of the experiment, an adaptation of the experiment or non-related to the experiment. Furthermore, the social learning taking place in the activities category should lead to another form of social change, that is: changed values and perspectives. A second indicator has therefore been designed to

measure these changed perspectives, focusing, much like built capacities, on both the perspectives of the private and public actors.

Physical change
The indicator for physical change is rather straightforward: it measures the physical change on a ratio scale. That is: how many measures, or square meters of measures were implemented. In order to allow for a comparison between the different experiments, three main categories for nature-based solutions are included in this scale: solutions addressing the issue with impermeable surfaces, solutions addressing issues with water absorption and solutions addressing issues with water storage.

Actionable knowledge
Three types of knowledge can be generated by transition experiments: knowledge about the problem, the normative sustainability goals and the methods to achieve these goals (Luederitz et al., 2017, p. 64). Ideally, all three types of knowledge would be generated. An ordinal scale has therefore been designed that measures for which of the three types knowledge was generated, with knowledge generated for all three types being the highest point of the scale. One can imagine, however, that it is not so much the question if this knowledge was generated, but rather what knowledge was generated that is interesting for this feature and the three features to follow. Aside from rating the features, additional attention will be paid to the qualitative insights that can be gained from these features.

Transferability
Transferability does not encompass the actual transferring of experiment lessons, but rather their potential for transferring. It is therefore relevant to not only look at if any lessons for transferring were learned, but also if they were validated, for example through theoretical research or practical application (Luederitz et al., 2017, p. 66). An ordinal scale assessing if any lessons were learned and if they were validated has therefore been designed.

Scalability
Similar to transferability does the feature of scalability not deal with the question if the experiment itself has been scaled up, but with the lessons or knowledge generated by the experiment that have the potential to facilitate upscaling (Luederitz et al., 2017, p. 67). An ordinal scale similar to the scale for transferability has therefore been designed, again assessing if any lessons were learned and if they were validated.

Accounting for consequences
The last of the "learning features", as identified in the causal diagram, is accounting for unintended consequences of implementing, transferring and upscaling the experiment. These are the different kinds of externalities identified in paragraph 4.4. Specifically rebound effects, offsetting sustainability gains and long-term impacts need to be considered. For accounting for consequences, a ratio scale has therefore been designed with the three aforementioned possible consequences as the possible categories.

Feature	Operationalization
Built capacities	**Ordinal Scale:** Low: neither private nor public actors report to have gained or enhanced skills Medium: private or public actors report to have gained or enhanced skills High: private and public actors report to have gained or enhanced skills
Accountability	**Ordinal Scale:** Low: stakeholders do not feel responsible for flood risk management

	Medium: stakeholders feel responsible for flood risk management, but want to remain reactive
	High: stakeholders feel responsible for flood risk management and want to become proactive
Social change	*Newly formed networks* **Ratio Scale:** Continuation of the experiment Adaptation of the experiment Not related to the experiment
	Changed perceptions **Ordinal Scale:** Low: neither citizens nor civil servants report to have changed their perspectives Medium: citizens or civil servants report to have changed their perspectives High: citizens and civil servants report to have changed their perspectives
Physical change	**Ratio Scale:** Impermeable surfaces Water absorption Water storage
Actionable knowledge	**Ordinal Scale:** Low: learning about 0 of the topics occurred Medium: learning about 1-2 of the topics occurred High: learning about 3 of the topics occurred
Transferability	**Ordinal Scale:** Low: no lessons for transferring have been formulated Medium: lessons for transferring have been formulated, but not validated High: lessons for transferring have been formulated and validated
Scalability	**Ordinal Scale:** Low: no lessons for upscaling have been formulated Medium: lessons for upscaling have been formulated, but not validated High: lessons for upscaling have been formulated and validated
Consequences	**Ratio Scale:** Rebound effects Long-term effects Offsetting sustainability gains

Table 5.3: Operationalization of output features

5.5. Operationalization of the outcome criteria

Socio-ecological integrity
The feature of socio-ecological integrity encompasses the understanding of the stakeholders of the importance of harmonizing ecosystems and human systems with the more traditional physical systems. For nature-based experiments, further implementation of nature-based solutions is a clear manifestation of this increased understanding. This understanding could lead to both top-down implementation as well as bottom-up implementation. Further implementation planned by the civil servants and the citizens are therefore the focus of the ordinal scale that has been designed for this feature. The indicator scales for socio-ecological integrity and all other outcome features are displayed in table 5.4.

Intergenerational equity
Intergenerational equity focuses on the question if the experiment's output led to opportunities for all stakeholder groups. As the number of less privileged stakeholder groups included in the experiment are already included in the activities feature, this feature measures the plans for their continued involvement in the future. The most unequal measure is no current or planned involvement of less privileged stakeholder groups. Next, there is a current or planned involvement, but this is the result of chance or general policies, rather than specific inclusion policies. Although this is seemingly a good example of intergenerational equity, one should be aware that a lack of inclusion policies means that there is no need to respond if these stakeholder

groups stop becoming involved in the future. This form of stakeholder inclusion has therefore been defined as medium. The highest point on the scale is planned or current stakeholder inclusion with specific inclusion policies in place. This is the form most likely to lead to long-term and sustainable inclusion of less privileged stakeholder groups.

Socio-ecological stewardship and democratic governance
As socio-ecological stewardship is already measured in the socio-ecological integrity feature, this feature focuses on democratic governance. Specifically, the continued inclusion of private actors in public decision making. A ratio scale measuring the number of new collaborative settings that were established as a result of the experiment has been designed as the indicator. As with social change, the three categories used to categorize the different participatory settings are: a continuation of the experiment, an adaptation of the experiment and non-related to the experiment.

Livelihood opportunity
Livelihood opportunity encompasses the economic opportunities that arise from the implementation of nature-based solutions, such as increased foot traffic or less damages to infrastructures. The economic benefits of implementing nature-based solutions can be measured on a ratio scale by looking at the money indirectly generated by the solution. Three categories of indirect benefits could be identified: benefits resulting from less damages to or need for infrastructure, benefits resulting from increased human wellbeing and benefits resulting from commercial activities and job opportunities.

Resource efficiency
For resource efficiency, especially the transport and implementation of nature-based solutions can lead to carbon dioxide emissions. As the calculations of transport emissions are outside the scope of this research, an ordinal scale has been designed instead. The most polluting option is if materials had to be transported from abroad, with the least polluting option being the use of local materials.

Precaution and adaption
The last feature, precaution and adaption, aims to minimize future flood risks. As concluded in paragraph 3.2, resilient solutions are multi-scale, modular, redundant and multifunctional. Climate change or flood resilient measures and approaches should therefore consist of varied solutions, addressing different parts of the system and overlapping on occasion (Ahern, 2011). If the experiment inspired the design of innovative measures or a new action plan containing such solutions, the experiment succeeded in realizing precaution and adaption. An ordinal scale has been designed to measure this outcome, ranging from no action plan or future measures formulated to a highly diverse, and thus resilient, action plan or set of measures formulated.

Feature	Operationalization
Integrity	**Ordinal Scale:** Low: no further implementation of NBS planned or undertaken Medium: further implementation of NBS planned or undertaken by citizens or civil servants High: further implementation of NBS planned or undertaken by citizens and civil servants
Equity	**Ordinal Scale:** Low: no planned or current involvement of less privileged stakeholder groups Medium: planned or current involvement or of less privileged stakeholder groups, but no specific policies designed High: planned or current involvement of less privileged stakeholder groups with specific policies designed
Governance	**Ratio Scale:**

	Continuation of the experiment Adaptation of the experiment Not related to the experiment
Opportunity	**Ratio Scale:** Infrastructure Human health and wellbeing Commercial and job opportunities
Efficiency	**Ordinal Scale:** Low: international materials used Medium: national materials used High: local materials used
Precaution	**Ordinal Scale:** Low: no flood risk mitigation plan or measures formulated Medium: narrow flood risk mitigation plan or measures formulated High: diverse flood risk mitigation plan or measures formulated

Table 5.4: Operationalization of outcome features

5.6. Assessment of the operationalized evaluative scheme

When applying the evaluative scheme, one should be aware that the designed indicators on their own provide little insight into the complex and interdependent nature of the features they represent (Conway, 2007). The number of participants, for example, can be used as an indicator for the collaborative process (Hartmann & Spit, 2016). This indicator, however, fails to factor in trade-offs between the number of participants and factors such as resource efficiency and manageability of the experiment. As these trade-offs differ per experiment, only counting and comparing the number of participants is therefore likely to generate an outcome that is of little meaning or use for evaluating the experiments (Larrue et al., 2013). This is a clear signal that focusing on *quantity* will not directly generate any insights in the *quality* of the feature. The operationalization of the criteria is therefore not a checklist of elements that need to be present at a specific level, but rather a guide to help focus the case studies and make it easier to identify, assess and compare elements of the experimentation process and their relations (Alexander et al., 2016).

Based on the lessons learned from operationalizing the framework, the research philosophy has shifted towards a more pragmatic approach. Highly structured indicators will be used to assess the transition experiments, but at the same time a more open, in-depth analysis of the framework's features will take place to identify any unexpected findings, nuances and trade-offs.

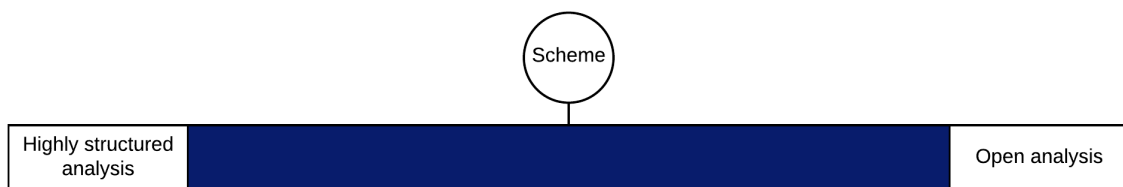*Figure 5.2: Research philosophy adapted based on literature review*

In short, the literature review proved some strengths of the evaluative scheme by Luederitz et al. (2017), but also exposed some possible weaknesses. In order to generate some more insights in these preliminary statements, in the next chapters, the evaluative scheme will be applied to real-world cases to also empirically assess its usefulness.

6. Research Design

In the previous chapters, a conceptual framework for evaluating nature-based FRM experiments has been constructed. Before this framework can be applied to real-world experiments, the research strategy will have to be developed. The goal of this design is to ensure that the data collected is suitable for answering the research question, as well as emphasize the need for a high quality, that is: valid and reliable, case study research. In this chapter, the research strategy will first be briefly introduced, followed by an identification of tactics for safeguarding the research quality. Paragraphs 6.3, 6.4 and 6.5 will detail the processes of case selection and data collection.

6.1. Research Strategy

Yin (2014, p. 29) identifies five components of the research design that can be used as the main guidelines. Each of the components will be briefly discussed. *Research questions* are considered to be the most important starting point for any research design. The research question in this report as described in chapter 1 aims to discover which method can be used to evaluate the governance nature-based flood risk management experiments in relation to the experiments' success. This research question is based on the general *proposition* that governance of transition experiments is an enabling factor in their success. More detailed propositions on the indicators for and relationships between experiments' inputs, activities, outputs and outcomes are described in chapter 5. As it is impossible to cover every little detail about the selected cases, these propositions are considered to be the most important and will provide the focal point for the research. Following the propositions, the *unit of analysis* is identified as the governance of a sustainability transition experiment as defined in chapter 2. The method to *link data to propositions* will be a combination of two different methods as identified by Yin (2014, p. 142). First, a logic model similar to the ones described in chapters 4 will be used to structure the chain of inputs, activities, outputs and outcomes as observed within the case. By comparing the empirical logic model to the theoretical logic model, patterns between theory and practice can be matched. By performing a cross-case comparative analysis next these patterns can either be confirmed or disconfirmed, which will provide guidance for the generalization of the research findings. The *criteria for interpreting the findings* allow the researcher to assess the robustness of the findings. These criteria are not as straightforward as they would be for a quantitative study, as commonly accepted thresholds or scales cannot be used for nominal analyses. Instead, the findings' so-called substantive significance will be used to assess the outcomes of the case studies (Patton, 2002). Bloomberg and Volpe (2008, p. 130) identify four questions that should be addressed when interpreting the meaning and relevance of the findings:

1. Are the findings solid and consistent? This includes the rejection of rival explanations.
2. Are the findings consistent with the existing body of knowledge? And what are the implications thereof?
3. How and to what extent do the findings increase the understanding of the object of analysis?
4. Are the findings useful for theory-building, policy or practice?

The first two questions are particularly interesting to assess the outcomes of the case studies: how reliable are they? What conclusions can be drawn from them? The last two questions, on the other hand, can be asked to assess the criteria used in the case studies: were they useful in evaluating the cases? Did they lead to findings that can inform theory or policy? Answers to these questions should lead to the answer to the main research question.

As the framework operationalized in chapter 5 identifies a large number of different elements of the experiment, the subunits, an embedded case design is most appropriate for this research. Additionally, a multiple-case design will allow for a cross-case comparison that will increase the generalizability of the findings (Yin, 2014). The research design for an embedded, multi-case study has been further developed in the research protocol as included in Appendix B. This protocol includes an overview of the case study, data collection procedures, data collection questions and guidelines for reporting on the results.

6.2. Tactics for Safeguarding Research Quality

Yin (2014, p. 45) identifies a number of tactics that safeguard the quality of the research. Each of the quality measures and the strategy used to ensure this quality has been summarized in table 6.1.

Measure	Description	Addressed in research
Construct Validity	Correct measurement of the concepts	Multiple sources of evidence Establish chain of evidence
Internal Validity	Correct identification of causal relationships	Pattern matching Logic models
External Validity	Correct generalization of findings	Theory Replication logic
Reliability	Research can be repeated with the same results	Case study protocol Case study database

Table 6.1: Tactics for safeguarding research quality based on Yin (2014, p. 45)

In order to increase construct validity, two tactics have been used. The first, using multiple sources of evidence by conducting multiple interviews per case as well as using case documents, will reduce the researcher bias in appraising, often highly subjective, concepts. The second tactic serves the same goal by building an evidence-based line of reasoning that can easily be traced and understood by an external observer. This process will allow an external observer to identify any subjectivity and offers the researcher an opportunity to critically assess her process and assumptions. By linking observed causal mechanisms to existing theory as well as 'opening up the black box' and identifying possible rival explanations, the risk of identifying false causalities is reduced. To increase the external validity, that is: the generalizability, of the research, findings are to be grounded in existing theory as well as be supported with similar findings in other cases. The more proof existing literature offers, or the more often a result is observed, the stronger the generalizability. The research reliability is increased by using the research protocol as described in appendix B as well as the development of a case study database that will allow external observers to access unpublished data relating to the case.

In addition to Yin's (2014) tactics for safeguarding the validity, both commonly accepted and unique rival explanations that threaten the research's internal validity can be identified and need to be addressed in the research design. Common craft rivals and the way they are or will be addressed in the research are summarized in table 6.2.

Craft rival	Description	Addressed in research
History	Context events influence outcomes	Select cases that did not experience any influential context events
Maturation	Changes in behavior not related to the experiment	Select cases that are at the same point in the transition process
Testing	Adjusted behavior due to awareness of testing	Select cases that were not subject to scientific observation or testing
Instrumentation	Changes in way phenomena are measured or evaluated	Include only one researcher in the research Define detailed evaluation criteria

| Statistical regression | Extreme values trend toward the more common values over time | Not relevant in this research |
| Mortality | Loss of participants | Select cases with the same duration and scale |

Table 6.2: Craft rivals (Christ, 2007)

As the craft rivals mainly relate to scientific experiments administering some sort of intervention, some liberty has been taken in translating these craft rivals to fit a qualitative research focus on a process of experimentation rather than a controlled scientific experiment. Most of these craft rivals can be largely excluded by carefully selecting the experiments. In order to reduce the impact of history, for example, one should select cases that experienced no external influences that could have influenced the outcomes, such as flooding, a financial crisis or political upheaval. To be able to perform a cross-case comparison, one should also look out for the maturation of the cases, as those designing and participating in a first experiment are probably less experienced than those participating in their third or fourth experiment. Similarly, participants might adjust their behavior if they know they are being observed or evaluated on specific points. This is why, ideally, cases that were not subject to any form of testing are selected. Instrumentation is the only craft rival that is related to the researcher and not the case. Inconsistencies in measurement should be mitigated by following predefined evaluation criteria. Last, mortality, or participant drop-out, appears to be mainly an issue for controlled scientific experiments. However, in the evaluation scheme, the number of times stakeholders participated in the experiment is included as an indicator. One can imagine that experiments with just two stakeholder meetings within two months probably have a much lower drop-out rate than experiments with ten stakeholder meetings spread out over three years. One should therefore try to correct for this by selecting cases with a comparable number of meetings over a comparable timespan.

In addition to craft rivals, one should also be considerate of unique rivals (Krathwohl, 2009). Unique, or confounding, rivals can be described as alternative hypotheses that could explain that what has been observed in the cases better than the main hypothesis (Yin, 2012b). However, as described in the introduction, due to the exploratory and qualitative nature of the research, no hypothesis to be proven or disproven has been formulated. Nevertheless, for interpreting the outcomes of the case studies, it can be highly relevant to compare the data to different theories that can explain the case observations. Due to the relatively new field of research of transition experiments, however, identifying these competing theories is a rather challenging task. Some theories on alternative explanations for the observations could be found, however. As a reminder, the focus of this research is based on the proposition that the governance of experiments is the key driver for the experiment's success (Weiland et al., 2017). Research by Antikainen et al. (2017), however, identifies the different input features as the key success factors. It can also be argued that context features, such as the political and institutional environment play a key role in the success of the experiments (Berkhout et al., 2010). As the evaluative scheme designed in chapter 5 already includes many of these possibly cofounding factors, it should be possible to assess their impact in comparison to the governance mechanisms that are to be studied.

6.3. Case Selection

The objective of the research is to assess if the criteria identified in chapter 5 can be used to evaluate transition experiments and can generate insights into the governance of such experiments. Of particular interest is the assumption that these criteria are broadly applicable to different transition experiments. This is best tested by selecting cases that took place under different circumstances and thereby likely encountered different barriers and enabling factors (Swanborn, 2008). This case selection method has been labeled the 'diverse-case method' and

aims to achieve variance in order to increase the representativeness of the cases selected (Gerring & Seawright, 2007).

Within this broad research goal, however, there is also a narrower research goal, aiming to distill governance elements that contribute to the success of transition experiments, which requires a high level of internal validity. This means that, while taking into consideration the diverse-case criterion, differences that could threaten the internal validity of the case study outcomes should be minimized. In order to ensure a proper assessment of the evaluation framework criteria, it is important that all of the experiments included in the case studies are, indeed, experiments with a focus on managing flood risks through nature-based solutions. Additionally, these experiments have to be focused on facilitating a transition towards a more resilient water management system. As the definition of such transition experiments formulated in chapter 2.2. still leaves room for quite some variation, this focus is narrowed down even further making use of the typology of experiments in sustainability science by Caniglia et al. (2017). The specific type of experiment that will be included in this research deals with sustainability solutions with participatory control over the interventions. Inherent to this type of experiment is that the experiment takes place in a real-world setting, and includes researchers, policy makers and citizens (Caniglia et al., 2017). Furthermore, some more pragmatic arguments, such as the accessibility to information, play a role in the selection of the cases.

Given the requirements described in the previous subparagraphs, the choice has been made to focus on experiments that took place within the Resilient Europe network. This network is part of the URBACT program, a program co-financed by the European Regional Development Fund and the member states of the European Union, Norway and Switzerland with the goal to facilitate cooperation and knowledge sharing between different European cities in order to solve urban challenges (URBACT, 2014). Not only do the experiments within the Resilient Europe network meet the criteria of being transition experiments as defined by Caniglia et al. (2017), direct access to the stakeholders within the network makes it easier to collect the required information. Additionally, the experiments taking place within the network ended in Spring 2018, coinciding with the start of this research, which makes the timing highly suitable for process evaluation (United States Office of Energy Efficiency and Renewable Energy, 2018).

Cases that represented different contexts while minimizing the occurrences of rival explanations experiment were selected. In chapter 5, different input features that influence the activities were identified. Many of them, such as trust, commitment and awareness, depend heavily on the political environment. The quality of democracy in a country determines the degree of access citizens have to government information, the freedom to express oneself and the trust in a fair process of decision-making (Sustainable Governance Indicators, 2018b). The quality of governance, on the other hand, encompasses the degree of societal consultation, accountability and adaptability in policy-making as well as the capacity of citizens to organize themselves and participate in the policy making process (Sustainable Governance Indicators, 2018a). The lower a country's score on one of both of these indicators, the more challenging conducting a participatory transition experiment is likely to be. Additionally, resources are considered to be relevant input features. The Gross National Income per capita reflects a county's citizens' average income and is correlated with other measures of a country's welfare, such as education and quality of life (The World Bank, 2018a). A low GNI could therefore indicate limited access to resources such as funding for the experiment and adequate time and knowledge for citizens to participate. The scores of the countries in which the experiments took place on each of these indicators were therefore used as guidelines for selecting diverse cases. In order to achieve maximum variance, a case with high scores as well as a case with low scores were selected (Gerring & Seawright, 2007). The selected cases also had to be similar on a

number of characteristics to minimize the impact of craft rivals as described in table 6.2. The selected cases and their scores on the indicators are summarized in table 6.3.

	Burgas, Bulgaria	Antwerp, Belgium
Democracy score[2]	5.72	7.35
Governance score[3]	5.15	6.37
GNI per capita[4]	7,580 USD	41,860 USD
Inhabitants	6.500[5]	6.000[6]
History	No major events	No major events
Maturation	First experiment	First experiment
Testing	None	None
Mortality	Duration: 2015-2018 Scope: visioning and one demonstration project	Duration: 2015-2018 Scope: visioning and one demonstration project

Table 6.3: Selected Cases

6.4. Case Introduction

The city of *Burgas* is located in the southeast of Bulgaria, flanked by three large lakes and the Black Sea. It is the fourth largest city in Bulgaria and experiences a continued population increase. In the last years, Burgas has experienced a number of floods caused by heavy rainfall, causing major economic damages and even human casualties (Burgas Municipality, 2018). This has been a clear demonstration that existing, predominantly hard-engineered infrastructure is currently no longer able to defend the city from flooding and changes to the water management system are needed. However, the GNI per capita of Bulgaria is with 7580 US Dollars one of the lowest in the European Union and allows for the categorization of Bulgaria as a developing country (The World Bank, 2018b). Similarly, Bulgaria is one of the weakest countries in the European Union in terms of democracy and governance scores (Sustainable Governance Indicators, 2018a, 2018b). As a result, public participation in policy making processes and trust in the government are low and limited resources for experimentation are available. All of these barriers needed to be overcome before the municipality and citizens of Burgas could even start to design a transition experiment towards resilient flood risk management.

The city of *Antwerp* is the capital of the Flanders province and is located in the north of Belgium, at the outlet of the Scheldt river. Its city center has a high building density that limits rainwater absorption. Some issues with rainwater runoff caused by heavy rainfall already occur on a small scale and are expected to increase in scale and occurrence in the future (Antwerp Municipality, 2018) . With a GNI per capita of over 5 times that of Bulgaria, Belgium can be classified as a developed country (The World Bank, 2018a). Societal consultation is common and citizens are relatively well organized and equipped to participate in policy processes (Sustainable Governance Indicators, 2018b, 2018a). Subsequently, it is assumed that

---

[2] (Sustainable Governance Indicators, 2018b)

[3] (Sustainable Governance Indicators, 2018a)

[4] (The World Bank, 2018a)

[5] (Wikipedia, 2018)

[6] (Antwerp Municipality, 2018)

stakeholders in the city of Antwerp should experience less barriers for experimentation than those in Burgas. The major challenge for Antwerp, however, was to align resilient infrastructure with resilient citizens and institutions in order to set the stage for a city-wide transition towards resilient water management.

6.5. Data Collection

After the cases have been selected, a number of steps need to be taken prior to the data collection. These sequential steps are: defining which information to subtract from the cases, selecting participants, designing research questions and defining the data collection procedures. In this research, the *information* that is required from each case is already extensively defined and operationalized in chapter 5. Each of the four categories of the logic model framework is considered to be useful for answering the main research question, although the activities and output categories are the focus of this research. For the *selection of participants,* it is important to select participants that have a good understanding of these two categories. Ideally, all different types of stakeholders involved in the experiment, such as citizens, policymakers and researchers, would be interviewed as their backgrounds could give them different perspectives on the object of research. For Antwerp, it was possible to interview citizens to corroborate the findings of the policymakers and provide a different perspective. For Burgas, direct interviews with citizens were deemed not feasible due to both language barriers and difficulties with making contact with the citizens involved in the experiment. In that case, email interviews, making use of open ended questionnaires, could be conducted to gain insights into the citizens' perspectives. In order to conduct *ethical research* each of the participants received an email detailing the interview process and goals before the interview. This information was repeated again at the beginning of the interview to allow the interviewee to address any questions or objections he or she might have. All interviews were recorded with the consent of the interviewee. These recordings were destroyed after the interviews had been transcribed. All the interviewees received a report of the interview and were given the opportunity to comment on the researcher's interpretation of the interview if desired. All interviewees received a copy of the final report.

An overview of the conducted stakeholder interviews is displayed in table 6.4. In order to protect the interviewee's privacy, names have been redacted. A more extensive explanation of the data collection phase is included in the research protocol in appendix B.

No.	Experiment	Role	Date	Medium
1.	Antwerp	Policymaker	April 25, 2018	Skype
2.	Antwerp	Process counsellor	April 25, 2018	Skype
3.	Burgas	Policymaker	April 26, 2018	Skype
4.	Burgas	Policymaker	April 26, 2018	Skype
6.	Antwerp	Citizen	April 28, 2018	Face-to-face
7.	Antwerp	Citizen	April 28, 2018	Face-to-face
8.	Antwerp	Citizen	April 28, 2018	Face-to-face
9	Burgas	Citizen	May 15, 2018	Survey
10.	Burgas	Citizen	May 15, 2018	Survey
11.	Burgas	Citizen	May 15, 2018	Survey

Table 6.4: Stakeholder interviews

7. Empirical Research

In this chapter, the findings from the two cases will be analyzed. First, a brief, chronological overview of the case will be given, followed by a summary of the findings for each of the evaluation framework features. Last, some preliminary observations about the case and the framework will be made. The experiment in Burgas will first be discussed, with the experiment in Antwerp being discussed second.

7.1. Individual Case Analysis: The 'Green Belt' Experiment, Burgas

The Dolno Ezerovo neighborhood is located at the outskirts of Burgas, bordering Lake Vaya. Most of the neighborhood is located below sea level. Since 2010, the neighborhood experienced 5 major floods, causing significant damage to properties and infrastructure and even human deaths (Burgas Municipality, 2018). Climate change is expected to increase the risk of flooding and it is therefore important to increase the neighborhood's ability to withstand these threats. The goal of the experiment was therefore related to increasing the neighborhood's resilience. Not by focusing on hard-engineered measures, as is common in Burgas, but by focusing on the relation between infrastructures, people and ecosystems. After receiving help from the Resilient Europe network project leaders, a collaboration process was designed, the main goal of which was to co-create a detailed action plan to increase the neighborhood's resilience. A risk assessment was performed by an external company in order to identify the most important challenges related to climate change and flooding. This assesment served as the main input for the design of the action plan. A budget of €45.500 was made available through EU funding and municipal funding.

Preparation Phase
In order to determine the focus of the experiment, a first meeting was organized with the municipality's consultative council. Stakeholders included representatives of the harbor and airport, local businesses, the Red Cross and the local police and fire departments. During this meeting, the contours of the action plan were defined. A second meeting was organized with the most relevant stakeholders from the consultative council, including different municipal departments and NGO's like the Red Cross and environmental organizations, as well as residents from the neighborhood. During this meeting, the idea of implementing a 'Green Belt' of nature-based solutions was agreed upon. The idea of experimenting with the small-scale implementation of such a nature-based solution in collaboration with the residents was also conceptualized during this meeting.

Experimentation Phase
Initial attempts to engage citizens in the implementation of a small-scale solution had been unsuccessful. Overall, citizens felt let down by the municipality. The past floods proved to them that the municipal policies to increase flood risk resilience were ineffective. Furthermore, previous negative experiences with public consultations led them to believe that their input would not be appreciated by the municipality. The head of the local center for administrative services was therefore asked to act as a mediator between the civil servants and the citizens. With her help, the civil servants succeeded in engaging some local key stakeholders who engaged their networks. Additionally, the civil servants held door-to-door conversations with the citizens to discuss the experiment and action plan with them. This method of direct contact proved to be a useful tool in engaging all stakeholders, including less privileged stakeholder groups, in the experiment. A meeting was organized where the citizens could discuss where to implement the small-scale solution. A festive event, kicked off by the deputy mayor, was organized to plant a symbolic first five willow trees from a local nursery at this chosen location. A very high number of residents attended the event. The present civil servants made use of this

high turn-out to inform residents about the initiative, receive input on the action plan and build public support. Residents expressed their satisfaction with the nature-based solution, which would not only help to protect against floods, but also improved the local living environment.

Conclusion phase
After the small-scale solution had been implemented, four more meetings with the stakeholders from the consultative council and some residents were organized. During these meetings, the stakeholders reflected on the lessons learned from the small-scale implementation and designed different pathways for achieving urban resilience. These pathways consisted of different, successive, actions to take to achieve different goals, such as citizen awareness and the implementation of nature-based solutions. An external consultancy company was asked to design a detailed action plan based on these pathways, which was discussed and refined during the meetings. The last of the four meetings was aimed at identifying the next actions to take and to inspire all stakeholders involved to proactively work on increasing resilience in the neighborhood

Outcomes
Both the citizens and the civil servants reported to be very satisfied with the process of collaboration. The civil servants stated to have learned new methods to involve citizens in the process of policymaking and have started to see the potential of the residents becoming more proactive in climate change mitigation. The citizens, on the other hand, reported to have regained some trust in the municipality by actively being involved in the design of the action plan. They also have an increased awareness of the measures that can be taken to protect themselves and their neighborhood from floods. However, actual changes in behavior were only visible at a municipal level, indicating that the experiment failed to empower the citizens. Plans are being made to repeat the process of experimentation in other areas of the city. An increased awareness of the importance of collaboration between municipal departments inspired further collaboration. The citizens, however, have not yet started to change their behavior, although the indicate a high willingness to collaborate in a similar fashion with the municipality for the implementation, and if needed: adaptation, of the action plan.

Lessons learned
For the municipality, the most important lessons learned were on how to engage citizens, by using a mediator as well as intensive, face-to-face contact, and on how to increase a neighborhood's climate change resilience, by including different aspects that enhance each other's impact. Additionally, they learned that small-scale demonstration projects are a good way to test different policies as well as show citizens that their input leads to tangible results. However, none of these lessons have been applied elsewhere yet. The experiment succeeded in its goal of designing an action plan to increase the resilience of the neighborhood and learning lessons on how to implement the action plan. As for the neighborhood's social resilience, citizens have not been activated yet, but a basis for further collaboration between the municipality and the citizens has been built. The municipality did learn, however, that, once contact has been made, engaging citizens in improving their own safety and living environment is relatively easy, but engaging them in projects where the benefits are less obvious, for example by repeating the experiment in a neighborhood where flooding has not yet occurred, might be more challenging. Consequently, a downside of this form of collaboration between citizens and the municipality is the intensity of the process, which requires a lot of time and effort from the civil servants. Because of high local awareness of the need for change and the small scale of the experiment, these were no issues in this experiment, but they might become issues when the experiment is being transferred or scaled up.

7.2. Enabling factors and barriers for experimentation Burgas

In this paragraph, the case will be structured according to the logic model evaluation framework. Extensive argumentation for choices made and examples are included in Appendix E.

7.2.1. Input features (Appendix E.1)

Awareness

Due to the experience with the floods, there was a high awareness of flood risks in the neighborhood. Initially, this awareness was offset by a low awareness of the role of citizens in protecting themselves against the flood risks. In the framework, awareness has therefore been classified as: medium.

As the experiment progressed and the citizens learned more about their role in the experiment, this flood awareness became an important driver for commitment. Additionally, the awareness of the opportunity to improve their living environment was a second reason for citizens to commit to the experiment.

Commitment

The civil servants from the different departments were the only group that was actively involved throughout the experiment. Then there were two other groups of stakeholders that were highly committed to either the meetings to design the action plan or the small-scale implementation. The majority of the citizens participated only in one or two of the meetings or events. Commitment has therefore been defined as medium. As the group of relevant stakeholders was quite large, however, working with a small core group was considered more efficient than putting efforts in engaging all stakeholders for all actions or events.

Trust

Trust between the citizens and the municipality has been classified as: low. The citizens felt like their input was not appreciated and their needs were not taken into consideration by the municipality. This created a barrier for them to join the experiment. The municipality on the other hand had limited faith in the citizen's capacity to become more involved in flood risk management, which meant that the civil servants had limited experience with engaging citizens. This too posed a barrier.

The collaboration during the small-scale implementation helped build trust between the both parties. The citizens felt taken seriously and the civil servants saw that the citizens were willing and able to contribute to flood risk management. This removed a barrier for future collaboration.

Social capital

In addition to making use of the existing consultative council to engage stakeholders in the experiment, two local networks were used. These two networks proved instrumental in spreading information and encouraging citizens to join the experiment. Even more vital was the head of the local center for administrative services. As she had a good relation with both the civil servants and the citizens, she acted as a mediator between the two groups, connecting key stakeholders and encouraging citizens to join the experiment. In terms of local knowledge, the citizens had little experience with bottom-up initiatives.

Expertise

A few experts were included in the experiment. Their roles ranged from collecting the visions and writing the action plan to organizing and presiding over the meetings. Additionally, the

experiment received help from the Resilient Europe network, which proved to be very useful in providing the tools to design the process as well as the action plan.

Support
The most important driver was the financial support. The funding from the EU allowed the civil servants to conduct the experiment. A clear indicator of the importance of financial support is the fact that the further implementation of the action plan has been put on hold while new sources of funding are being identified. Additionally, the experiment participants received support in the form of a risk assessment that served as a starting point for the design of flood risk mitigating solutions. This was an important tool for defining the measures to be taken in order to increase the resilience in the neighborhood.

Institutional alignment
Because the experiment was led by the public actors instead of the private actors, no institutional barriers were observed. Because of its small scale, the implemented solution remained within the boundaries of the institutional framework. Institutional alignment is therefore defined as: high.

At the tree planting event, the deputy mayor was also present, as well as a number of civil servants. This was highly appreciated by the citizens, as it showed them that their actions were noticed and appreciated by the entire municipality. This further increased the trust in the municipality.

Leadership
The entire experiment was led by a small group of civil servants. Because the municipality and the citizens did not have any experience with bottom-up experimentation, the civil servants were vital in coordinating and steering the process of experimentation.

7.2.2. Activities features (Appendix E.2)
Transparency
As all experiment participants were satisfied with the communication, transparency is defined as: high. This contributed to building trust between the experiment participants and increased the participants' satisfaction with the collaboration process.

Collaboration
The process of collaboration was set up in such a way that the civil servants from the different municipal departments were the most active stakeholders. The stakeholders participating in the 6 meetings also took on an active role by helping design the action plan. In doing so, they became more committed to implementing the action plan, as it reflected their own interests and ideas. In order to engage the citizens, two methods were used. Local schoolchildren were taught about urban resilience and the experiment by their teachers. The children, in turn, relayed the information to their parents. The second method was going door-to-door to discuss the topic with the citizens and make them aware of the experiment, aided by some local citizens. Neither of these forms was as much collaboration as it was information sharing. It was a necessary method to build awareness, however. A benefit of such an intensive approach, was that most of the citizens in the neighborhood, including the less privileged stakeholder groups, came into contact with the experiment. Stakeholder inclusion is therefore defined as: high. In order to involve the citizens more actively in the experiment, they were allowed to choose where the first solution was to be implemented by participating in a neighborhood meeting. They were then also asked to help plant the trees during a festive event. This turned out to be a good method to put neighborhood resilience on the citizens' agendas and show them how they could

contribute to small-scale solutions. The approach was highly appreciated by the citizens, as they felt their needs were taken seriously by the municipality. Both the design of the action plan and the decisions about the small-scale implementation were largely based on stakeholder consensus, which was highly novel for the Burgas municipality. Additionally, the civil servants also appreciated all the help and input from the Resilient Europe network, which allowed them to adopt best practices from cities with more experience with this type of collaborative experimentation.

Reflexivity and Learning
Throughout the experiment, there was no explicit focus on formalized learning goals and mechanisms. Rather, learning took place quite intuitively. The design of the action plan was a learning process in which the system was defined and normative ideas about the problems and solutions were discussed and combined into a widely-accepted action plan. The small-scale implementation of a solution was a method to learn more about possible approaches towards implementing the action plan and the outcomes were used to refine the action plan. The Resilient Europe network meetings also inspired the civil servants to try different approaches over the course of the experiment. These meetings were also opportunities for reflection on the process itself and the direction of the process.

Sequence of actions
The most actions of the process were sequenced in a meaningful manner. The first meeting dealt with a very broad topic, and over the course of the meeting, this topic was narrowed down further. A small-scale experiment was planned in the middle of the process to help with this process of narrowing down. The only remarkable aspect of the sequencing was the inclusion of the neighborhood's residents. Most of them were only involved after the plan for the experiment and the general direction of the action plan had already been defined. Sequence of actions has therefore been defined as: medium

Methodology
Detailed methods for conducting the experiment were used and applied with the help from the Resilient Europe network and the municipality's own experience with public consultation processes. However, there were some elements of the experiment that were the result of a trial-and-error approach rather than a predefined method. Especially the approaching and engaging of stakeholders required practical experience rather than theoretical knowledge. Methodology has therefore been defined as: medium.

7.2.3. Output features (Appendix E.3)
Built capacities
For the civil servants, the process of experimentation was a useful experience, which helped them gain skills such as facilitating consensus-based decision-making and engaging citizens in policymaking. This was the result of learning-by-doing and built confidence, because of the success of the experiment. Citizens, on the other hand, do not feel they built any capacities. This can be linked to their passive role in the experiment. Built capacities has therefore been defined as: medium.

Accountability
The citizens reported that they would like to continue to collaborate with the municipality to mitigate flood risks as they feel their knowledge can contribute to the implementation of solutions that can successfully mitigate local flood risks. This is a change from their previous unengaged attitude and a sign that some sense of problem ownership has been created.

However, this engagement still relies on a top-down approach towards flood risk management, no bottom-up initiatives have been taken yet. Accountability has therefore been defined as: medium.

Social change
Relating to accountability, no new networks have been formed in order to continue the experiment. Unofficially, a new network between the citizens and the civil servants can be observed, but this network has not been active since the end of the experiment. The experiment was highly successful in changing the stakeholders' perspectives, however, as both citizens and civil servants report they now have a more positive view of each other and see opportunities for further collaboration. Both stakeholder groups now have a better concept of neighborhood resilience.

Physical change
During the small-scale implementation of a solution, 5 willow trees were planted at the neighborhood's central square. This is just one of the nature-based solutions identified in the action plan and the implementation was confined to one location. Nevertheless, this change helped to demonstrate what change could be achieved and helped build public support for the action plan.

Actionable knowledge
A highly-detailed action plan has been defined that has been adapted according to the small-scale solution implementation's outcomes. Concrete and measureable actions to increase the neighborhood's resilience have been defined. Actionable knowledge has therefore been classified as: high.

Transferability
Lessons for transferring the experiment, such as personal stakeholder contact and small-scale solutions, have been defined and shared with the different municipal departments. So far, they have not yet been applied. Transferability has therefore been defined as: medium. Some possible challenges for transferring the lessons could be a lack of interest from the citizens and the high intensity of the experiments, which might require too much time and effort from the civil servants.

Scalability
Much like the lessons for transferring the experiment, lessons for scaling up the experiment have been defined, but not yet validated. Scalability has therefore been classified as: medium. The most important lesson for scaling out defined by the civil servants is repeating the same process of experimentation on different locations in the neighborhood in order to facilitate a step-by-step implementation of the action plan. For scaling out, the neighborhood's residents have already been contacted. For scaling up, however, more and new stakeholders would have to become involved. A worry is that this would require more time and effort from the civil servants, who were unsure they would be able to put in all this time and effort.

Accounting for consequences
Two unintended consequences have been explicitly considered and mechanisms for their mitigation have been defined. No explicit activities to identify unintended consequences could be identified, however.

7.2.4. Outcome features (Appendix E.4)

Socio-ecological integrity
The civil servants are highly committed to expanding the number of nature-based solutions in the neighborhood in addition to other flood risk mitigation measures. Their targets have been formalized in the action plan. They have also shared their positive experiences with focusing on different aspects of an issue with their colleagues from different municipal departments. The citizens, on the other hand, although pleased with the restoration of the neighborhood's ecosystems, have not yet planned or undertaken any action to contribute to these solutions on their own. Socio-ecological integrity is therefore defined as: medium.

Intergenerational equity
Due to the effortless inclusion of the entire neighborhood during the experiment, no active plans have been made to specifically include less privileged stakeholder groups in the future. Implicitly, however, they already are included in the ambitious indicators of citizen reach and inclusion defined in the action plan, which would make it near impossible to exclude these stakeholder groups. Intergenerational equity has therefore been defined as: medium.

Socio-ecological stewardship and governance
Although both the citizens and the civil servants indicate that a bridge of trust has been built between the two parties and a willingness to collaborate in the future has been expressed, no formal initiatives or agreements have been made yet. For now, these plans to involve citizens in future, resilience-related decision-making remain informal.

Resource efficiency
Trees from a local nursery were used for the implementation of one of the solutions. As the energy needed to transport the trees was limited, the resource efficiency can be classified as: high. Additionally, the trees chosen are aquatic species that thrive in the, often waterlogged, location. They also require minimal upkeep, which contributes to the resource efficiency.

Precaution and adaption
The action plan created focuses on a number of different flood risk mitigation solutions, often overlapping or complementing each other. This is a risk-averse approach. Additionally, indicators for monitoring the impact of the solution have been defined, which should allow for timely adaption in case of any issues. Precaution and adaption is therefore defined as: high.

7.3. Intermediate conclusion Burgas
The goal of the experiment was to design a consensus-based action plan to increase the resilience against climate change threats in the neighborhood. By facilitating intensive collaboration between different stakeholders, the experiment has succeeded in doing so. The action plan is a useful tool for increasing the neighborhood's resilience by focusing on different solutions, not just the most common ones. Additionally, the small-scale implementation of the solution was meant to actively engage the neighborhood's residents in building resilience. Although the experiment did succeed in increasing the citizens' interest in the topic and their willingness to collaborate with the municipality, no active empowerment could be observed. Their limited active involvement in designing the action plan and the related solutions could have caused this outcome. However, the citizens' and municipality's lack of experience with bottom-up approaches could also have been a factor in the experiment's failure to empower the citizens. The role of the context and input factors could have played a major role in the outcome of the experiment.

The indicators used for the appraisal of the experiment did not always succeed in capturing the essence of the features. The feature intergenerational equity, for example, has

been classified as medium. However, this is not the result of a conscious exclusion or oversight, but of a lack of need to actively include less privileged stakeholder groups based on positive previous experiences. Likewise, the features for transferability and scalability indicate how reliable to lessons learned are, but fail to identify the actual lessons learned, even though these lessons are much more interesting from a governance perspective than a 'ticked box' of their presence. Similarly, some features contain a wealth of information, whereas other features are merely supporting features of those information-rich features. These are all indicators that the operationalization of the evaluative framework needs to be carefully assessed when analyzing and comparing the two cases.

Evaluation scheme Green Belt experiment, Burgas			
Input	**Activities**	**Output**	**Outcome**
Awareness: • Medium Commitment: • Medium Trust: • Low Social capital: • 3 Networks included Expertise: • 3 Experts included Support: • €45.500,- • Risk assessment Institutional alignment: • High Leadership: • High	Transparency: • High Collaboration: • 6 Council meetings • 1 Neighborhood meeting • 4 Network meetings Decisions: • 3 Major decisions made Stakeholder inclusion: • High Stakeholder satisfaction: • High Reflexivity and learning: • Medium Sequence of actions: • Medium Methodology: • Medium	Built capacities: • Medium Accountability: • Medium Social change: • 0 New networks Changed perception: • High Physical change: • 5 Trees Actionable knowledge: • High Transferability: • Medium Scalability: • Medium Accounting for consequences: • 2 Consequences accounted for	Socio-ecological integrity: • Medium Intergenerational equity: • Medium Socio-ecological stewardship and democratic governance: • 0 New participatory settings Livelihood opportunity: • Not identified Resource efficiency: • High Precaution and adaption: • High

Table 7.1: Evaluation of Burgas experiment

7.4. Individual Case Analysis: The 'Green Corridor' Experiment, Antwerp

The Sint-Andries neighborhood is located in Antwerp city center, bordering the Scheldt river. It has a high building density, with few public and private green spaces available. Although the neighborhood has currently had limited experiences with floods, its location near the river and the high building and people density makes it vulnerable to the impact of climate change. In order to increase the neighborhood's resilience to these threats, the area had been selected for an experiment. The main goal of the experiment was to not only increase the resilience of the place, but also of the people. The experiment was therefore heavily focused on stakeholder empowerment. Civil servants, project leaders of the Resilient Europe network and an experienced process councilor collaborated to design the basic set up and goals of the experiment. Specific learning questions and methods were designed in advance and linked to different aspects of the experiment in order to increase the experiment's impact. A group of experts was asked to identify the main issues and related goals in the neighborhood. This information could be used to support the co-creation process. A budget of €46.000 was made available through EU funding and municipal funding.

Preparation Phase

In order to determine which topic to focus on, informal meetings were first held with different stakeholder groups in the neighborhood, such as the social housing association and youth organizations. This helped define the contours of the experiment. The first challenge then was to engage the neighborhood's residents. Although the process councilor already had an extensive network in the neighborhood, engaging the residents still required a lot of effort. By approaching interesting stakeholders or stakeholder groups individually, a group of 44 residents could be convinced to participate in the first meeting. Their motivations for joining this meeting varied, but many of them were more concerned about their direct living environment than about the threats of climate change. They reported to have joined the experiment because they felt the municipality could not be trusted to address living environment-related issues in the neighborhood. During this first meeting, the residents were informed about the experiment and were asked to identify the issues that were most important to them. Two follow-up meetings were then organized to allow the residents to identify actions and pathways to address these issues. These pathways were aimed at addressing different aspects of climate resilience, such as social and infrastructural resilience. The idea of the Green Corridor, a corridor connecting different nature-based solutions throughout the neighborhood, was designed.

Experimentation Phase

A group of residents volunteered to be responsible for the implementation of the Green Corridor. This project group met five times over the course of the experiment to discuss the project and plan the implementation of different solutions. The project group was coached by the process councilor and the civil servants. The process councilor applied for and received and additional €3.000 from the city's participatory budget to make a start at implementing the Green Corridor. A meeting to Rotterdam was organized to inspire both the experiment participants and local politicians. Throughout the experiment, the citizens and process councilor tried to involve the local politicians, but there was limited interest for the project or the issue of climate change. This only enforced the citizens' belief that the municipality would be of little help to them when making their neighborhood climate resilient. In collaboration with the project group, civil servants and process councilor, a "dream day" was organized in the local community center where residents were invited to design their ideal neighborhood and link these dreams to the Green Corridor project. This event was followed up with a "do day", where all residents were invited to help implement some small-scale solutions, such as depaving squares and building planters. In addition to permanent solutions, some temporary solutions, such as

artificial grass and a fake pond were used to show what would be possible in the future. Plants from local nurseries were used and some plants from other areas of the city were 'recycled'. This event was organized at the same time as the city-wide "car free day", which drew a lot of visitors to the neighborhood. Both events were highly appreciated by the neighborhood. They helped people visualize what could be achieved in the neighborhood and how they could contribute to it. It helped to commit residents to the idea of a Green Corridor and created a sense of problem ownership. The experiment was concluded with an evaluation, conducted by an independent evaluator.

Outcomes
Both the citizens and the civil servants reported that they were very satisfied with the process of collaboration and the outcomes. Participating in the experiment was a useful experience for all stakeholders. The civil servants reported to have developed the skills to facilitate co-creation processes and indicated to have gained a new perspective on the capabilities of citizens to tackle such abstract issues as climate change. Citizens, on the other hand, learned how to organize themselves and to design and implement solutions that mitigate climate risks. They indicated that participating in the experiment made them aware of the threats of climate change and the issue is now high on their agendas. The citizens have established an association, Klimaatrobuust Sint-Andries ("Climate Resilient Sint-Andries"), in order to continue to implement climate risk mitigating solutions. In addition, stakeholders report that, inspired by the experiment, residents have also started to take private measures to mitigate the threat of climate change, for example by considering rainwater drainage when remodeling their house. A monthly meeting with the civil servants has been established to discuss progress and issues. The pathways and actions have been adjusted based on the outcomes of the experiment and both citizens and civil servants are committed to meeting the predefined goals.

Lessons learned
Lessons from the experiment, like giving the citizens the freedom to design their own solutions and conducting small-scale demonstration projects, have already successfully been applied in different projects. In addition to the citizens planning to scale out their current plans, a co-creation approach similar like the one used in this experiment will be used to design the city-wide water plan. Overall, the experiment has been quite successful in both engaging and activating citizens in climate change mitigation activities and in generating actionable knowledge the municipality can apply in different projects. Some issues with the experiment could be identified, however. Firstly, less privileged stakeholder groups were not involved in the experiment. This was the result of a conscious choice to focus on the frontrunners and achieve quick results. There are no plans for their involvement in the future, but their interests are being considered in by the Klimaatrobuust Sint-Andries association, for example by including the. Second, the experiment exposed a number of issues previously not identified. These include unclear ownership rights and maintenance responsibilities as well as a highly restrictive institutional environment that limits the possibilities to experiment. The experiment was able to bypass these issues by focusing on small-scale solutions, but for the upscaling of the experiment, these could become more important.

7.5. Enabling factors and barriers for experimentation Antwerp
In this paragraph, the case will be structured according to the logic model evaluation framework. Extensive argumentation for choices made and examples are included in Appendix F.

7.5.1. Input features (Appendix F.1)

Awareness

Citizens were largely unaware of the threat of floods and other climate change related disasters in the neighborhood. In the framework, awareness has therefore been classified as: low. However, the reason most citizens initially joined the experiment, was because it gave them an opportunity to improve on their living environment. The lack of awareness of the threat of climate change was neutralized by the awareness of the need to adjust the living environment.

Over the course of the experiment, being exposed to information about the threat of climate change and discussing the topic with citizens that did find climate change resilience and important issue, many of the citizens changed their perspective.

Commitment

As only a relatively small group of citizens was involved throughout the process, with other citizens joining at different moments, commitment has been defined as: medium. This set-up worked very well, however, with the relatively small project group being able to take decisions quickly and a larger group of citizens helping out when needed or possible. The experiment participants plan to continue in a similar fashion, with a small group of 'planners' and a larger group of 'doers'.

Throughout the experiment, new stakeholders started to join the meetings. The experiment participants attributed this largely to word-of-mouth advertising and highly visible changes that were being implemented.

Trust

There was a high level of mistrust in the neighborhood, due to previous bad experiences with the municipality. Trust has therefore been classified as: low. However, this was a driver for participation in the experiment rather than a barrier, because citizens felt compelled to join the experiment to make sure their needs were met.

This feeling did not change over the course of the experiment due to the lack of interest from the city council in the experiment. They did start to trust the civil servants, however, due to their satisfaction with the collaboration. The civil servants, on the other hand, came to realize how large the gap was between the city council's interests and their plans to increase the neighborhood's resilience, making them even more committed to facilitating bottom-up approaches.

Social capital

A lot of different networks were included in the experiment, including a number of networks that were already active in the field of either bottom-up initiatives or climate change resilience, like, for example, the local gardening and biodiversity associations. This not only helped to engage citizens, but also meant that a lot of knowledge and skills were already present in the neighborhood. The process councilor, who had a lot of knowledge of the different local networks, was vital in engaging and connecting all of the relevant networks and stakeholders and seeing opportunities for the experiment.

Throughout the experiment, the networks proved useful in reaching and activating a large number of citizens. The process councilor's knowledge of the neighborhood and the municipality helped solve a number of issues that someone with only top-down or bottom-up knowledge could not have solved.

Expertise

A number of experts were included in the experiment, with their roles ranging from helping to design the experiment to providing input or guiding the process. Again, the process councilor,

with ample experience with co-creation processes, was a vital expert in designing and conducting the experiment.

Support
The financial support was an important driver for the experiment, which was clearly visible when new sources of funding had to be found in order to implement some solutions. Additionally, the coaching sessions supported the citizens to start acting independent of the civil servants and the process councilor. These were a form of support that directly influenced their decision to establish a new neighborhood association to continue the experiment.

Institutional alignment
The experiment encountered a number of institutional barriers when planning to implement some solutions, mainly due to local regulations. By negotiating, adapting and keeping the solutions small, they were able to implement some of their planned solutions, however. Institutional alignment is therefore defined as: medium. For the continuation of the experiment, it is expected that this highly regulated environment will make it even more difficult to implement solutions.

   The lack of interest from local politicians in the experiment was highly frustrating for all experiment participants, both top-down and bottom-up. On the one hand, this was experienced as a barrier, as an alignment with the city council's policies could enhance the impact of the experiment and make it easier to experiment in the future. On the other hand, however, the lack of political interest did convince the participants to commit to the continuation of the experiment.

Leadership
Initially, the experiment was led by the process councilor in collaboration with the civil servants. As the experiment progressed, however, the citizens started to take on more responsibility. This was a result of the experiment design, which was set-up in a way to promote problem ownership by giving citizens a blank slate to design their own solutions and encouraging them to start taking responsibility for their own solutions.

   At the end of the experiment, the citizens have organized themselves in an association and have taken over the experiment. The association is led by a board that now acts as the leader.

7.5.2. Activities features (Appendix F.2)
Transparency
As all experiment participants were satisfied with the communication, transparency is defined as: high. This contributed to building trust between the experiment participants. A barrier for the experiment, however, was the lack of transparent municipal information. The experiment participants indicated that it was very difficult for them to make sense of funding opportunities, permits and regulations on their own, which slowed down the process.

Collaboration
The process of collaboration was set up in such a way that the citizens took on the most active role, making the important decisions, and the civil servants and the process councilor took on a more facilitating role. This was done in order to build a sense of problem ownership and an increased willingness to take responsibility for the implementation of different measures. An action plan was designed to identify actions, goals and pathways to make the neighborhood climate resilient and was used as a tool for guiding the collaboration and assessing the designed solutions. In order to engage more residents in the experiment, a "dream day" was organized as

an approachable way to link their dreams for the neighborhood with climate change resilience. Subsequently, a "do day" was organized to show what a climate resilient neighborhood could look like and to emphasize the potential of the experiment. Both days were good mechanisms for building public support and informing and inspiring residents. Involving citizens in the Resilient Europe network meetings was not only a good way for them to learn, reflect and get inspired, it also emphasized that they were taken seriously by the civil servants, which helped to commit them to the experiment. Overall, the experiment participants were highly satisfied with the process of collaboration.

The active citizens, however, were a largely homogeneous group, stakeholder inclusion has therefore been defined as: low. Less privileged stakeholder groups did not participate in the experiment. This was the result of a trade-off between focusing on generating output or focusing on inclusion.

Reflexivity and Learning
Detailed learning goals and related mechanisms for learning had been designed. Both the formulation of learning goals and the formulation of mechanisms for learning have therefore been classified as: high. This helped the civil servants to focus their efforts on learning about the most relevant aspects of the experiment, which led to a number of detailed lessons for upscaling or transferring. Additionally, social learning took place through the processes of collaboration in which sharing information and discussing different perspectives led to a shared vision on a climate resilient neighborhood. Aside from an evaluation at the end of the experiment, there was little room for reflection. The Resilient Europe network meetings, however, provided some opportunities for reflection that were highly appreciated by the experiment participants.

Sequence of actions
Each of the actions or collaboration mechanisms built on the previous actions. Additionally, they were timed each month, which made sure the experiment stayed high on the participant's agendas, without overburdening them. Some actions were timed to coincide with other events in order to enhance their impact. This is why sequence of actions is classified as: high.

Methodology
Detailed methods for conducting the experiment were used an applied. However, there were some elements of the experiment that were the result of a trial-and-error approach rather than a predefined method. Especially the approaching and engaging of stakeholders required practical experience rather than theoretical knowledge. Methodology has therefore been defined as: medium.

7.5.3. Output features (Appendix F.3)
Built capacities
Both the citizens and the civil servants indicated to have gained or enhanced skills and capacities. This was mainly the result of learning-by-doing, which helps build skills as much as it builds confidence. Built capacities is defined as: high.

Accountability
The citizens indicated that participating in the experiment and defining a shared perspective on the importance of climate change resilience helped build a sense of accountability. An external driver, the lack of trust in the city council, however, could also be identified as an important enabling factor in their sense of problem ownership. Accountability is defined as: high.

Social change
The citizens have united themselves in an association in order to continue the experiment. Learning about the future challenges and social learning to come to a shared vision both helped them change their perspective on climate change resilience and their role in it. The positive experiences from the experiment also helped to change the perspective of the civil servants on the abilities of citizens to tackle these issues and how they could be involved in future projects. Social change is therefore defined as: high.

Physical change
At the end of the experiment, a number of different small-scale solutions had been implemented. This turned out to be instrumental in keeping the participants committed, as their efforts were translated in tangible results, as well as building public support, awareness and interest in the experiment.

Actionable knowledge
The result of the methods used and the mechanisms for collaboration was a highly-detailed action plan, translating the shared vision for the neighborhood into concrete and measurable actions. Actionable knowledge has therefore been defined as: high.

Transferability
Lessons learned have already been successfully applied in different projects. Especially the open agendas and small-scale demonstration projects were considered to be best practices that can be applied elsewhere. Transferability has therefore been defined as: high. Some challenges for transferring the lessons, however, could be a lack of frontrunners in a neighborhood and the high intensity of the process of engaging stakeholders, which requires a lot of time and effort.

Scalability
Lessons for scaling up have also been defined and are mainly focused on repeating the experiment in the neighborhood to take small, cumulative steps to change. Collaboration between the municipality and the citizens is seen as even more important as larger projects are thought to encounter more issues with ownership, regulations and funding. Neither of these lessons have been validated yet, however. Scalability is therefore defined as: medium.

Accounting for consequences
Some unintended consequences have been considered. Some were the result of the learning questions, whereas others became apparent during the process of experimentation. There were no specific mechanisms for identifying and accounting for unintended consequences included in the experiment, however.

7.5.4. Outcome features (Appendix F.4)
Socio-ecological integrity
Both the citizens and the civil servants indicated to be committed to implementing more nature-based solutions in the neighborhood, by being active in the new association or supporting this association. The action plan will be used as a guiding document for new initiatives. Citizens also reported that residents started making changes to their private properties as a result of being exposed to information about or participating in the experiment. Socio-ecological integrity is therefore defined as: high

Intergenerational equity

No plans have been made to actively include the less privileged stakeholder groups in the further expansion of the 'Green Corridor' and the implementation of the action plan. Instead, the participants plan to continue to focus on frontrunners and quick actions. Intergenerational equity has therefore been classified as: low. These less privileged groups, however, will receive the same information and invitations as other residents in the neighborhood and their living environments will still be included in the implementation of the action plan.

Socio-ecological stewardship and governance

A monthly meeting between the citizens and civil servants has been set up for to continuation of the experiment. Although the experiment was aimed at empowering the citizens to take responsibility for climate change resilience, the process of collaboration thought that a coalition of top-down and bottom-up actors could enhance each other's capacities and mandates and a continuation of collaboration is therefore desired.

Resource efficiency

As local materials were used for the implementation of the solutions, the resource efficiency has been classified as: high. However, much of the plants have died since and will need to be replaced by new plants. For the upscaling of the experiment, more resilient plants will therefore be selected in order to increase the resource efficiency.

Precaution and adaption

As the action plan, as a result of the method to focus on different aspects of resilience, contains a broad range of complemental and overlapping actions, the action plan provides a risk-averse solution to climate change threats. Precaution and adaption is therefore defined as: high.

7.6 Intermediate conclusion Antwerp

The goal of the experiment was to empower citizens to become more proactive in helping their neighborhood become more climate change resilient. By giving the citizens the freedom to design their own measures and helping them to implement these solutions themselves, the experiment has succeeded in doing so. Aligning the experiment with existing networks and combining it with other wishes from the citizens, the experiment succeeded in ensuring public support and committing citizens to the experiment. A shared vision document turned out to be a useful tool in communication the experiment's goals and guiding the upscaling of the experiment. Although much of the outputs could be traced back to carefully designed learning and collaboration mechanisms and methods, influential context factors could also be identified. The clearest example of which is the lack of trust in the city council, which was a strong driver for citizens to join the experiment. This is a warning signal that input factors can play a big role in the success of the experiment, no matter how carefully designed.

When assessing the indicators used to evaluate the framework's features, some observations can be made. In some cases, the scales appear to not adequately reflect the feature. One such example is the commitment scale, which classifies full commitment from all stakeholders as the highest scale. This case demonstrated, however, that this scale does not reflect the trade-off between process manageability and inclusion. A similar observation could, for example, be made for the scale of stakeholder inclusion. Other features failed to look at aspects that were marked as relevant in the cases. The feature of social capital, for example, fails to highlight the importance of a mediator linking the municipality and the citizens. Additionally, some features appear to have been more influential than others. All of these assessments are warning signals that the operationalization of the evaluation framework might not yet be satisfactory.

Evaluation scheme Green Corridor experiment, Antwerp			
Input	**Activities**	**Output**	**Outcome**
Awareness: • Low Commitment: • Medium Trust: • Low Social capital: • 5 Networks included Expertise: • 4 Experts included Support: • €49.000, - • 3 Coaching sessions Institutional Alignment: • Medium Leadership: • High	Transparency: • High Collaboration: • 7 Public meetings • 5 Private meetings • 4 Network meetings Decisions: • 3 Major decisions made Stakeholder inclusion: • Low Stakeholder satisfaction: • High Reflexivity and learning: • High Sequence of actions: • High Methodology: • Medium	Built capacities: • High Accountability: • High Social change: • 1 New network Changed perception: • High Physical change: • 30m2 depavement • 4,5m2 living pavement • 20m2 plants • 120L water reservoir Actionable knowledge: • High Transferability: • High Scalability: • Medium Accounting for consequences: • 2 Consequences accounted for	Socio-ecological integrity: • High Intergenerational equity: • Low Socio-ecological stewardship and democratic governance: • 1 New participatory setting Livelihood opportunity: • Not identified Resource efficiency: • High Precaution and adaption: • High

Table 7.2: Evaluation of Antwerp experiment

8. Cross-case analysis

Now that the two cases have been introduced and individually analyzed, the cases will be compared to each other in order to confirm or contest preliminary findings. As the goal of the framework should be to facilitate the appraisal of sustainability experiment and identify governance practices that contributed to the success or failure of the experiments, the analysis will start with the outputs and work back from there to identify different practices that contributed to these outputs. The focus of the analysis is therefore not only on the features, but also on the relationship between the features. The *process* of experimentation will be analyzed in addition to the *elements* of the experiment. After the cross-case comparison, some observations about the governance of transition experiments as well as the evaluation framework will be discussed.

8.1. Experiment output

8.1.1. Built capacities

Burgas			
Input	*Activities*	*Output*	*Outcome*
Lack of social capital	Collaboration allowed citizens to play a passive role "Learning-by-doing"	Medium: civil servants did build capacities, citizens did not	Medium socio-ecological integrity: civil servants continue experimentation, citizens do not
Antwerp			
High social capital	Collaboration forced citizens to take an active role Coaching "Learning-by-doing"	High: civil servant and citizens built capacities	High socio-ecological integrity: civil servants and citizens continue experimentation

Table 8.1: Analysis built capacities

Compared to Antwerp, Burgas was not as successful in building the capacities of the stakeholders involved in the experiment. As a result, the citizens in Burgas did not acquire any skills that would have helped them to take any bottom-up initiatives to continue to build resilience in the neighborhood. The citizens in Antwerp, on the other hand, reported to have acquired skills that made them feel more confident in their own ability to scale out the experiment. They indicated to have already been applying these skills within their new association. When tracing back these outputs to the activities, the main difference that can be observed are the mechanisms of collaboration between the two experiments. In Antwerp, citizens were involved from the first meeting. They were given the freedom to design their own action plan and were encouraged to take on an active role in the implementation of the action plan. As a result, the citizens in Antwerp had a lot of hands-on experience with designing and implementing sustainable solutions, which they stated helped them build their capacities. Additionally, the citizens received coaching to help them develop the skills needed to do so. In contrast, in Burgas, only about 15% of the consultative council responsible for the design of the action plan and the small-scale implementation was made up of citizens. For the most part, the experiment was the responsibility of the civil servants. The citizen's active involvement in the experiment was highly limited and largely symbolic. Their passive participation meant that they did not build any new capacities, because they had no need to apply them.

The civil servants participating in both experiments reported to have built similar capacities. The civil servants reported to now have the skills to engage citizens, encourage them to think about the topic of resilience and guide a process of consensus-building. In both cases, they reported that they gained these skills by adopting a "learning-by-doing" approach. Similarly, they reported that successfully applying these skills has helped to build the

confidence to try to apply them elsewhere. When comparing these statements to the activities, one can indeed conclude that the civil servants in both cases performed largely similar activities, like organizing and guiding the process of collaboration and engaging citizens.

It should be noted, however, that a major difference between the two experiments exists. Where the citizens in Antwerp had a lot of experience with bottom-up initiatives, the citizens and civil servants in Burgas had virtually none. This not only resulted in a top-down focused experiment approach, but might have led to different outcomes had the experiment in Burgas applied similar collaboration mechanisms as the experiment in Antwerp. One can pose the question if the less successful output of the experiment in Burgas can be attributed to the governance of the experiment, or to the context in which the experiment took place.

8.1.2. Accountability

Burgas			
Input	*Activities*	*Output*	*Outcome*
Lack of social capital	Collaboration allowed citizens to play a passive role Low visibility experiment Social learning took place	Medium: citizens feel more responsible, but remain responsive	Medium socio-ecological integrity: civil servants continue experimentation, citizens do not
Antwerp			
High social capital Low level of trust Medium institutional alignment	Collaboration forced citizens to take an active role High visibility experiment Social learning took place	High: citizens feel more responsible and have become proactive	High socio-ecological integrity: civil servants and citizens continue experimentation

Table 8.2: Analysis accountability

As a reminder: accountability in this context is interpreted as a feeling of responsibility for achieving sustainable change. Similar to the built capacities, the experiment in Antwerp was more successful in building a feeling of responsibility than the experiment in Burgas. The citizens in Burgas stated that they would like to continue to collaborate with the civil servants on increasing climate change resilience in the neighborhood, but the civil servants reported that, so far, the citizens have no plans to take on a more active role. In Antwerp, on the other hand, the citizens reported that they feel a strong sense of accountability, which is a driver for them to continue the experiment. Additionally, they observed that individual citizens have started to take responsibility for their own role in climate change resilience by adapting their private spaces, for example by considering the water drainage when remodeling their home.

When comparing the activities that could have led to this difference, the methods used for collaboration are the first feature to be assessed. The citizens in Antwerp were allowed to design their own action plan, which they reported helped them build a strong sense of responsibility for also implementing it. Additionally, the experiment was highly visible, with public events, flyers, posters and articles and items in the local news outlets. By emphasizing that the citizens themselves are the driver of change, the neighborhood's residents became aware of their own role in climate change resilience. In Burgas, on the other hand, the experiment was organized top-down, which emphasized the role of the municipality as opposed to the role of the citizens. Additionally, the citizens were only confronted with the experiment at a few events, all of which were facilitated top-down. Although the citizens did learn about

climate change resilience and intensified collaboration through this exposure, it was only partially successful in building a strong sense of accountability.

However, when looking at the input, the same observation as for built capacities can be made: there was a large difference in the social capital between the two neighborhoods. In Burgas, the citizens had no experience with taking responsibility for public issues, whereas the citizens in Antwerp had plenty. The citizens in Antwerp also reported they felt they had to take on an active role, as they learned during the experiment that their wishes and needs were not aligned with the vision of the city council. This might have contributed to the different outcomes in addition to the differences in activities.

8.1.3. Social change

Burgas			
Input	*Activities*	*Output*	*Outcome*
Lack of bottom-up knowledge, but use of a mediator Medium awareness Low level of trust High institutional alignment	Collaboration allowed citizens to play a passive role Low barrier for participation Transparency Methodology aimed at social learning	0 New networks High changed perception	Medium socio-ecological integrity, 0 new participatory settings High equity
Antwerp			
High bottom-up knowledge and use of a mediator Low awareness Low level of trust Medium institutional alignment	Collaboration forced citizens to take an active role Higher barrier for participation Transparency Methodology aimed at social learning	1 New network High changed perception	High socio-ecological integrity, 1 new participatory setting Low equity

Table 8.3: Analysis social change

In terms of social change, the experiment in Antwerp achieved a higher level of social change than the experiment in Burgas. The experiment in Burgas was highly successful in changing the perspectives of the participants. It was less successful, however, in establishing new, permanent, networks and participatory settings. Although the indicator for intergenerational equity classifies the output of equity in the Burgas experiment as medium, the experiment did succeed at including and changing the perceptions of all stakeholders, including the less privileged groups. The experiment in Antwerp succeeded in changing the perceptions of all participants. Additionally, one new association was founded during the experiment and a formal collaboration between the association and the civil servants has been established. The experiment in Antwerp failed, however, to include less privileged stakeholder groups in the experiment and the experiment participants have not expressed an interest in including those groups in the future.

In both cases, some of the methods used were aimed at social learning. By informing citizens about climate change resilience and encouraging them to think about and discuss the topic, citizens became aware of the importance neighborhood resilience. Open collaboration between the citizens and the civil servants and consensus based decision-making helped build trust between the two stakeholder groups, with the citizens observing that their needs are taken seriously and the civil servants observing the willingness of the citizens to contribute to resilience. Similar to the built capacities and accountability, however, the changed perspectives in Burgas did not lead to any tangible changes in the behavior of the citizens, whereas in Antwerp, a new neighborhood association was created. Although the civil servants in Burgas

indicate that they would like to continue to involve citizens more actively in policymaking, no formal participatory settings have been established. In Antwerp, on the other hand, a monthly meeting between the civil servants and the new neighborhood association has been established. This difference can, again, be attributed to the difference methods of collaboration, with the methods in Burgas focusing on inclusion and those in Antwerp on empowerment. In terms of stakeholder inclusion, the experiment in Burgas was therefore much more successful than the experiment in Antwerp. A possible explanation for this difference can be that the methods of collaboration in Antwerp, which required commitment and effort from the stakeholders, might have acted as a barrier for participation. This is illustrated by the high interest of the neighborhood's residents to participate in the open "dream" and "do" days, but a much lower interest in committing to the more intensive design process. Similarly, the more symbolic, less taxing collaboration between the civil servants and the citizens in Burgas might have removed barriers for participation.

When comparing the different outcomes to the input features, some observations can be made. The differences between local knowledge of bottom-up initiatives seemed to have influenced the outcomes. When expanding the definition of social capital, however, in both cases a local mediator appeared to have played an important role in facilitating the collaboration by engaging citizens. Similarly, networks played an important role in both cases in spreading information and gaining credibility, although more effective in Antwerp than in Burgas.

The high awareness of flood risk in the neighborhood helped citizens in Burgas to commit to the experiment and over the course of the experiment, they learned they could link the experiment to their living environment as well. The lack of awareness of flood risks in Antwerp, however, was not a barrier for commitment, as citizens did have a high awareness of other factors relating to flood risk mitigation solutions, such as the need for more green spaces in the neighborhood. Over the course of the experiment, they learned about linking those solutions to climate change resilience. Different awareness levels therefore led to similar outcomes through similar methods. Although in both cases a similar low level of trust could be observed, the impact of this trust level differed between the cases. In Burgas, it created an additional barrier that prohibited citizens from participating in the experiment. By collaborating with the citizens as well as by showing the municipality's interest in their wellbeing, a bridge of trust could be built that encourages citizens to participate in future experiments. In Antwerp, the low level of trust in the municipality was actually a driver for participation. Although the trust between the citizens and the civil servants actually increased during the experiment, both the citizens and the civil servants reported that they felt let down by the lack of interest from the city council and the institutional barriers they encountered. This, however, only made them more determined to take action on their own.

8.1.4. Physical change

Burgas			
Input	*Activities*	*Output*	*Outcome*
Commitment Expertise Support	Collaboration Methodology	5 Trees planted	High resource efficiency Socio-ecological integrity
Antwerp			
Commitment Expertise Support	Collaboration Methodology	55m2 of various solutions implemented	Medium resource efficiency Socio-ecological integrity

Table 8.4: Analysis physical change

In Burgas, 5 trees were planted, with an additional 35 planned to be planted. The trees came from local nurseries and were specifically selected to survive in wet environments, which will help them withstand future floods. In Antwerp, a number of different measures, relating to depavement, the use of plants and trees and water storage, have been implemented, adding up to about 55 square meters of physical change. A number of new projects aiming to increase the number of nature-based solutions in the neighborhood have already been planned. In contrast to Burgas, however, the experiment participants failed to consider the durability of the plants when designing the solutions. Although the plants were local and, in some cases, even recycled from other projects in the city, they were not suitable for the environment and many of the plants have already died as a result. The indicator for resource efficiency fails to include this aspect of efficiency, which is why the resource efficiency outcome of the experiment in Antwerp has been, unjustly, classified as high by the indicators designed in this research.

Despite the differences in outputs and outcomes, the activities that led to these outcomes were largely similar. In both cases, implementing a small-scale solution to demonstrate the potential of the project was part of the methodology. In both cases, this demonstration was successful in helping to build awareness and gaining public support for a larger-scale implementation. Similarly, in both cases, the design of the solution was the result of a process of collaboration and social learning, where the stakeholders came to a joint vision for the solution to implement. The differences in scale can be attributed to the different implementation goals. In Burgas, the implementation took place at the beginning of the design of the action plan as a way to test some initial assumptions about the desirability of NBS and methods to involve stakeholders and build awareness. As a result, the implementation was kept rather small. Based on the outcomes of the implementation, the action plan was designed. In Antwerp, on the other hand, the implementation took place at the end of the process to start implementing the action plan on a small scale. At this point, there was a more developed action plan in which different nature-based solutions were designed, which allowed for a more diverse implementation of nature-based solutions. When comparing the differences between the two experiments, one could state that Antwerp has accomplished more and more diverse outputs, whereas the experiment in Burgas achieved better outcomes. Given the differences in scale and goals of the different implementations of nature-based solutions, however, one could argue that it would be unjust to compare the two experiments at all.

When looking at the input that contributed to the physical change, some features can be identified as relevant. In both cases, the choice for nature-based solutions was based on input from experts and scientific reports on the risks and issues in the neighborhoods. In both cases, a similar commitment construction could be observed with a small group of 'thinkers' and 'planners' that designed and planned the implantation, aided by a larger group of less-committed 'doers' that assisted in actually implementing the solution. In both cases, this was a construction that was preferred by the experiment participants, which indicates that, for experiments relating to such a high number of stakeholders, high commitment of all stakeholders is not a prerequisite for success. Lastly, especially in Antwerp, the importance of financial support was emphasized by some challenges with obtaining funding for the implementation.

8.1.5. Actionable knowledge

Burgas			
Input	*Activities*	*Output*	*Outcome*
Expertise from Resilient Europe network	Methodology Collaboration within Resilient Europe network	Highly detailed action plan	Precaution and adaption Democratic governance

	Collaboration with stakeholders for social learning		
Antwerp			
Expertise from Resilient Europe network	Methodology Collaboration within Resilient Europe network Collaboration with stakeholders for social learning	Medium detailed action plan	Precaution and adaption Democratic governance

Table 8.5: Analysis actionable knowledge

In both cases, a detailed action plan was designed, taking into consideration the different challenges the neighborhoods face as well as the wishes and needs of the neighborhood's residents. By making sure the action plan focuses on different, overlapping aspects of climate change resilience, including infrastructural resilience, societal resilience and ecosystem resilience, the action plans represent a highly risk-averse approach when compared to the traditional focus on just one of the aspects. Additionally, the indicators included in the action plan should encourage adequate monitoring of the progress and success of the implementation and should lead to timely adaption when needed. Lastly, the action plans serve as a tool for committing the different stakeholders to the consensus-based visions, both in the short and long-term.

For both cities, this design approach was the result of the influence of the Resilient Europe network. Within the Resilient Europe program, the cities were asked to apply a transition management approach that helped them design the action plans and transition pathways. They were assisted in doing so by the transition management experts active in the network. The methodology was therefore highly influenced by the participation in the program. The contents of the action plan were also influenced by the program, as especially the civil servants in Burgas indicated that they adopted many of the approaches they observed in other cities in their own action plan, such as the method of direct contact. The collaboration with the different stakeholders, however, had an even larger influence on the contents of the action plan, as the broadly-shared visions on the actions and pathways were the results of social learning. These contents appear to have been influenced by the stakeholder constellations. In Burgas, many stakeholders dealing with flood risks on a professional basis, such as civil servants, the Red Cross and the fire department, were involved in the design of the action plan, whereas the citizens played a relatively small role. In Antwerp, on the other hand, the citizens designed the action plan, with the civil servants and professionals only providing information and assistance. When comparing the action plans, one can observe that the Burgas action plan is much more focused on detailed actions and indicators to mitigate flood risks, whereas the Antwerp action plan leaves much more room for interpretation and focuses more on the living environment as a whole. Aside from this possibly being the result of the actors and actions that were involved in designing these action plans, the differences in flood risks between the two neighborhoods might also have influenced the focus and detail of the action plans. As the flood risk in Burgas is much higher than in Antwerp, the need for a highly detailed a specific risk mitigation plan is also much higher than in Antwerp.

It should be noted that, although the output from both experiments have been classified as high, the output from the Burgas experiment is much more detailed and task-focused than the output from the experiment in Antwerp. The indicator currently used to appraise the output fails to take into consideration these different levels of detail, which hampers comparison between the different outputs. Nevertheless, one can question if a high level of detail is desirable in all cases or if variation is allowed, depending on the cases' context and goals.

8.1.6. Transferability

Burgas			
Input	*Activities*	*Output*	*Outcome*
Commitment	Collaboration Reflexivity and learning	Medium: lessons learned, but not validated	Socio-ecological integrity Socio-ecological stewardship and gov.
Antwerp			
Commitment	Collaboration Reflexivity and learning	High: lessons learned and validated	Socio-ecological integrity Socio-ecological stewardship and gov.

Table 8.6: Analysis transferability

The civil servants in Burgas drew a number of lessons for transferring from the experiment. Their best practices included personal contact with citizens to engage them and build trust, the use of small-scale experiments to be able to quickly demonstrate results and the integration of the different aspects of resilience in future action plans to reduce the risk of failure. Additionally, the civil servants identified two possible challenges. The first challenge could be engaging citizens when their connection to the experiment is less obvious than in Dolno Ezerovo. The second challenge could be the high intensity of the process, which requires more manpower than the municipality can provide. Although these lessons have not yet been validated through research or application elsewhere, the civil servants indicated they were highly motivated to repeat such experiments within their municipality.

The civil servants in Antwerp also identified a number of lessons for transferring. The first lesson is to approach citizens with an open agenda and allow them to 'dream' about solutions that they would like to see, as opposed to asking citizens to comment on existing plans. This helps build support for the plan as well as provides fresh perspectives. The second lesson is the organize highly public "do days" where a large group of citizens can become involved in the experiment and build awareness and problem ownership. These lessons have already been validated by successfully applying them elsewhere. Similar to Burgas, the civil servants in Antwerp recognized that different stakeholder constellations and the high intensity of the collaboration process might be barriers for transferring the experiment.

Most of the lessons learned in both cases were the result of the process of collaboration, or 'learning-by-doing', or forecasting activities. In both cases, it was beneficial that the civil servants were committed throughout the process to observe and extract the different lessons. In Antwerp, specific learning questions were formulated and linked to different learning mechanisms, which helped to focus the experiment on the lessons that were most important to them. In Burgas, a more open approach towards learning was adopted, which still led to the deduction of a number of lessons.

When analyzing the indicator used for the appraisal of transferability, one could argue that the question *if* validated lessons for transferring have been formulated contains very little information. Rather, asking *which* lessons have been formulated provides the information that can be interesting for the governance of experiments.

Burgas			
Input	*Activities*	*Output*	*Outcome*
Commitment	Collaboration Reflexivity and learning	Medium: lessons learned, but not validated	Socio-ecological integrity Socio-ecological stewardship and democratic governance
Antwerp			
Commitment	Collaboration Reflexivity and learning	Medium: lessons learned, but not validated	Socio-ecological integrity Socio-ecological stewardship and democratic governance

Table 8.7: Analysis scalability

Similar to the lessons for transferability, in both cases lessons for scalability have been defined. They either focus on scaling out, horizontal scaling, or scaling up, vertical scaling. The civil servants in Burgas indicated they aim to repeat the experiment, arguing that the approach will still be applicable when scaling out step-by-step. The intensity of the process, having to involve more and different stakeholders, was a point of concern, however.

The civil servants in Antwerp also identified the repeating of the experiment as the best practice for scaling out. In addition, the participants of the experiment in Antwerp learned that, although the responsibility for the scaling out has been largely transferred to the citizens, maintaining the current coalition of stakeholders was vital in actually achieving change, due to the complementary skills. Citizens, for example, are better positioned to engage local stakeholders and design tailored solutions, whereas civil servants are in a better position to understand the nuances of the institutional environment. Two issues with scaling out or up were identified, the first of which relates to the maintenance of the nature-based solutions, which is currently largely the responsibility of the citizens. However, a larger responsibility might become a barrier for commitment and participation. Secondly, local regulations are likely to become an issue if the nature-based solutions transfer from small-scale experiments to large-scale changes in the public domain. The lessons have not yet been validated, however.

Similar input and activity features could be identified when comparing the two cases. The opportunities and challenges identified in the previous subparagraphs, however, are not necessarily the result of lessons learned from the experiment, but rather from forecasting activities of the possible impact of implementing the lessons that were learned from the experiment. One could argue that the features transferability and scalability are both lessons learned from the experiment, only with a slightly different focus. In line with that statement, one can also conclude that for the feature of scalability, the question *which* lessons were learned is more important than the evaluation *if* any lessons were learned.

Burgas			
Input	*Activities*	*Output*	*Outcome*
Awareness	Collaboration leads to learning Reflexivity and learning	2 Consequences accounted for	Precaution and adaption
Antwerp			
Awareness	Collaboration leads to learning Reflexivity and learning	2 Consequences accounted for	Precaution and adaption Resource efficiency

Table 8.8: Analysis accounting for consequences

In addition to the lessons formulated for the transferring and upscaling of the experiments, some unintended consequences from these processes of transferring or upscaling could be identified. In Burgas, future maintenance and offsetting sustainability gains through littering have been identified as possible unintended consequences. In both cases, measures have been defined to mitigate any negative impact from these two consequences. In Antwerp, maintenance has also been identified as a challenge, but no clear-cut answer to this issue has been found yet. The process councilor and the civil servants are aware of the challenge, however, and are considering different solutions. The second consequence identified in Antwerp is the offsetting of sustainability gains by using tap water to water the nature-based solutions. Alternatives focusing on rainwater use have been included in the action plan to mitigate this consequence.

In both cases, the identification of these consequences was a result of the collaboration process. The consequences were identified either naturally, through the awareness of the participants of the possibility of the consequence occurring, or through special evaluation activities aimed at identifying these consequences. The consequences identified are partially overlapping with the lessons for scalability, which indicates that accounting for consequences is a third aspect of the lessons learned from and throughout the experiment. Here too, the lessons learned are more interesting than assessing if any lessons were learned.

8.2. The governance of transition experiments

Based on the comparison between the outputs and outcomes of the two experiments in Burgas and Antwerp, some observations can be made about the governance of both experiments. In the next paragraph, the similarities and differences between the two cases will be briefly highlighted and their impact on the outcomes will be assessed.

8.2.1. Similarities

Both the experiment in Burgas and the experiment in Antwerp adhered to a number of good governance principles, the most dominant of which are inclusion, transparency and responsiveness. In both cases, the open and collaborative process facilitated the formulation of a shared vision of the pathways towards resilience in the neighborhoods. Stakeholders highly appreciated the opportunity to ensure their interests were represented in the action plan. Additionally, equal access to information and equal opportunities to influence the decision-making increased the stakeholder satisfaction with the process and helped build trust and commitment. Although no other literature on transition experiments has been published yet to contradict or support this claim, the relationship between good governance and factors such as trust and satisfaction has been supported by literature on other topics (e.g. Bouckaert & van de Walle, 2003; Ott, 2011; Yousaf, Ihsan, & Ellahi, 2016).

Another dominant feature that could be observed is the importance of "learning-by-doing". In both cases, putting theory to practice helped the civil servants gain new skills, capacities and confidence. Similarly, by discussing the elements and goals of neighborhood resilience, the experiment participants reached a consensus on the vision for the neighborhood. This is a form of social learning that emerged as a result of actively working with the topic of resilience. Additionally, by sharing their lessons with other civil servants within the Resilient Europe network and actively contributing to these other experiments, civil servants were forced to reflect on their own processes. These activities therefore encouraged learning about the process of collaboration itself. Again, although no research has been conducted on this phenomenon occurring within transition experiments, the observations seem to align with the so-called Experiential Learning Theory, which emphasizes the importance of learning-by-doing (Buri et al., 2012). Although experiential learning has been researched in many different fields, the focus on the impact of experiential learning on civic responsibility is especially interesting

for transition experiments (e.g. Billig, 2000; Hatcher, 1996). The presumed relationship between learning-by-doing and building accountability can be highly relevant for the future design of transition experiments.

It should be noted that *equal opportunities to influence the decision-making* mentioned earlier does not imply that all stakeholders need to be involved in this process of decision-making. For both experiments, a small core group of stakeholder representatives was responsible for the majority of the decisions made, with a larger group of stakeholders being given the opportunity to provide input on these decisions at highly accessible events. By organizing the experiments this way, the process remained adaptive and manageable, while ensuring responsiveness to the population's needs. This led to quick results as well as broad public support for the action plans. One should be careful however that long-term commitment is considered to be a prerequisite for higher order learning in transition experiments, so the level and type of learning will likely differ depending on the stakeholders' commitment to the process (Antikainen, Alhola, & Jääskeläinen, 2017).

The action plans themselves serve as governance tools for future collaboration. They provide clear actions to be taken to increase neighborhood resilience, the ways these actions are to be assessed and the goals these actions should serve. In doing so, they serve as an informal agreement between the stakeholders involved in the experiment to commit to the shared vision in both the short and long-term.

A stakeholder that plays an important role in the experiment is the local mediator. This mediator serves as a neutral party between the citizens on the one hand and the civil servants on the other hand and can use her local knowledge to identify challenges and opportunities for the experiment. In both cases, such a local mediator was active and proved to be instrumental in engaging and supporting citizens throughout the different phases of the experiment.

8.2.2. Differences

The main difference between the two cases that kept reoccurring when comparing the two experiments, is the difference between a top-down and a bottom-up approach. Emphasizing the importance of learning-by-doing, most of the learning in a top-down focused experiment took place at the level of the civil servants. In the bottom-up focused experiment, on the other hands, citizens were given the opportunity to become actively engaged in the experiment, thus engaging in learning activities. Additionally, because the citizens played such a big role in the design and first implementation of the action plan, they reported to feel a strong sense of accountability for the continued implementation of what they considered to be *their* action plan. The downside of such a bottom-up focus is the question of legal accountability, however. For a top-down approach, the legal accountability for the outcomes remains with the municipality. For a bottom-up approach, on the other hand, the boundaries become blurry. New, clearly-defined rules for sharing the legal responsibility between the municipality and the citizens will have to be designed in such cases.

In addition, an externality can be observed when comparing the citizen empowerment. In Burgas, the citizens involved in the experiment were not empowered or inspired to take individual actions to protect themselves and their neighborhood from flood risks. In Antwerp, on the other hand, both citizens directly involved in the experiment and citizens that were not directly involved in the experiment have started to take individual actions. Remarkably, the residents in Burgas have had more direct contact with the civil servants and should be better informed about the importance of neighborhood resilience. The citizens in Antwerp, however, had the opportunity to observe their neighbors take collective and individual actions to increase the resilience in the neighborhood. It can be hypothesized that awareness of possible individual actions to take is not enough to actually empower citizens, but that copycat behavior can be a driver for change. Psychologists refer to this phenomenon as social proof: people are more

likely to change their perception or behavior if they observe others in their network doing so (e.g. Cialdini, 2009; Contractor & DeChurch, 2014). Empowering local citizens can therefore be an important goal if one wants to impacts the experiment 'spillover' into the wider neighborhood.

Another difference between the two cases is the difference in stakeholder inclusion. Where the experiment in Burgas succeeded in involving less privileged stakeholder groups, the experiment in Antwerp did not. It can be argued that the difference between a top-down and a bottom-up approach could have contributed to this difference. A bottom-up approach requires a high level of commitment from the citizens, which might act as a barrier for participation. An observation should be made, however, that, despite the exclusion of less privileged stakeholder groups in the experiment in Antwerp, efforts were made to ensure the experiment was responsive to their needs.

The two cases represented two highly different approaches when it came to stakeholder inclusion in the design of the action plan. In Burgas, a large number of experts were included in the design of the action plan, which resulted in a highly detailed, flood risk-mitigating action plan. In Antwerp, the citizens were given the opportunity to design the action plan, which led to an action plan that was less detailed and more focused on holistic improvement of the living environment. Depending on the gravity of the flood risks to be mitigated, either approach could yield desirable results.

Furthermore, differences could be observed between the input features. The most dominant of which is the difference in social capital between the two cases. Where the citizens in Antwerp had a lot of experience with and knowledge about bottom-up experiments, the citizens in Burgas had none. This raises the question to what degree the outcomes of the experiments can be contributed to the design and governance of the experiments and to what degree the outcomes can be contributed to the context.

Some remarkable differences between the input features and their impact on the experiment could furthermore be observed. Where the lack of trust posed a barrier for collaboration in Burgas, it acted as a driver in Antwerp. Similarly, the citizens in Burgas felt encouraged to participate in the experiment by the high level of institutional alignment, whereas the institutional misalignment in Antwerp achieved the same goal. It could also be observed that a high level of awareness of flood risks could act as a driver for participation in the experiment, but that a lack of awareness does not need to be a barrier for participation, as long as the experiment can be linked to other issues that the neighborhood's residents are aware of.

8.2.3. Intermediate reflection on the governance of transition experiments
When analyzing the experiment outcomes relating to the governance of transition experiments, a few observations can be made. Firstly, a number of the similarities and differences observed all lead to a similar conclusion: the more hands-on experience the experiment participants gained, the more likely they are to continue with either the experiment or the broader transition towards a more resilient water system. In the experiments, a sense of responsibility, a new understanding of the importance of resilient systems and experience with transition experiments all appear to contribute to the confidence and will to contribute to a transition. These elements of social change can be enhanced by a collaborative, responsive and transparent process of experimentation.

However, the differences observed between the two cases highlight the difficulties to make generalized statements about different transition experiments. Firstly, a number of trade-offs could be observed between the different experiments, such as a top-down or a bottom-up approach and the implementation of physical change at the start or the end of the process of collaboration. These trade-offs inevitably lead to different outcomes that, given the context of the experiments, can be equally suitable for the experiments. Furthermore, similar features

could be observed to lead to different outcomes in the two experiments, which raises the question how large the influence of the context is on the experiment outcomes. Both findings seem to indicate that standardized evaluation of different experiments and their success is not advisable, given the large differences these experiments represent.

8.3. The evaluation of transition experiments
8.3.1. Observations
Aside from studying the different mechanisms, processes and institutions and their role in the experiment, the case studies served a second goal of validating the evaluation framework and its operationalization. Based on the cross-case analysis, some statements regarding the validity of the framework can be made.

Firstly, three input features were added to the framework. The input feature of social capital turned out to be a highly relevant feature in both the cases, with local knowledge, networks and mediators playing instrumental roles in the experiments. The feature of institutional alignment, on the other hand, also proved to be a relevant input feature. Although the impact differed, in both cases, a relation between the support from the municipality and stakeholder commitment and satisfaction could be observed. Additionally, in Antwerp, it could be observed that local laws and regulations can threaten the implementation and upscaling of the experiment. The third feature, leadership, could also be observed in both cases, with the leadership in Antwerp even transferring from the civil servants to the citizens. Together with the local mediator, a leader or leadership entity responsible for coordinating and facilitating was assessed by the stakeholders as an important prerequisite for a successful process.

Some positive observations about the framework can be made. Firstly, it can be concluded that the framework is sensitive to many different experiment outcomes, including some that are often ignored in policy evaluations. Outcomes such as trust, built capacities and changed perceptions are all hard to measure quantitatively, but can have a significant impact on the transition process. This was emphasized by the civil servants that participated in the experiment, with one of the civil servants in Antwerp stating: "Empowerment, local knowledge and built capacities are all impressive results. But politicians have a tendency to focus on the tangible results. The question is: how can we make these results more visible?". By using this framework for the evaluation of transition experiments, these less visible results are also included in the evaluation. Additionally, the holistic nature of the framework helps to structure an abundance of information and provides a research focus that takes into consideration the many possible relations, including some less obvious ones.

This is a departure from the way flood risk management has previously been evaluated and studied. Examples of such frameworks include the analysis of the actors and their roles in the experiment (e.g. Doorn, 2016), the analysis of highly general and isolated concepts, such as goodness of fit or collaborative design (e.g. Priest et al., 2016; van Herk et al., 2011), each of which tends to focus on just a narrow aspect of FRM. Some authors have started to design and apply more holistic frameworks, aiming to expose the relations between different aspects, but these are often still focused on a narrow aspect of FRM, such as legitimacy or resilience (e.g. Alexander et al., 2016; H. Mees, Crabbé, & Driessen, 2017). The framework adapted in this research can provide a valuable tool for the comprehensive evaluation of FRM experiments as it not only exposes features of the experiment and relationships between these features that were not considered in previous evaluation methods, but also provides a format for the comparison of the findings from different experiments.

However, the application of the framework on two real-world cases also exposed a number of weaknesses. The first weak spot of the framework is the operationalization. As highlighted in Chapter 7, a large number of times additional relevant information could be observed within the cases that had not been properly included in the indicators. Examples

include, but are not limited to, the failure to include local mediators and local knowledge in the indicator for social capital and the actionable knowledge feature's inability to assess the level of detail of the actionable knowledge generated by the experiment. Although this could be the result of poorly designed indicators, the literature study conducted in chapter 3 already showed that most of the features likely consist of a diverse range of variables. As this is validated by the range of variables observed in the cases, it seems more likely that a small number of indicators simply does not offer enough variation to cover all possibly relevant aspects of the experiment. Given the already high information load, it is reasonable to assume that a framework including indicators for all variables will become unmanageable for any researcher not familiar with the experiment. One could furthermore pose the question if the costs of such an extensive evaluation are proportional to the size of the, often rather small, transition experiments.

Additionally, the evaluation of the experiments took place based on ex-post interviews with a few representative stakeholders. In order to conduct a truly reliable evaluation, especially considering outputs dealing with stakeholder satisfaction or social change, one would need to gain input from a much larger sample of the population. Furthermore, the changes could be evaluated even more thoroughly if a baseline study was performed ex ante in order to be able to quantitatively and qualitatively assess the differences before and after the experiment. Again, this would further increase the evaluator's workload.

Next, one could observe some differences in the way the features could be used. The features included in the framework appear to be of different quality. In the activities category, for example, two actual activities, collaboration and learning, are included, whereas the other three features, sound methodology, meaningful sequence of actions and transparency appear to be requirements for the activities rather than activities on their own. Similarly, built capacities, accountability, social change and physical change in the output category can all be compared and appraise quite straightforwardly. Actionable knowledge, transferability, scalability and accounting for consequences, on the other hand, do not appear to add much information to the evaluation when determining *if* and *to which degree* these features were achieved. Rather, *which lessons* can be learned from these four features appears to add more value to the evaluation.

In addition, some of the indicators, such as equity and democratic governance are as much an outcome as they are a requirement for performing the activities. In a similar fashion, activities lead to outcomes, but also influence input features, making the input features an outcome of the experiment as well. The success of the experiment, however, is only evaluated based on the outcome features and not on the input or activities features. One could argue that the reality is too chaotic to be adequately captured by a logic model framework for evaluation. This is supported by the use of logic models, which have mainly been applied to more straightforward program evaluations (Newcomer et al., 2015). The research findings combined with the traditional use of logic models cast doubts on the assumption that these models can also be used for the evaluation of more complex transition experiments.

Last, one can question if, even in the case of a perfectly operationalized framework, evaluation between different experiments is actually possible. In the cases of Burgas and Antwerp, in many instances, the outcomes were the result of different goals, trade-offs and, arguably, contexts. The differences in physical change, for example, can be attributed to the different moments in the collaboration process the solutions were implemented and the different goals they served. One should furthermore notice that the outcome features themselves are not objective goals to be achieved, but rather value-laden propositions what an experiment should achieve. Although the framework does assist in exposing these different goals, contexts and trade-offs, these are not adequately reflected in the outputs and outcomes when using the framework for the positivist evaluation and comparison of different experiments.

8.3.2. Intermediate reflection on the evaluation of transition experiments

Overall, both the findings from analyzing the governance of the experiments as well as the evaluation itself seem to indicate that the positivist framework for the evaluation and comparison of a number of transition experiments is not fit-for-purpose. Even when comparing experiments that took place within the same policy program, major differences in context and experiment goals could be observed. As the framework is already quite extensive in its current form, adding each of these different possible elements to the framework seems highly unfeasible. Applying a restricted and standardized framework to evaluate and compare these experiments would therefore not only lead to an appraisal that fails to include the specific characteristics of each experiment, but also ignores a wealth of information about the aspects of the experiment that led to the different outcomes. It would, in short, likely increase the risk of 'tick-box' evaluations that give little incentive for in-depth analysis and learning.

Nevertheless, it can be stated that, despite the positivist nature of the framework, in-depth analysis did take place when studying the two cases. This can be attributed to the designed use of the framework, applying a pragmatic approach combining both positivist and interpretivist research. Although the indicators were used for evaluation, they were just a small part of the analysis. A more open analysis, taking into consideration the different possible aspects of each of the features as described in chapter 3, took place in addition to the positivist evaluation. This more interpretivist approach did not focus on what the experiment *should* achieve, but rather on *what* was achieved and *why* this was achieved. For this interpretivist analysis, the framework's general features were highly useful to focus and structure the analysis and comparison, while remaining abstract enough to discourage superficial analysis.

9. Conclusions and discussion

In this chapter, the research questions will be answered, conclusions will be drawn and the findings will be discussed.

9.1. Research questions and framework design

As climate change and rapid urbanization increase the risk of floods in cities all over Europe, increasing the cities' resilience against these threats is becoming a progressively important issue. As traditional measures to mitigate flood risk are unable to accommodate the uncertainty and rapid change inherent to climate change, the water management system will not have to just be adapted, it will need to be *transformed*. Transition experiments can play an important role in this transformation, as they allow for learning about the system, its problems and possible solutions, without creating a lock-in. However, little research has been done on the governance of transition experiments, which means few tools and guidelines are available for the design of such experiments. This research aims to contribute to bridging this knowledge gap by providing a validated framework for the evaluation of the governance of transition experiments. The main research question to be answered in this research has been formulated as follows:

How can the governance of transition experiments be evaluated in relation to the experiment's success, focusing on flood risk management?

A number of sub questions were devised to help answer this main research questions. Each of these sub questions will be answered below. The first three sub questions were aimed at facilitating the design of an evaluation framework. The last sub question was aimed at validating the designed framework. By revisiting the first three sub questions after the validation, the main research question can be answered.
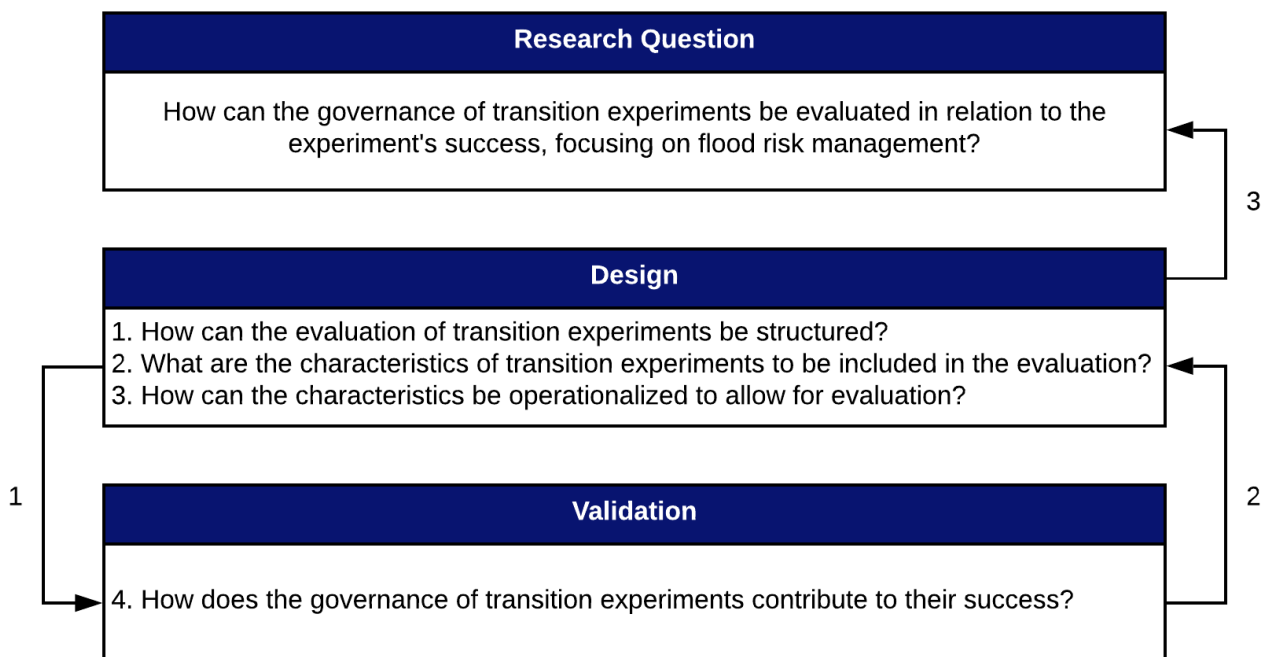

Figure 9.1: Method of answering the main research question

1.1. How can the evaluation of transition experiments be structured?

Transition experiments are experiments aimed at facilitating a transition. By adapting the definition by Van den Bosch and Rotmans (2008, p.13), the following definition of transition experiments has been defined: "Sustainability transition experiments are innovation projects with a societal challenge as a starting point for learning, aimed at contributing to a sustainability transition by employing a multi-stakeholder approach."

As these experiments are often highly complex, consisting of different phases, sub-goals and related sub-governance features, any evaluation activities will have to focus on separating these different aspects first. Logic model frameworks are traditionally used in program evaluations, but can be adapted to fit the research focus on transition experiments. They consist of four categories, inputs, activities, outputs and outcomes, which are used to chronologically structure the evaluation.

| Inputs *What was invested?* | → | Activities *How was it done?* | → | Outputs *What was generated?* | → | Outcomes *What was accomplished?* |

Figure 9.2: Logic model of evaluation

The major strength of the logic model when compared to other evaluation methods is the focus on not just the results, but the relationships between the results and the context and experiment activities. This allows for a more thorough understanding of which outcomes were achieved and why they were achieved. This makes the logic model framework highly useful when evaluating the governance of transition experiments in relation to the experiment's success. Based on the evaluation of a large number of real-world cases, Luederitz et al. (2017) have designed a logic model framework for the evaluation of transition experiments. However, this framework needs to be validated before it can be used on a large scale.

1.2. What are the characteristics of transition experiments to be included in the evaluation?

In order to gain a better understanding of the possible challenges, opportunities and governance of transition experiments in the water system, three concepts related to the system were analysed. Urban Resilience was used to further specify the desired output and outcomes of the transition experiments, whereas Flood Risk Management and Reflexive Governance were used to identify governance mechanisms and goals that could or should be observed within the experiments. In order to increase resilience, a number of outputs were identified as important in the literature on urban resilience. Especially technological and institutional innovation as well as the stakeholders' capacity to self-organize were deemed to be important outputs of the experiment. In terms of governance mechanisms, literature on urban resilience identified intensive collaboration as an important activity, that could lead to learning, not just about the system and the innovation, but also about shared visions and values.

Literature on Flood Risk Management underwrites the importance of collaboration and emphasizes that governance mechanisms should be focused on reaching a stakeholder consensus rather than just stakeholder consultation. In addition, practices of good governance, such as transparency, fairness and legitimacy are identified as relevant governance practices. Here too, the literature states that stakeholder empowerment should be seen as one of the main goals of any process of collaboration.

Lastly, literature on Reflexive Governance again emphasizes the importance of collaboration in all aspects of the experiment. The process of collaboration does not just serve

a specific goal, it is also an opportunity for learning about the process itself. Stakeholder empowerment is the only output of reflexive governance that could be identified.

Overall, it can be concluded that the literature on the three concepts provide a general direction for the governance of transition experiments, that can be summarized as consensus-focused collaboration that is not only aimed at learning about the innovations, but also at stakeholder empowerment and changing perspectives.

1.3. How can the characteristics of transition experiments be operationalized to allow for evaluation?

When comparing the logic model of evaluation as designed by Luederitz et al. (2017) to the findings from the literature review, one can firstly note that many of the evaluation model's features are also highlighted in the literature on the related topics. An additional 3 features are tentatively added to the inputs category: social capital, institutional alignment and leadership.

However, despite the similarities in the features identified, each of the concepts highlighted different aspects of the features that could be observed and evaluated when comparing different transition experiments. Taking into consideration that the logic model already includes 27 features, or main categories, including each of these aspects in the evaluation was thought to lead to an evaluation framework of unmanageable proportions. Instead, each of the aspects was compared to each other and, where possible, combined into more general indicators. Where possible, quantitative indicators where designed. For features with a more qualitative nature, ordinal scales were designed to be able to compare the degree to which an experiment succeeded in realizing a specific activity or goal.

This resulted in a final framework including 31 indicators, capturing the most dominant characteristics as identified in the literature. It was noted, however, that by reducing the number of indicators, much of the detail of the framework was lost. This was ex-ante marked as an important limitation of the framework and the data collection was therefore designed in such a matter that aspects not included in the predefined indicators could be observed as well.

Input	Indicator
Awareness	Stakeholder awareness of flood risks and their own role in FRM
Commitment	Participation throughout the experiment
Trust	Willingness to rely on other participants' judgment and capacities
Social Capital	Number and type of pre-existing networks included in the experiment
Expertise	Number and type of expertise included in the experiment
Support	Number and type of support included in the experiment
Ins. Alignment	Experience of institutional barriers and impact on experiment
Leadership	Effective leadership entity
Activities	
Transparency	Stakeholder satisfaction with transparency
Collaboration	Mechanisms for collaboration Number of big decisions made through collaboration Stakeholder representation and diversity Stakeholder satisfaction with collaboration
Reflexivity	Presence of learning goals and learning mechanisms
Sequence	Presence of careful planning
Methodology	Presence of methods for achieving outcomes
Outputs	
Built capacities	Reported gained or enhanced skills by both citizens and civil servants
Accountability	Stakeholder feeling of responsibility and willingness to become proactive
Social change	Newly formed networks Changed perceptions
Physical change	Number and type of physical changes

Actionable knowledge	Number and type of lessons learned
Transferability	Presence of validated lessons for transferring
Scalability	Presence of validated lessons for upscaling
Consequences	Number and type of consequences accounted for
Outcomes	
Integrity	Further implementation of NBS planned or undertaken by citizens and civil servants
Equity	Planned or current involvement of less privileged stakeholder groups with specific policies designed
Governance	Number and type of new participatory settings
Opportunity	Number and type of economic opportunities
Efficiency	Source of materials used
Precaution	Formulation of flood risk mitigating plan

Table 9.1: Indicators designed ex-ante

9.2. Results

Making use of the framework described in the previous paragraph, two cases have been evaluated. Based on the outcomes of the case studies, a number of observations about the evaluation model can be made. In order to structure this evaluation, the design questions 1.1, 1.2 and 1.3 will be revisited to assess if the answers to these questions still hold up after the empirical application of the framework

2.1. How does the governance of transition experiments contribute to their success?
Some interesting differences and similarities could be observed between the two cases that can likely be attributed to the governance of the experiment. The most remarkable difference in the outcomes of the two experiments is the social change that could be achieved. Where the experiment in Antwerp did achieve tangible change in terms of stakeholder empowerment and institutional innovation, the experiment in Burgas was less successful in doing so. As in both cases the experiment participants emphasized the importance of learning-by-doing, it seems likely that the differences in the governance mechanisms of both experiments contributed to this difference. This is also supported by literature on experiential learning (Billig, 2000; Buri et al., 2012; Hatcher, 1996). The Antwerp experiment placed the citizens at equal footing with the civil servants, coached them on how to conduct such experiments and made them responsible for most of the decisions made during the process of experimentation. This helped build a sense of accountability and helped the citizens gain practical experience. In Burgas, on the other hand, the involvement of the citizens was limited to information-sharing and some, largely symbolic, actions and did not result in this social change.

The importance of taking action can be further emphasized by the observation that citizens in Antwerp that were not directly involved in the experiment did start to take individual actions by observing their neighbours doing it, whereas citizens in Burgas, although better informed about neighbourhood resilience, did not. It can be hypothesized that the social proof phenomenon also plays a role in achieving wider social change in neighbourhood transition experiments (Cialdini, 2009; Contractor & DeChurch, 2014). It can therefore be hypothesized that the governance of transition experiments should not just be focused on stakeholder inclusion. Rather, it should also focus on transferring the responsibility of the design and implementation of the experiment to the less-empowered participants, with the already empowered participants taking on a facilitating role. This can not only help empower experiment participants, but has the potential to spill over into the wider neighbourhood through leading-by-example.

Some of the principles of good governance, specifically transparency, inclusion and responsiveness, could be observed in both cases as boundary conditions that served the goal of increasing stakeholder satisfaction and thereby building trust, commitment and social capital

(Bouckaert & van de Walle, 2003; Ott, 2011; Yousaf et al., 2016). Additionally, vision documents can be useful governance tools for committing stakeholders to a joint vision. Furthermore, they safeguard the implementation of this joint vision on the short and long-term and serve as a reminder of the desired system changes. Mediators with ample knowledge of the local community and a strong local network can be instrumental in designing and facilitating the processes of collaboration and implementation.

Overall, one can carefully conclude that the governance of transition experiments does contribute to the success of such an experiment, as already hypothesized by other scholars (Kivimaa et al., 2017; Weiland et al., 2017). Especially for stakeholder empowerment as well as stakeholder satisfaction, governance can play an important role. This indicates that the governance mechanisms will need to be carefully considered when designing and conducting a transition experiment in order to maximize the experiment's output. As a result, the governance of transition experiments can be interpreted as a fertile topic for further research. This research has provided some initial avenues for further research into the topic.

However, some caution should be observed when interpreting these statements. Firstly, it could also be observed that the governance of experiments is by no means straightforward. More likely, the choices made and the outcomes observed are the result of trade-offs between different experiment goals. An increase of specific features is therefore not always desirable as it could lead to the deterioration of other features. High stakeholder inclusion, for example, will lead to the empowerment of a larger number of stakeholders. However, it can also hamper the manageability and progress of the experiment, which is why in both experiments the choice was made to focus on a small, but manageable core group over a highly inclusive, but unmanageable larger group. The governance best practices as described in the previous paragraphs should therefore not be interpreted as silver bullet solutions. Furthermore, some interesting observations about the input features could be made. Context features, such as trust and institutional alignment, that served as barriers in Burgas, served as drivers for participation and empowerment in Antwerp. The fact that their different impacts on the cases could be clearly observed, raises the question how large their influence is on the experiment outcomes and how they interrelate with these outcomes. Especially for the feature of social capital it can be hypothesized that this feature served as a confounding variable for the causalities observed. Confounding variables are outside influences that have an impact on an observed relationship, in this case the relationship between the governance and experiment outcomes. This causality has been neither confirmed nor rejected by this research and needs to be researched before definitive statements about the relationship between governance practices and the outcomes of a transition experiment can be made.

2.2. How can the evaluation of transition experiments be structured?
The logic model turned out the be useful for conducting the evaluation. First, by assigning the different characteristics of transition experiments to the four categories, a wealth of features can be structured into smaller, more manageable points of focus for the evaluation. Second, the logic model allows for making the assumptions about the relationship between different features more explicit, encouraging the evaluator to open up the hypothetical black box to identify causalities and allowing for a more thorough understanding of the experiment. Third, the inclusion of the input features in the logic model forces the evaluation to consider rival explanations for the causalities observed. As the case studies indicated that input features can have a big impact on the outcomes of the experiment, explicitly focusing on these features and their role in the experiment can help craft or debunk alternative explanations for the research findings. Similarly, the outcome category serves as a reminder to not only consider the direct outputs of the experiment, but also the more indirect outcomes and externalities. This allows for a better appraisal of the, potential, impact of the experiment.

In conclusion, the holistic, yet structured approach of the logic model was useful in analysing the cases and interpreting the outcomes. One could therefore conclude that, in addition to program evaluation, logic models can also be used for the evaluation, and likely the design, of transition experiments. This confirms earlier findings by, for example, Forrest & Wiek (2014) and Alexander, Priest & Mees (2016).

However, Luederitz et al. (2017) already warned that the logic model categories were not to be interpreted as linear, but rather as parallel and interdependent. This was also indicated by the literature on logic models, which tend to focus on programmes over experiments (Newcomer et al., 2015). This was confirmed by the case studies, as input features could be observed as outputs of the activities and activities could also serve as outcome features. The assumption that the output and outcome features can serve as indicators for the success of an experiment therefore is too simplistic and does not hold when tested against the less straightforward reality of transition experiments. When applying a logic model for the evaluation of transition experiment, all categories should therefore be included in the appraisal of the success of the experiment.

2.3. What are the characteristics of transition experiments to be included in the evaluation?

The literature review unveiled three new characteristics, social capital, institutional alignment and leadership, that were tentatively added to the logic model. Each of the three features could not only be observed in the two experiments, but turned out to have an impact on the experiment. It is therefore proposed that these features are added to the evaluative scheme when using the scheme for the evaluation of transition experiments aimed at increasing climate change resilience. The inclusion of less tangible features, such as built capacities and social change was much appreciated by the experiment participants, as these features are often excluded from experiment evaluations. Apart from livelihood sufficiency and opportunity, all features of the logic model could be observed in the two cases. Although for some features, their role in the experiment did not appear to be as impactful as others, it is likely that this is the result of the limited selection of experiments the framework was to be tested on. For now, no clear indicators that any of the features will have to be removed from the framework can be observed.

One can therefore conclude that the features included in the adapted evaluative scheme are indeed relevant when analyzing or evaluating different transition experiments. This indicates that these features can serve as good starting point when designing such an experiment ex ante or designing an analysis ex post.

However, it should be noted that the features currently included in the framework are not all of the same quality. Some differences in their role in the framework and the appraisal of the outcomes could be observed. These different roles and qualities make it highly challenging to determine how to evaluate and appraise these different characteristics. Additionally, one can argue that social capital and institutional alignment are especially relevant for experiments in the public-private sphere with a large number of possible stakeholders. It can be questioned if they will be equally relevant for experiments with a smaller number of stakeholders or a focus on the private sphere. This is something that will have to be confirmed or disproven by further research.

2.4. How can the characteristics of transition experiments be operationalized to allow for evaluation?

Application of the evaluative scheme to two real world cases exposed a serious, but already anticipated, weakness in the framework. The indicators designed for the evaluation, often limited to one or two per indicator to minimize the size of the framework, did not allow for a

proper evaluation of the experiments. On many occasions, the indicators appeared to be too narrow to properly reflect all different aspects and nuances of a feature. Even based on two relatively small case studies, multiple important aspects could be observed per feature. This finding is supported by the outcomes of the literature review, in which similar extensive lists of aspects could be identified.

For future research making use of the evaluative scheme, one could consider including each of these aspects in the indicators. However, evaluating the cases making use of the already quite extensive evaluative scheme showed that preparing and conducting such an extensive evaluation, especially as an external researcher, requires a lot of time and effort. Similarly, in order to properly measure features that represent change, one should include baseline studies and make sure interview sample sizes adequately reflect the population. This only further increases the researcher's workload. A worry can be voiced that expanding the number of indicators even further would lead to an evaluative scheme of almost monstrous proportions. Furthermore, one can notice that any indicators designed for the different features are highly normative and depend on a researcher's perspective on what such an experiment *should* achieve. Even when making use of commonly accepted literature on the experiment, reducing the impact of the researcher herself, one can argue that these indicators do not necessarily represent the norms and values of the experiment participants. This might lead to a disproportionate focus on features and indicators that were not relevant for the experiment and a neglect of features or indicators that were. It can also discourage evaluators from analyzing aspects of the element that are not included in the 'tick-box' evaluation framework, which could possibly lead to the exclusion of interesting research findings and can hamper learning.

This leads to the most important finding for answering the main research question. Based on this research, it appears that it would be unwise to use the evaluation framework for positivist research when evaluating a number of transition experiments ex post. Such experiments appear to be too complex to adequately capture in a manageable evaluative scheme. One could argue that the indicators used in the research were poorly designed which led to these findings. However, the vast amount of literature summarized in chapter 3 supports the statement that different aspects of each of the features can be relevant, depending on the experiment's goal and context. It should also be noted that a highly-operationalized framework could fail to identify new and interesting aspects and interactions, as the operationalization is often based on existing literature and therefore tends to exclude the unknown. Based on these research findings, one can question if exploration of the positivist use of the framework for ex post evaluation provides a fruitful avenue for further research.

3. How can the governance of an experiment be evaluated in relation to the experiment's success, focusing on flood risk management?

Overall, the logic model has been identified as a useful framework for structuring the evaluation. The characteristics included in the framework by Luederitz et al. (2017) can be recognized as features that are all likely to play a role in the success of transition experiments, with the addition of three new features: social capital, institutional alignment and leadership. However, operationalizing the framework and applying it to real-world cases exposed some shortcomings of the scheme. When comparing the answers to all three design questions, there appears to be some friction between the positivist nature of the way the framework has been used and the less straightforward nature of the real-world cases. Issues with how to determine what should be included in the definition of success, how to deal with features of different quality and how to operationalize the framework in a way that reflects the nature of the cases emerged. Furthermore, concerns were expressed about the manageability of an even more extensive framework and the time and effort a properly designed evaluation would require. However, applying both a positivist and an interpretivist approach in this research has provided

some insights into the advantages and drawbacks of both approaches. Based on these insights, a manual can be designed detailing when it is appropriate to apply which approach.

In the previous paragraph, it has been concluded that applying a positivist approach when aiming to gain in-depth insights into the governance of transition experiments appears to be highly ineffective. That is not to say, however, that a positivist evaluation of transition experiments making use of the evaluative scheme is infeasible by default. It can be proposed that the framework can be used ex ante to guide the design of the transition experiment and to design and prioritize indicators that are to be used for the, internal, ex post evaluation. Doing so would increase the applicability of the indicators, especially if the features in all four categories are considered in the experiment's definition of success. Additionally, the application of the framework by the experiment organizers themselves would also remove a number of barriers an external evaluator encounters, such as gaining access to information and experiment participants, which would increase the manageability of the evaluation. If the indicators of the evaluative scheme are designed ex ante and with the specific experiment goal and context in mind, it seems highly likely that the scheme can indeed be used for a positivist evaluation. The framework would help to appraise if and to what degree different predefined goals have been met. In this case, the appropriate research philosophy for applying the framework would be more towards the highly structured, positivist end of the spectrum (figure 9.3). It should be noted, however, that such a structured form of evaluation likely leaves limited opportunities for recognizing and analyzing unexpected outcomes and behavior. Predefined indicators also make the framework less applicable to a larger number of experiments, due to the contextual differences.

Figure 9.3: Recommended research philosophy for ex ante experiment and evaluation design

Using the framework in a positivist manner for an ex post analysis of multiple transition experiments to gain an in-depth insight into the governance of transition experiments, however, would not be advised based on the outcomes of this research. Alternatively, it is proposed to employ an interpretivist approach when applying the framework. Instead of treating the framework's features as elements that need to be present in an experiment to a certain degree, the features can be used to guide and structure the analysis of an experiment in a more open-ended fashion. This would allow for the comparison of different elements and their relationships, while accommodating the different experiment focuses and goals. As already described in the answer to question 2.3, further research in such an open and explorative manner could help confirm the relationships found in this research or identify new relationships. This would help add to the existing body of literature on the governance of transition experiments. Furthermore, based on the cumulative findings of further research, adjustments to the framework, for example by grouping and prioritizing features and relationships, can be made. In such cases, the appropriate research philosophy for applying the framework shifts towards the open, interpretivist, end of the spectrum (figure 9.4). A drawback of this approach is that it allows for learning and analysis but, due to the lack of direction from indicators, is less suitable for evaluation and comparison.

Figure 9.4: Recommended research philosophy for ex post analysis

The recommended research approaches apply to two highly different situations: evaluating experiment goals and generating knowledge. There is, however, a third possible use of the framework that combines both approaches: the use for program evaluation. Both cases analyzed in this research were part of a policy program. Within these policy programs, there is often a need to asses if the participants met the agreed requirements as well as a need for policy learning for further implementation or future programs. For those cases, a pragmatist approach adapted from the approach applied in this research can be recommended. The experiment goals or priorities can be agreed upon ex ante by the program manager and participants. General indicators applicable to all cases can be designed based on these goals and priorities. These general measurements of success should serve as an invitation to reflect on the observed differences between the experiments included in the program and the causes of these differences. However, this approach does have a drawback. As it contains neither detailed indicators nor completely open features, it is tailor-made for appraisal nor learning. It is therefore unlikely to lead to straightforward answers if requirements have been met or which in-depth lessons have been learned. It is, however, the most pragmatic approach to combining both program evaluation goals. On the research philosophy spectrum, this approach can be placed in the middle of the spectrum (see figure 9.5).
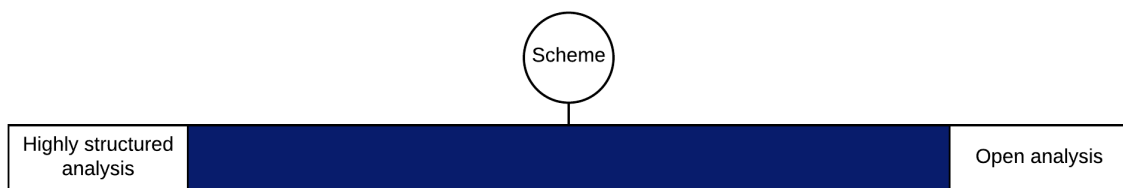
Figure 9.5. Recommended research philosophy for ex post evaluation

9.3. Discussion
In the previous paragraph, the individual results have already been discussed and possible knowledge gaps and alternative explanations have been described and, where possible, refuted. In this paragraph, a more general discussion of the findings will take place.

9.3.1. Research Methods and Design
A first research element to be discussed is the design of the research and the research methods used. For the operationalization of the evaluative scheme, a literature review has been performed. Given the various aspects of the object of research, three main fields of literature were included in the review. Because of this relatively large research scope, a limited number of papers could be included on each topic. This means that the literature review findings only provide a general insight into the different fields of literature and relevant details could have been excluded from the review. This, in turn, could have negatively impacted the operationalization of the evaluative scheme, which could have led to the scheme's poor performance when evaluating the experiments. However, it can also be noted that even such a

limited literature review already led to a large number of aspects that could not be included in the operationalization. This was already marked as an issue during the operationalization of the framework and measures were taken to allow for analysis of aspects of the cases that were excluded from the indicators. The outcomes confirmed these initial concerns by showing that a large number of aspects not included in the framework's indicators could be identified. It seems therefore unlikely that the, relatively superficial, literature review has played an important role in the inadequate operationalization of the evaluative scheme.

The indicators designed can also be discussed. Because of the limited access to highly detailed information, a three-point Likert scale was designed for a number of indicators. Although this did allow for some variations in the ratings of the cases, it should also be noted that mid-point scales often encourage a neutral, or mid-point, ratings (Nadler, Weston, & Voyles, 2015). Looking at the ratings given to both cases, slightly over half of the ratings of the Burgas experiment were medium, whereas slightly less than 20 percent of the ratings in Antwerp were medium. Combining these statistics, one can observe that, in total, only about a third of the ratings given was medium, which makes it unlikely that there was a bias towards the mid-point answer in this research. Nevertheless, designing a four-point scale for future research could minimize the risk of biased ratings. One can comment, however, on the lack of detail in the three-point scales. Most of the scales consisted of composed answers, which gave some insights in if an answer could be observed, but limited insights into the degree to which they can be observed. In this research, this drawback could be mitigated by the more interpretivist approach that was applied in addition to the positivist approach, but this is something that should be considered when designing indicators for future, positivist evaluations.

A third element that can be discussed is the way the case studies have been designed and conducted. The cases were selected to represent both commonalities and major differences. Overall, one can conclude that the cases did indeed represent some variety, especially given the context of both cases. As for the similarities, it turned out that the cases, despite their apparent commonalities, were rather different in terms of goals and experiment focus. These differences created a challenge for comparing and interpreting the outcomes. Ideally, they would have been more similar in order to increase the external validity of the results (Yin, 2014). The differences did, however, expose a clear issue with the generalizability of the operationalized evaluative scheme. Additionally, the small number of cases selected implies that it is challenging to generalize the research findings. It is highly likely that repeating this research with different and more cases will yield different outcomes. The research findings should therefore not be interpreted as generalized truths, but rather as possible directions for further research.

In addition to the cases being more different than expected, the accessibility of the information turned out the be a challenge as well. Although the contact person within the Resilient Europe network was instrumental in providing contacts and informative documents, especially the citizens in Burgas were difficult to reach. As a result, brief surveys had to be used to gain insights on their perspectives, which greatly reduced the amount of available information when compared to interviews. Furthermore, for the experiment in Burgas, a language barrier existed. The surveys had to be translated to and from Bulgarian by a third party, the interviewees sometimes had difficulties expressing themselves in English and a large number of policy documents could not be translated in time. When compared to the case study in Antwerp, in which all stakeholders were native Dutch speakers, the case study in Burgas yielded more generic and significantly less information. As a result, interesting aspects or nuances of the experiment in Burgas might have been overlooked.

The interview protocol, however, turned out to be highly useful in guiding the data collection. By using different information sources, construct validity could be safeguarded. The interview protocols helped structure the interviews and made sure the questions asked were

non-leading and clear. However, it should be noted that, due to the interviews often being conducted on the same day, there was no time to improve the interview protocols based on experiences with previous interviews. Subsequently, less clear or useful interview questions could not be adjusted to make the interviews more effective in gathering information, which could have led to blind spots in the cases. The case study protocol and case database, however, do ensure that the material supporting the claims made can be accessed by those interested and the research can be repeated if needed. This increases the reliability of the findings.

Furthermore, one should note that these types of evaluations are especially sensitive to social desirability bias. The experiments' successes might therefore have been over-evaluated, whereas the experiments' failures might have been under-evaluated or even left out. Although the neutral phrasing of the interview questions is thought to have had a mitigating impact on this bias, it cannot be excluded altogether. This too has led to the very cautious interpretation of the research findings.

A last element of the research methods and design that should be discussed is the choice to include all features of the framework in the research. As already discussed extensively before, this led to the use of highly generic indicators that were found to fail to adequately capture the experiment. Furthermore, one can state that, even when limiting the number of indicators, using the framework for the evaluation of the experiments still required a lot of information. Given the limited time and resources available for this research, this led to results that are still rather general. It did not allow for focusing on possible interesting topics, nor did it allow for a thorough analysis of any identified relationships. One could state that the extensiveness of the evaluative scheme was actually a constraining factor as much as it was an enabling factor in this research. On the other hand, a case could also be made that such a superficial, yet comprehensive analysis is a necessary starting point for any developing field of research. Overall, one can conclude that, despite the lack of detail, the research did yield some interesting insights and possible directions for further research.

9.3.2. Research findings
The case study results are a first attempt at analyzing the governance of transition experiments in such a comprehensive manner. As a result of these pioneering efforts, limited literature is available to compare the results to. Nevertheless, theories and findings from different fields of research could be linked to some of the outcomes of the case studies. For the empowerment of stakeholders and the diffusion of innovation, the importance of learning-by-doing and social proof has already been extensively researched (e.g. Billig, 2000; Buri et al., 2012). Similarly, the importance of good governance, in this case specifically transparency, responsiveness and inclusion has also been supported by a large body of literature (e.g. Ott, 2011; Yousaf et al., 2016). Furthermore, the replication of patterns and outcomes in both cases suggests some level of external validity that should allow for the generalization of the findings. One should emphasize, however, that the differences between the context of the findings in the literature and this research indicate that this is by no means definitive proof of a relation between the literature and the research outcomes. Due to the small number of cases, the observed replicated patterns could be a result of chance rather than causal relationships. More research into both the relationship between the experiments and the different bodies of literature is therefore required, as well as an increased number of case studies researching the role of governance in transition experiments.

Despite these limitations, the tentative conclusion drawn from this research is that governance features do indeed play a role in the outcomes of a transition experiment, especially influencing the social change resulting from participating in the experiment. Two hypotheses can be posed based on the findings that would warrant future research:

H1: *The more responsibility is transferred to the experiment participants, the larger the social change that can be achieved.*

H2: *The better the principles of transparency, responsiveness and inclusion safeguarded in the experiment, the larger the social change that can be achieved.*

By researching these hypotheses, the 'black box' of transition experiments and their success can be unpacked further. Both hypotheses, if proven, could influence the design and management of transition experiment and could support and accelerate transitions to more resilient water systems.

However, some caution should be observed when interpreting these hypotheses. For the first hypothesis, it should be emphasized that governance is predominantly a concept developed and applied in developed countries with a culture of bottom-up initiatives and more egalitarian power arrangements (Haque et al., 2017). One can therefore question if this hypothesis still holds for more authoritative, often developing, societies. When testing this hypothesis, it is therefore important to take into consideration the dominant cultural and political aspects that could influence the outcomes of a transition experiment. This could possibly lead to insights into the need for different governance approaches depending on the context in which the transition experiment takes place.

The second hypothesis currently only includes the three principles of good governance that were observed in this research. For further research, it can be interesting to not just test this hypothesis, but add more principles of good governance to this hypothesis to confirm if these have any impact on the outcomes of the transition experiment as well. These findings could help identify which governance practices should be included in the experiment as well as help prioritize the features of the evaluative scheme.

In addition to these hypotheses that warrant further research, some interesting gaps in the findings could be observed. Firstly, the influence of input features, but specifically social capital, on the outcomes of the experiment should be researched to confirm or oppose the hypothesis posed in this research that it could serve as a confounding variable in the observed relationship between the governance and the outcomes of the experiment. Secondly, it could be observed that similar input features had completely different impacts on the experiments and the other way around. In order to understand the workings of the experiment and add to the body of literature, further research into why these differences occur is therefore desired.

As for the evaluative scheme, both the theoretical and empirical findings strongly indicate that a positivist use of the framework for ex-post analysis should be discouraged. Instead, it can be recommended to use the framework in an interpretivist manner when applying it to analyze different transition experiments. The positivist use of the framework for the ex-ante design of a transition experiment and its evaluation, however, has not been ruled out as a possible use of the evaluative scheme based on this research and could be investigated further. The features of the framework are highlighted as relevant by different fields of literature relating to transition experiments. The outcomes of both case studies also showed that the majority of the features could be observed in the real-world experiments. Despite this external validity, the features' positions within the framework should not be interpreted as fixed. Further research using the framework should be able to expose patterns regarding the importance or lack of importance of the different features.

9.4. Reflection
9.4.1 Academic reflection
In this research, the use of an evaluative scheme for the appraisal of transition experiments has been validated. Previously, common evaluation and research methods tended to focus on a narrow aspect of transition experiments, such as legitimacy, resilience or the technical

implications of the experiment (e.g. Alexander et al., 2016; H. Mees et al., 2017; van Slobbe et al., 2013). The literature review described in chapters 2 and 3, however, exposed a large number of factors that are thought to be relevant when designing and conducting experiments for transformational system change. Based on the findings from this literature review, it can be argued that previous methods for the evaluation of transition experiments are too limited to gain a thorough understanding of the complexity of the experiments. This should send a warning signal to the academic community that previous research methods for the analysis of transition experiments could have been inadequate for gaining a thorough understanding of the dynamics of the experiments.

The framework adapted in this research is aimed at providing a comprehensive evaluative scheme, combining the ecological, institutional, economic and social aspects all present in the system the transition experiment takes place in. By combining these physical and immaterial parts of the system, a holistic evaluation or analysis can be conducted. Based on the experiences with applying this holistic framework, three different methods to apply this framework have been identified, each adhering to a different research goal. If applied in future research, these methods can lead to a significant improvement in the evaluation and analysis of transition experiments.

Additionally, the case study results have provided some interesting avenues for further research. As the topic of the governance transition experiments was sorely under researched, these tentative hypotheses and research avenues can provide a valuable starting point for further research. Furthermore, the findings from this research have added the first pages to the body of literature on the topic of the governance of transition experiments.

9.4.2. Reflection on societal relevance

The tentative statements about the governance of transition experiments made in this research have the clearest impact on society. Knowledge on how to design and conduct transition experiments can be instrumental in realizing transitions towards more resilient systems that are able to withstand the impact of climate change. Even more so than the tentative statements made in this research, future research based on the method and avenues for further research identified in this research can add to the practical knowledge on how to increase the effectiveness of transition experiments.

Indirectly, this research comments on the way program or experiment evaluations are typically conducted. This research showed that a positivist approach with predefined indicators is not only likely to exclude important experiment elements, but also provides little incentive to analyze the experiment beyond the 'tick-box' indicators. This could be why, despite a large number of transition experiments being conducted worldwide, the knowledge on the design and governance of these experiments is still very limited (Kivimaa et al., 2017). By including a manual for program evaluation making use of the evaluative scheme, this research offers policy makers the tools to combine both 'tick-box' evaluation and in-depth analysis to increase the policy learning about the governance of transition experiments.

9.4.3. Recommendations

This research uncovered a number of directions for future research. Interesting avenues for future research into the topic of the governance of transition experiments are the relationship between hands-on experience with the design and implementation of the experiment and the social change that can be achieved. The impact of the inclusion of good governance principles in transition experiments on the experiment outcomes can also be researched. The impact of context or input features on the experiment's activities and outcomes is also a fertile direction for further research that could contribute to a better understanding of the transition experiments and the desired forms of governance in different contexts. Additionally, the current evaluation

took place right after the experiment had ended, which means only the short-term outcomes of the experiment could be adequately assessed. Evaluation of the experiments at a later moment in time could help generate insights into the impact of the experiment on the long-term and the role governance played in achieving these long-term outcomes.

Some evaluative scheme-specific recommendations for further research can also be made. Firstly, the use of the evaluative scheme for the ex-ante design of experiments and evaluations still needs to be validated. Secondly, the continued application of the scheme to real-world cases can help refine the framework's features and identify features that should be added or excluded from the scheme. This includes the validation of the three added input features. Thirdly, applying a holistic approach when analyzing different transition experiments can help to gain a more thorough understanding of these experiments and how to design and conduct them in order to contribute to a transition towards a more resilient system.

Additionally, some tentative policy recommendations can be made based on the research findings. First, it appears that the governance of transition experiments could have a large impact on the outcomes. The careful design of the governance mechanisms and tools to be included in the experiment based on local best practices could increase the impact of the experiment. Furthermore, the experiment itself can serve as an opportunity for governance experimentation to add an additional layer of learning to the experiment. Next, it seems like intangible outcomes, such as changed perspectives and confidence, are as important as tangible outcomes, but much more difficult to evaluate and demonstrate. By considering these outcomes during the design phase of the experiment and designing indicators and baseline studies, they can be made more tangible. The evaluative scheme can be used as a source of inspiration for this process. Last, it can be advised to critically assess the methods for the evaluation of an experiment or broader program and to employ a pragmatic approach that encourages learning as well as evaluation.

9.4.4. Conclusion
Overall, it can be concluded that the evaluative scheme for the evaluation of transition experiments requires further development. This research exposed the challenges of applying the scheme in a positivist manner, but also poses that the use of the scheme in an interpretivist could assist in opening up the 'black box' of transition experiments. In addition, the research findings indicate that there is indeed a relation between governance and the outcomes of the experiment. This research contributed to the body of literature by identifying a method to analyze transition experiments as well as provide some directions for future research. In doing so, it has hopefully laid a stepping stone towards not only a developed body of knowledge on the governance of transition experiments, but also towards a more resilient planet. However, future research is needed to prove the hypotheses posed in this research and to identify new relations and confounding variables. On that note, it seems appropriate to conclude this report with the same words Barack Obama used to end his speech at the 2015 United Nations Climate Change Conference:

"Let's get to work!"

References

Abrams, P., Borrini-Feyerabend, G., Gardner, H. J., & Heylings, P. (2003). *Evaluating governance: A handbook to accompany a participatory process for a protected area*. Ottawa. Retrieved from https://portals.iucn.org/library/sites/library/files/documents/Rep-2003-022.pdf

Ahern, J. (2011). From fail-safe to safe-to-fail: Sustainability and resilience in the new urban world. *Landscape and Urban Planning*, *100*(4), 341–343. https://doi.org/10.1016/j.landurbplan.2011.02.021

Alexander, M., Priest, S., & Mees, H. (2016). A framework for evaluating flood risk governance. *Environmental Science and Policy*, *64*, 38–47. https://doi.org/10.1016/j.envsci.2016.06.004

Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory*, *18*(4), 543–571. https://doi.org/10.1093/jopart/mum032

Antikainen, R., Alhola, K., & Jääskeläinen, T. (2017). Experiments as a means towards sustainable societies – Lessons learnt and future outlooks from a Finnish perspective. *Journal of Cleaner Production*, *169*, 216–224. https://doi.org/10.1016/j.jclepro.2017.06.184

Antwerp Municipality. (2018). *Integrated Action Plan for Urban Resilience*. Antwerp.

Baarda, B., Van der Hulst, M., & De Goede, M. (2012). Wat is een interview? [What is an interview?]. In *Basisboek Interviewen* (pp. 11–39). Groningen: Noordhoff Uitgevers bv.

Bellamy, R. (2016). A Sociotechnical Framework for Governing Climate Engineering. *Science, Technology, & Human Values*, *41*(2), 135–162. https://doi.org/10.1177/0162243915591855

Béné, C., Mehta, L., Mcgranahan, G., Cannon, T., Gupte, J., & Tanner, T. (2017). Climate and Development Resilience as a policy narrative: potentials and limits in the context of urban planning. *Climate and Development*, *10*(2), 116–133. https://doi.org/10.1080/17565529.2017.1301868

Benson, D., Lorenzoni, I., & Cook, H. (2016). Evaluating social learning in England flood risk management: An "individual-community interaction" perspective. *Environmental Science and Policy*, *55*, 326–334. https://doi.org/10.1016/j.envsci.2015.05.013

Berkhout, F., Verbong, G., Wieczorek, A. J., Raven, R., Lebel, L., & Bai, X. (2010). Sustainability experiments in Asia: Innovations shaping alternative development pathways? *Environmental Science and Policy*, *13*(4), 261–271. https://doi.org/10.1016/j.envsci.2010.03.010

Bhamra, R., Ab, D., & Burnard, K. (2011). Resilience: the concept, a literature review and future directions. *International Journal of Production Research*, *49*(18), 5375–5393. https://doi.org/10.1080/00207543.2011.563826

Billig, S. (2000). Research on K-12 School-Based Service-Learning. *Phi Delta Kappan*, *81*(9), 658–664. Retrieved from https://www.researchgate.net/publication/283763081

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

Bouckaert, G., & van de Walle, S. (2003). Comparing Measures of Citizen Trust and User Satisfaction as Indicators of 'Good Governance': Difficulties in Linking Trust and Satisfaction Indicators. *International Review of Administrative Sciences*, *69*(3), 329–343. https://doi.org/10.1177/0020852303693003

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brenner, M. (1985). Intensive Interviewing. In M. Brenner, J. Brown, & D. Canter (Eds.), *The Research Interview: Uses and Approaches* (pp. 147–161). London: Academic Press.

Buizer, M., Elands, B., & Vierikko, K. (2016). Governing cities reflexively—The biocultural diversity concept as an alternative to ecosystem services. *Environmental Science and Policy*, *62*, 7–13. https://doi.org/10.1016/j.envsci.2016.03.003

Burgas Municipality. (2018). *Integrated Action Plan for Urban Resilience*. Burgas.

Buri, J. R., Gunty, A., Seel, N. M., Dukas, R., Even, R., Scheuer, O., … Kehoe, E. J. (2012). Experiential Learning Theory. In *Encyclopedia of the Sciences of Learning* (pp. 1215–1219). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-1428-6_227

Butler, C., & Pidgeon, N. (2011). From "flood defence" to "flood risk management": Exploring governance, responsibility, and blame. *Environment and Planning C: Government and Policy*, *29*(3), 533–547. https://doi.org/10.1068/c09181j

Caniglia, G., Schäpke, N., Lang, D. J., Abson, D. J., Luederitz, C., Wiek, A., … von Wehrden, H. (2017). Experiments and evidence in sustainability science: A typology. *Journal of Cleaner Production*, *169*, 39–47. https://doi.org/10.1016/j.jclepro.2017.05.164

Carpenter, S., Walker, B., Anderies, J. M., & Abel, N. (2001). From Metaphor to Measurement: Resilience of What to What? *Ecosystems*, *4*(8), 765–781. https://doi.org/10.1007/s10021-001-0045-9

Chen, H. T. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In *Evaluation von Programmen und Projekten für eine demokratische Kultur* (pp. 17–40). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-19009-9_2

Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools*, *44*(5), 451–459. https://doi.org/10.1002/pits.20237

Cialdini, R. (2009). Social Proof: Truths Are Us. In *Influence: The Psychology of Persuasion* (pp. 87–125). New York: HarperCollins. Retrieved from http://elibrary.bsu.az/books_400/N_232.pdf

Collins, B., Boyd, D., & Curzon, R. (2017). Exploring local projects for sustainable energy in system transition: local perceptions of success. *Technology Analysis and Strategic Management*, *29*(9), 1076–1088. https://doi.org/10.1080/09537325.2016.1268684

Contractor, N. S., & DeChurch, L. A. (2014). Integrating social networks and human social motives to achieve social influence at scale. *Proceedings of the National Academy of Sciences of the United States of America*, *111 Suppl*(Supplement 4), 13650–7. https://doi.org/10.1073/pnas.1401211111

Conway, G. (2007). Monitoring the state of the Solent. *Marine Policy*, *31*(5), 632–637. https://doi.org/10.1016/j.marpol.2007.03.009

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, *24*(2), 119–128. https://doi.org/10.1016/S0149-7189(01)00003-9

Doorn, N. (2016). Governance Experiments in Water Management: From Interests to Building Blocks. *Science and Engineering Ethics*, *22*(3), 755–774. https://doi.org/10.1007/s11948-015-9627-3

Dryzek, J. S., & Pickering, J. (2017). Deliberation as a catalyst for reflexive environmental governance. *Ecological Economics*, *131*, 353–360. https://doi.org/10.1016/j.ecolecon.2016.09.011

Du, J., Zhang, S., Wu, G., Moura, J. M. F., & Kar, S. (2017). Adaptive governance of social-ecological systems. *Annual Review of Environment and Resources*, *30*(1), 441–473. https://doi.org/10.1146/annurev.energy.30.050504.144511

European Climate Adaptation Platform. (2018). Uncertainty Guidance. Retrieved February 23, 2018, from http://climate-adapt.eea.europa.eu/knowledge/tools/uncertainty-guidance/topic2

European Commission. (2013a). *Impact Assessment*. Brussels.

European Commission. (2013b). *The EU Strategy on adaptation to climate change*. Brussels.

European Commission. (2014). *In-depth Report: E-Consultation on Nature-Based Solutions*. Retrieved from https://ec.europa.eu/research/environment/pdf/e-consultation_on_nature-based_solutions_in-depth_report.pdf#view=fit&pagemode=none

European Commission. (2015). *Towards an EU Research and Innovation policy agenda for Nature-Based Solutions & Re-Naturing Cities*. Brussels. https://doi.org/10.2777/765301

European Commission. (2018). Nature-Based Solutions. Retrieved February 1, 2018, from https://ec.europa.eu/research/environment/index.cfm?pg=nbs

European Environment Agency. (2016). *Urban adaptation to climate change in Europe 2016: Transforming cities in a changing climate*. Luxembourg.

Evers, M., Jonoski, A., Almoradie, A., & Lange, L. (2016). Collaborative decision making in sustainable flood risk management: A socio-technical approach and tools for participatory governance. *Environmental Science and Policy*, *55*, 335–344. https://doi.org/10.1016/j.envsci.2015.09.009

Farrelly, M., & Brown, R. (2011). Rethinking urban water management: Experimentation as a way forward? *Global Environmental Change*, *21*, 721–732. https://doi.org/10.1016/j.gloenvcha.2011.01.007

Feola, G., & Nunes, R. (2014). Success and failure of grassroots innovations for addressing climate change: The case of the transition movement. *Global Environmental Change*, *24*(1), 232–250. https://doi.org/10.1016/j.gloenvcha.2013.11.011

Feyen, L., Dankers, R., Bódis, K., Salamon, P., Barredo, J. I., Feyen, L., … Bódis, K. (2012). Fluvial flood risk in Europe in present and future climates. *Climatic Change*, *112*(1), 47–62. https://doi.org/10.1007/s10584-011-0339-7

Foley, R. W., Bernstein, M. J., & Wiek, A. (2016). Towards an alignment of activities, aspirations and stakeholders for responsible innovation. *Journal of Responsible Innovation*, *3*(3), 209–232. https://doi.org/10.1080/23299460.2016.1257380

Folke, C. (2006). Resilience: The emergence of a perspective for social-ecological systems analyses. *Global Environmental Change*, *16*(3), 253–267. https://doi.org/10.1016/j.gloenvcha.2006.04.002

Folke, C., Carpenter, S., Elmqvist, T., Gunderson, L., Holling, C. S., & Walker, B. (2002). Resilience and Sustainable Development: Building Adaptive Capacity in a World of Transformations. *Journal of Human Environment*, *31*(5), 437–440. https://doi.org/10.1579/0044-7447-31.5.437

Folke, C., Carpenter, S. R., Walker, B., Scheffer, M., Chapin, T., & Rockström, J. (2010). Resilience thinking: Integrating resilience, adaptability and transformability. *Ecology and Society*, *15*(4). https://doi.org/10.5751/ES-03610-150420

Forrest, N., & Wiek, A. (2014). Learning from success - Toward evidence-informed sustainability transitions in communities. *Environmental Innovation and Societal Transitions*, *12*, 66–88. https://doi.org/10.1016/j.eist.2014.01.003

Forrest, N., & Wiek, A. (2015). Success factors and strategies for sustainability transitions of small-scale communities - Evidence from a cross-case analysis. *Environmental Innovation and Societal Transitions*, *17*, 22–40. https://doi.org/10.1016/j.eist.2015.05.005

Fortier, F. (2010). Taking a climate chance: A procedural critique of Vietnam's climate change strategy. *Asia Pacific Viewpoint*, *51*(3), 229–247. https://doi.org/10.1111/j.1467-8373.2010.01428.x

Geaves, L. H., & Penning-Rowsell, E. C. (2016). Flood Risk Management as a public or a private good, and the implications for stakeholder engagement. *Environmental Science and Policy*, *55*, 281–291. https://doi.org/10.1016/j.envsci.2015.06.004

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, *31*(8–9), 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

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

Gerring, J., & Seawright, J. (2007). Techniques for choosing cases. In J. Gerring (Ed.), *Case Study Research: Principles and Practice* (pp. 86–150). New York: Cambridge University Press.

Gibson, R. B. (2006). Sustainability assessment: Basic components of a practical approach. *Impact Assessment and Project Appraisal*, *24*(3), 170–182. https://doi.org/10.3152/147154606781765147

Gill, S. ., Handley, J. ., Ennos, A. ., & Pauleit, S. (2007). Adapting Cities for Climate Change: The Role of the Green Infrastructure. *Built Environment*, *33*(1), 115–133. https://doi.org/10.2148/benv.33.1.115

Grabs, J., Langen, N., Maschkowski, G., & Schäpke, N. (2016). Understanding role models for change: a multilevel analysis of success factors of grassroots initiatives for sustainable consumption. *Journal of Cleaner Production*, *134*, 98–111. https://doi.org/10.1016/j.jclepro.2015.10.061

Gralepois, M., Larrue, C., Wiering, M., Crabbé, A., Tapsell, S., Mees, H., … Szwed, M. (2016). Is flood defense changing in nature? Shifts in the flood defense strategy in six European countries. *Ecology and Society*, *21*(4), art37. https://doi.org/10.5751/ES-08907-210437

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.2753/MIS0742-1222240302

Halliday, A., & Glaser, M. (2011). A Management Perspective on Social Ecological Systems: A generic system model and its application to a case study from Peru. *Human Ecology Review*, *18*(1). Retrieved from http://www.humanecologyreview.org/pastissues/her181/halliday.pdf

Haque, M., Shyaka, A., & Mudacumura, G. (2017). Rethinking governance and democracy in developing countries. In M. Haque, A. Shyaka, & G. Mudacumura (Eds.), *Democratizing Public Governance in Developing Nations* (pp. 13–23). London: Routledge.

Hartmann, T., & Driessen, P. (2017). The flood risk management plan: towards spatial water governance. *Journal of Flood Risk Management*, *10*(2), 145–154. https://doi.org/10.1111/jfr3.12077

Hartmann, T., & Spit, T. (2016). Legitimizing differentiated flood protection levels - Consequences of the European flood risk management plan. *Environmental Science and Policy*, *55*, 361–367. https://doi.org/10.1016/j.envsci.2015.08.013

Hatcher, R. G. B. and J. A. (1996). Implementing Service Learning in Higher Education. *The Journal of Higher Education*, *67*(2), 221–239. Retrieved from http://www.jstor.org/stable/2943981

Hegger, D. L. T., Driessen, P. P. J., Dieperink, C., Wiering, M., Raadgever, G. T. T., & van Rijswick, H. F. M. W. (2014). Assessing stability and dynamics in flood risk governance: An empirically illustrated research approach. *Water Resources Management*, *28*(12), 4127–4142. https://doi.org/10.1007/s11269-014-0732-x

Hegger, D. L. T., Driessen, P. P. J., Wiering, M., Rijswick, H. F. M. W. Van, Kundzewicz, Z. W., Matczak, P., … Ek, K. (2016). Toward more flood resilience: Is a diversification of flood risk management strategies the way forward? *Ecology and Society*, *21*(4). Retrieved from http://hdl.handle.net/2066/162768

Heintz, M. D., Hagemeier-Klose, M., & Wagner, K. (2012). Towards a risk governance culture in flood policy-findings from the implementation of the "floods directive" in Germany. *Water (Switzerland)*, *4*(1), 135–156. https://doi.org/10.3390/w4010135

Hendriks, C. M., & Grin, J. (2007). Contextualizing reflexive governance: The politics of Dutch transitions to sustainability. *Journal of Environmental Policy and Planning*, *9*(3–4), 333–350. https://doi.org/10.1080/15239080701622790

Heusinkveld, B. G., Steeneveld, G. J., van Hove, L. W. A., Jacobs, C. M. J., & Holtslag, A. A. M. (2014). Spatial variability of the Rotterdam urban heat island as influenced by urban land use. *Journal of Geophysical Research: Atmospheres*, *119*(2), 677–692. https://doi.org/10.1002/2012JD019399

Hildén, M., Jordan, A., & Huitema, D. (2017). Special issue on experimentation for climate change solutions editorial: The search for climate change and sustainability solutions - The promise and the pitfalls of experimentation. *Journal of Cleaner Production*, *169*, 1–7. https://doi.org/10.1016/j.jclepro.2017.09.019

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., … Kanae, S. (2013).

Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

Holling, C. S. (1973). Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics*, *4*, 1–23. Retrieved from http://www.jstor.org/stable/2096802

Johnson, C., Penning-Rowsell, E. C., & Parker, D. (2007). Natural and imposed injustices: the challenges in implementing 'fair' flood risk management policy in England. *Geographical Journal*, *173*(4), 374–390. https://doi.org/10.2307/30130632

Julian, D. A. (1997). The utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning*, *20*(3), 251–257. https://doi.org/10.1016/S0149-7189(97)00002-5

Kabisch, N., Frantzeskaki, N., Pauleit, S., Naumann, S., Davis, M., Artmann, M., … Bonn, A. (2016). Nature-based solutions to climate change mitigation and adaptation in urban areas: Perspectives on indicators, knowledge gaps, barriers, and opportunities for action. *Ecology and Society*, *21*(2). https://doi.org/10.5751/ES-08373-210239

Keesstra, S., Nunes, J., Novara, A., Finger, D., Avelar, D., Kalantari, Z., & Cerdà, A. (2018). The superior effect of nature based solutions in land management for enhancing ecosystem services. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2017.08.077

Kivimaa, P., Hildén, M., Huitema, D., Jordan, A., & Newig, J. (2017). Experiments in climate governance – A systematic review of research on energy and built environment transitions. *Journal of Cleaner Production*, *169*, 17–29. https://doi.org/10.1016/J.JCLEPRO.2017.01.027

Krathwohl, D. (2009). Experimentation and Experimental Design. In *Methods of Educational and Social Science Research: The logic of methods* (pp. 477–518). Long Grove: Waveland Press, Inc.

Krieger, K. (2013). The limits and variety of risk-based governance: The case of flood management in Germany and England. *Regulation and Governance*, *7*(2), 236–257. https://doi.org/10.1111/rego.12009

Kuklicke, C., & Demeritt, D. (2016). Adaptive and risk-based approaches to climate change and the management of uncertainty and institutional risk: The case of future flooding in England. *Global Environmental Change*, *37*, 56–68. https://doi.org/10.1016/j.gloenvcha.2016.01.007

Lafortezza, R., Chen, J., van den Bosch, C. K., & Randrup, T. B. (2018). Nature-based solutions for resilient landscapes and cities. *Environmental Research*. https://doi.org/10.1016/j.envres.2017.11.038

Larrue, C., Hegger, D., & Trémorin, J. (2013). *Researching Flood Risk Governance in Europe : background theories*. Utrecht.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., … Zhu, D. (2017). Global Carbon Budget 2017. *Earth System Science Data Discussions*, 1–79. https://doi.org/10.5194/essd-2017-123

Lebel, L., Anderies, J. M., Campbell, B., & Folke, C. (2006). Governance and the Capacity to Manage Resilience in Regional Social-Ecological Systems. *Marine Sciences Faculty Scholarship*, *11*(1).

Leskens, J., Boomgaard, M., Zuijlen, C. Van, & Hollanders, P. (2013). A multi-layer flood safety approach towards resilient cities. In *International Conference on Flood Resilience, Experiences in Asia and Europe, 5-7 September 2013, Exeter UK* (Vol. 5, pp. 1–8). Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/5525448

Loorbach, D., & Shiroyama, H. (2016). The Challenge of Sustainable Urban Development and Transforming Cities. In D. Loorbach, J. Wittmayer, H. Shiroyama, J. Fuijino, & S. Mizuguchi (Eds.), *Theory and Practice of Urban Sustainability Transitions Governance of Urban Sustainability Transitions* (1st ed., pp. 3–12). Tokyo: Springer Japan. Retrieved from https://link-springer-com.focus.lib.kth.se/content/pdf/10.1007%2F978-4-431-55426-4.pdf

Luederitz, C., Schapke, N., Wiek, A., Lang, D. J., Bergmann, M., Bos, J. J., … Westley, F. R. (2017). Learning through evaluation - A tentative evaluative scheme for sustainability transition experiments. *Journal of Cleaner Production*, *167*, 61–76. https://doi.org/10.1016/j.jclepro.2016.09.005

Maes, J., & Jacobs, S. (2017). Nature-Based Solutions for Europe's Sustainable Development. *Conservation Letters*, *10*(1), 121–124. https://doi.org/10.1111/conl.12216

McLaughlin, J., & Jordan, G. (2015). Using logic models. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 62–87). John Wiley & Sons. https://doi.org/10.1002/9781119171386.ch3

Meerow, S., Newell, J. P., & Stults, M. (2016). Defining urban resilience: A review. *Landscape and Urban Planning*, *147*, 38–49. https://doi.org/10.1016/j.landurbplan.2015.11.011

Mees, H., Crabbé, A., & Driessen, P. (2017). Conditions for citizen co-production in a resilient, efficient and legitimate flood risk governance arrangement. A tentative framework. *Journal of Environmental Policy & Planning*, *16*(6), 827–842. https://doi.org/10.1080/1523908X.2017.1299623

Mees, H. L. P., Driessen, P. P. J., & Runhaar, H. A. C. (2014). Legitimate adaptive flood risk governance beyond the dikes: The cases of Hamburg, Helsinki and Rotterdam. *Regional Environmental Change*, *14*(2), 671–682. https://doi.org/10.1007/s10113-013-0527-2

Mentens, J., Raes, D., & Hermy, M. (2006). Green roofs as a tool for solving the rainwater runoff problem in the urbanized 21st century? *Landscape and Urban Planning*, *77*(3), 217–226. https://doi.org/10.1016/j.landurbplan.2005.02.010

Nadler, J., Weston, R., & Voyles, E. (2015). Stuck in the Middle: The Use and Interpretation of Mid-Points in Items on Questionnaires. *The Journal of General Psychology*, *142*(2), 71–89.

NASA. (2018). Climate Change: Vital Signs of the Planet: Effects. Retrieved February 1, 2018, from https://climate.nasa.gov/effects/

Nesshöver, C., Assmuth, T., Irvine, K. N., Rusch, G. M., Waylen, K. A., Delbaere, B., … Wittmer, H. (2017). The science, policy and practice of nature-based solutions: An interdisciplinary perspective. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2016.11.106

Newcomer, K., Hatry, H. P., & Wholey, J. S. (2015). Planning and Designing Useful Evaluations. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation: Fourth Edition* (4th ed., pp. 7–35). John Wiley & Sons. https://doi.org/10.1002/9781119171386.ch1

Newig, J., Kochskämper, E., Challies, E., & Jager, N. W. (2016). Exploring governance learning: How policymakers draw on evidence, experience and intuition in designing participatory flood risk planning. *Environmental Science and Policy*, *55*, 353–360. https://doi.org/10.1016/j.envsci.2015.07.020

Olsson, P., Folke, C., & Berkes, F. (2004). Adaptive Comanagement for Building Resilience in Social-Ecological Systems. *Environmental Management*, *34*(1). https://doi.org/10.1007/s00267-003-0101-7

Ott, J. C. (2011). Government and Happiness in 130 Nations: Good Governance Fosters Higher Level and More Equality of Happiness. *Social Indicators Research*, *102*(1), 3–22. https://doi.org/10.1007/s11205-010-9719-z

Pahl-Wostl, C., Palmer, M., & Richards, K. (2013). Enhancing water security for the benefits of humans and nature — the role of governance Introduction: water security definition and scope of the review. *Current Opinion in Environmental Sustainability*, *5*, 676–684. https://doi.org/10.1016/j.cosust.2013.10.018

Patton, M. Q. (2002). Two Decades of Developments in Qualitative Inquiry. *Qualitative Social Work: Research and Practice*, *1*(3), 261–283. https://doi.org/10.1177/1473325002001003636

Pickett, S., Cadenasso, M. L., & Grove, J. M. (2004). Resilient cities: Meaning, models, and metaphor for integrating the ecological, socio-economic, and planning realms. *Landscape and Urban Planning*, *69*(4), 369–384. https://doi.org/10.1016/j.landurbplan.2003.10.035

Porter, J., & Demeritt, D. (2012). Flood-risk management, mapping, and planning: The institutional politics of decision support in England. *Environment and Planning A*, *44*(10), 2359–2378. https://doi.org/10.1068/a44660

Posavac, E. (2016). Developing and Using a Theory of the Program. In *Program Evaluation: Methods and Case Studies* (pp. 54–67). New York: Taylor & Francis.

Priest, S. J., Suykens, C., Rijswick, H. F. M. W. Van, Schellenberger, T., Goytia, S., Kundzewicz, Z. W., … Homewood, S. (2016). The European Union approach to flood risk management and improving societal resilience lessons from the implementation of the Floods Directive in six European countries. *Source: Ecology and Society*, *21*(4). Retrieved from http://www.jstor.org/stable/26270028

Ravesteijn, W., & Kroesen, O. (2015). Design for values in water management. In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 847–867). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6970-0_31

Redman, C. L., Grove, J. M., & Kuby, L. H. (2004). Integrating Social Science into the Long-Term Ecological Research (LTER) Network: Social Dimensions of Ecological Change and Ecological Dimensions of Social Change. *Ecosystems*, *7*(2), 161–171. https://doi.org/10.1007/s10021-003-0215-z

Reed, M., Fraser, E., & Dougill, A. (2006). An adaptive learning process for developing and applying sustainability indicators with local communities. *Ecological Economics*, *59*(4), 406–418. https://doi.org/10.1016/j.ecolecon.2005.11.008

Reed, S., Friend, R., Jarvie, J., Henceroth, J., Thinphanga, P., Singh, D., … Sutarto, R. (2015). Resilience projects as experiments: implementing climate change resilience in Asian cities. *Climate and Development*, *7*(5), 469–480. https://doi.org/10.1080/17565529.2014.989190

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. https://doi.org/10.1007/BF01405730

Rotmans, J., Kemp, R., & Van Asselt, M. (2001). More evolution than revolution: Transition Management in Public Policy. *Foresight*, *3*(1), 15–31. https://doi.org/10.1108/MBE-09-2016-0047

Rotmans, J., & Loorbach, D. A. (2008). Transition management: reflexive governance of societal complexity through searching, learning and experimenting. In J. Van den Bergh & F. Bruinsma (Eds.), *Managing the transition to renewable energy* (pp. 15–46). Cheltenham: Edward Elgar Publishing. Retrieved from https://www.narcis.nl/publication/RecordID/oai%3Arepub.eur.nl%3A37236/coll/person/id/18/Language/nl

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, *14*(2), 131–164. https://doi.org/10.1007/s10664-008-9102-8

Saleh, F., & Weinstein, M. P. (2016). The role of nature-based infrastructure (NBI) in coastal resiliency planning: A literature review. *Journal of Environmental Management*, *183*, 1088–1098. https://doi.org/10.1016/j.jenvman.2016.09.077

Savaya, R., & Waysman, M. (2005). The logic model: A tool for incorporating theory in development and evaluation of programs. *Administration in Scoial Work*, *29*(2), 85–103. https://doi.org/10.1300/J147v29n02_06

Schanze, J. (2017). Nature-based solutions in flood risk management - Buzzword or innovation? *Journal of Flood Risk Management*, *10*(3), 281–282. https://doi.org/10.1111/jfr3.12318

Schäpke, N., Omann, I., Wittmayer, J. M., van Steenbergen, F., & Mock, M. (2017). Linking transitions to sustainability: A study of the societal effects of transition management. *Sustainability (Switzerland)*, *9*(5), 1–36. https://doi.org/10.3390/su9050737

Schot, J., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management*, *20*(5), 537–554. https://doi.org/10.1080/09537320802292651

Scott, C. (2010). Reflexive Governance, Regulation and Meta-Regulation: Control or Learning? In O. De Schutter & J. Lenoble (Eds.), *Reflexive Governance: Redefining the Public Interest in a Pluralistic World* (pp. 43–66). Oxford: Hart Publishing.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., … Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*(1), 10. https://doi.org/10.1186/1471-2288-7-10

Simmons, G., Giraldo, J. E. D., Truong, Y., & Palmer, M. (2018). Uncovering the link between governance as an innovation process and socio-economic regime transition in cities. *Research Policy*, *47*(1), 241–251. https://doi.org/10.1016/j.respol.2017.11.002

Sol, J., van der Wal, M. M., Beers, P. J., & Wals, A. (2017). Reframing the future: the role of reflexivity in governance networks in sustainability transitions. *Environmental Education Research*, 1–23. https://doi.org/10.1080/13504622.2017.1402171

Sustainable Governance Indicators. (2018a). Governance. Retrieved March 26, 2018, from http://www.sgi-network.org/2017/Governance

Sustainable Governance Indicators. (2018b). Quality of Democracy. Retrieved March 26, 2018, from http://www.sgi-network.org/2017/Democracy/Quality_of_Democracy

Swanborn, P. G. (2008). *Case-study's: Wat, Wanneer en Hoe? [Case studies: What, When and How?]* (4th ed.). Amsterdam: Boom Onderwijs.

Termeer, C. J. A. M., Dewulf, A., Karlsson-Vinkhuyzen, S. I., Vink, M., & van Vliet, M. (2016). Coping with the wicked problem of climate adaptation across scales: The Five R Governance Capabilities. *Landscape and Urban Planning*, *154*, 11–19. https://doi.org/10.1016/j.landurbplan.2016.01.007

Thaler, T., & Levin-Keitel, M. (2016). Multi-level stakeholder engagement in flood risk management-A question of roles and power: Lessons from England. *Environmental Science and Policy*, *55*, 292–301. https://doi.org/10.1016/j.envsci.2015.04.007

The World Bank. (2018a). GNI per capita, Atlas method (current US$). Retrieved March 26, 2018, from https://data.worldbank.org/indicator/NY.GNP.PCAP.CD

The World Bank. (2018b). World Bank Country and Lending Groups. Retrieved March 26, 2018, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Tyler, S., & Moench, M. (2012). Climate and Development A framework for urban climate resilience. *Climate and Development*, *4*(4), 311–326. https://doi.org/10.1080/17565529.2012.745389

UN Department of Economic and Social Affairs. (2004). *Guidelines for Reducing Flood Losses*. New York. Retrieved from www.un.org/esa/sustdev

United Nations Economic and Social Council. (2006). *Definition of basic concepts and terminologies in governance and public administration*. New York. https://doi.org/EB.AIR/WG.5/2001/7

United Nations Environment Programme. (2018). Nature-based solutions offer local benefits through local action. Retrieved February 23, 2018, from https://www.unenvironment.org/news-and-stories/story/nature-based-solutions-offer-local-benefits-through-local-action

United States Office of Energy Efficiency and Renewable Energy. (2018). Program Evaluation: Program Life Cycle. Retrieved March 26, 2018, from https://www.energy.gov/eere/analysis/program-evaluation-program-life-cycle

URBACT. (2014). URBACT at a glance. Retrieved March 26, 2018, from http://urbact.eu/urbact-glance

Van Buggenum, L. J. (2018). *Governance of Experimentation: a framework for evaluating transition experiments for urban flood risk management*. Delft University of Technology.

van Buuren, A., Ellen, G. J., & Warner, J. F. (2016). Path-dependency and policy learning in the dutch delta: Toward more resilient flood risk management in the Netherlands? *Ecology and Society*, *21*(4).

https://doi.org/10.5751/ES-08765-210443

Van De Meene, S. J., Brown, R. R., & Farrelly, M. A. (2011). Towards understanding governance for sustainable urban water management. *Global Environmental Change*, *21*, 1117–1127. https://doi.org/10.1016/j.gloenvcha.2011.04.003

Van Den Bosch, S., & Rotmans, J. (2008). *Deepening, Broadening and Scaling up: A Framework for Steering Transition Experiments*. Rotterdam. Retrieved from http://sus-i.nl/_files/KCT_transitieboekje_02.pdf

van den Heiligenberg, H. A. R. M., Heimeriks, G. J., Hekkert, M. P., & van Oort, F. G. (2017). A habitat for sustainability experiments: Success factors for innovations in their local and regional contexts. *Journal of Cleaner Production*, *169*, 204–215. https://doi.org/10.1016/j.jclepro.2017.06.177

Van Der Nat, A., Vellinga, P., Leemans, R., & Van Slobbe, E. (2016). Ranking coastal flood protection designs from engineered to nature-based. *Ecological Engineering*, *87*, 80–90. https://doi.org/10.1016/j.ecoleng.2015.11.007

van Herk, S., Zevenbergen, C., Ashley, R., & Rijke, J. (2011). Learning and Action Alliances for the integration of flood risk management into urban planning: A new framework from empirical evidence from The Netherlands. *Environmental Science and Policy*, *14*(5), 543–554. https://doi.org/10.1016/j.envsci.2011.04.006

van Popering-Verkerk, J., & van Buuren, A. (2017). Developing collaborative capacity in pilot projects: Lessons from three Dutch flood risk management experiments. *Journal of Cleaner Production*, *169*, 225–233. https://doi.org/10.1016/j.jclepro.2017.04.141

van Slobbe, E., de Vriend, H. J., Aarninkhof, S., Lulofs, K., de Vries, M., & Dircke, P. (2013). Building with Nature: In search of resilient storm surge protection strategies. *Natural Hazards*, *66*(3), 1461–1480. https://doi.org/10.1007/s11069-013-0612-3

Van Wesenbeeck, B. K., Mulder, J. P. M., Marchand, M., Reed, D. J., De Vries, M. B., De Vriend, H. J., … Van Wesenbeeck, B. K. (2014). Invited feature-FLA Damming deltas: A practice of the past? Towards nature-based flood defenses. *Estuarine, Coastal and Shelf Science*, *140*, 1–6. https://doi.org/10.1016/j.ecss.2013.12.031

Voss, J., & Bornemann, B. (2011). The Politics of Reflexive Governance: Challenges for Designing Adaptive Management and Transition Management. *Ecology and Society*, *16*(2), 9. https://doi.org/9

Voss, J., & Kemp, R. (2005a). Reflexive Governace: Learning to cope with fundamental limitations in steering sustainable development. Retrieved from http://kemp.unu-merit.nl/pdf/Voss Kemp Reflexive Governance FUTURES.pdf

Voss, J., & Kemp, R. (2005b). *Reflexive Governance for Sustainable Development – Incorporating feedback in social problem solving in Lisbon Special session on Transition management*. Lisbon. Retrieved from http://www.orobievive.net/conoscere/politiche_sostenibilita.pdf

Voss, J., & Kemp, R. (2006). Sustainability and reflexive governance: introduction. In J. Voß, D. Bauknecht, & R. Kemp (Eds.), *Reflexive Governance for Sustainable Development* (pp. 3–28). Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781847200266.00009

Voß, J. P., Smith, A., & Grin, J. (2009). Designing long-term policy: Rethinking transition management. *Policy Sciences*, *42*(4), 275–302. https://doi.org/10.1007/s11077-009-9103-5

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

Vuik, V., Jonkman, S. N., Borsje, B. W., & Suzuki, T. (2016). Nature-based flood protection: The efficiency of vegetated foreshores for reducing wave loads on coastal dikes. *Coastal Engineering*, *116*, 42–56. https://doi.org/10.1016/j.coastaleng.2016.06.001

Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, Adaptability and Transformability in Social – ecological Systems. *Ecology and Society*, *9*(2). https://doi.org/10.1103/PhysRevLett.95.258101

Walker, G., & Shove, E. (2007). Ambivalence, sustainability and the governance of socio-technical transitions. *Journal of Environmental Policy and Planning*, *9*(3–4), 213–225. https://doi.org/10.1080/15239080701622840

Walsh, N. (2017). Six reasons cities are located where they are. Retrieved February 22, 2018, from https://www.archdaily.com/866821/the-origin-of-cities-six-reasons-for-their-location

Ward, P. J., Pauw, W. P., van Buuren, M. W., & Marfai, M. A. (2013). Governance of flood risk management in a time of climate change: The cases of Jakarta and Rotterdam. *Environmental Politics*, *22*(3), 518–536. https://doi.org/10.1080/09644016.2012.683155

Watts, S., & Stenner, P. (2012). Basic Design Issues: Research Questions and Q sets. In *Doing Q Methodological Research: Theory, Method and Interpretation* (pp. 50–66). London: Sage Publications, Inc. https://doi.org/10.4135/9781446251911

Weiland, S., Bleicher, A., Polzin, C., Rauschmayer, F., & Rode, J. (2017). The nature of experiments for sustainability transformations: A search for common ground. *Journal of Cleaner Production*, *169*, 30–38.

https://doi.org/10.1016/j.jclepro.2017.06.182

Weiss, R. (1994). Preparation for Interviewing. In *Learning from Strangers: The Art and Method of Qualitative Interview Studies* (pp. 39–60). New York: The Free Press.

Werbeloff, L., & Brown, R. (2011). Working towards sustainable urban water management: The vulnerability blind spot. *Water Science and Technology*, *64*(12), 2362–2369. https://doi.org/10.2166/wst.2011.774

White House Office of the Press Secretary. (2015). Remarks by President Obama at the First Session of COP21. Retrieved February 22, 2018, from https://obamawhitehouse.archives.gov/the-press-office/2015/11/30/remarks-president-obama-first-session-cop21

Wiek, A., Kay, B., & Forrest, N. (2017). Worth the Trouble?! An Evaluative Scheme for Urban Sustainability Transition Labs (USTLs) and an Application to the USTL in Phoenix, Arizona. In N. Frantzeskaki, V. Castan Broto, L. Coenen, & D. Loorbach (Eds.), *Urban Sustainability Transitions* (pp. 227–256). Routledge. https://doi.org/10.1080/13604813.2014.892773

Wilby, R. L., & Keenan, R. (2012). Adapting to flood risk under climate change. *Progress in Physical Geography*, *36*(3), 348–378. https://doi.org/10.1177/0309133312438908

Wu, J., & Wu, T. (2013). Ecological Resilience as a Foundation for Urban Design and Sustainability. In S. T. Pickett, M. L. Cadenasso, & B. McGrath (Eds.), *Resilience in ecology and urban design : linking theory and practice for sustainable cities* (pp. 211–229). Dordrecht: Springer Netherlands. Retrieved from http://www.springer.com/series/8178

Xing, Y., Jones, P., & Donnison, I. (2017, January 20). Characterisation of nature-based solutions for the built environment. *Sustainability (Switzerland)*. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/su9010149

Yin, R. (2012a). *Applications of case study research*. *Applications of case study research*. Thousand Oaks, Calif: SAGE.

Yin, R. (2012b). Rival Explanations. In *Applications of case study research* (pp. 117–130). Thousand Oaks: Sage Publications, Inc.

Yin, R. (2014). Designing Case Studies: Identifying your case(s) and establishing the logic of your case study. In *Case Study Research: Design and Methods* (Fifth, pp. 27–70). Thousand Oaks, Calif: Sage Publications, Inc.

Yousaf, M., Ihsan, F., & Ellahi, A. (2016). Exploring the impact of good governance on citizens' trust in Pakistan. *Government Information Quarterly*, *33*(1), 200–209. https://doi.org/10.1016/j.giq.2015.06.001

Zainal, Z. (2007). Case study as a research method. *Jurnal Kemanusiaan Bil*, *9*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.956.301&rep=rep1&type=pdf

Governance of Experimentation
A framework for evaluating transition experiments for urban flood risk management
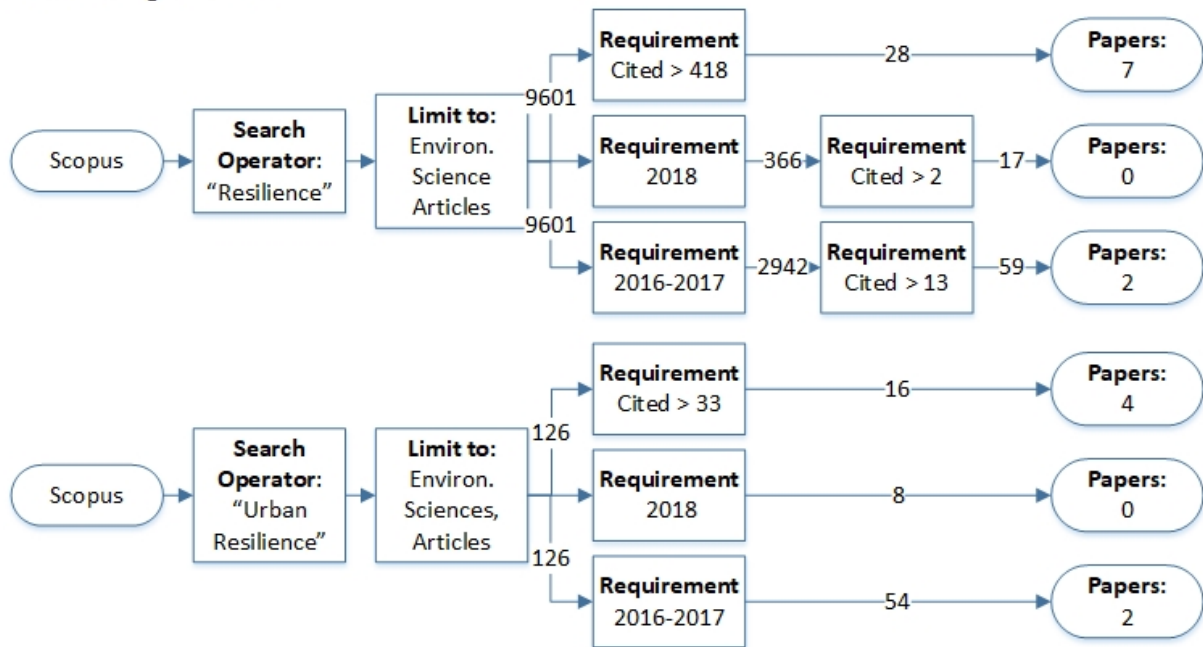
Appendix

Appendix A: Literature Search Protocol

In order to select the literature used in the literature review in chapter 3, the online search and database engine Scopus has mainly been used. Based on the object of research, different search operators have been used. A search flowchart has been defined as follows:

1. Does the search operator yield less than 100 documents?
   - Yes? Review documents
   - No? Continue to step 2
2. Does limiting the search to the relevant scientific fields and document types yield less than 100 documents?
   - Yes? Review documents
   - No? Continue to steps 3, 4 and 5
3. Limit the search to [number of citations = MAX] and [number of citations = 0.2*MAX]
4. Limit the search to papers published in 2018
   Does limiting the search to papers published in 2018 yield less than 100 documents?
   - Yes? Review documents
   - No? Continue to step 4.1.
   4.1. Limit the search to [number of citations = MAX] and [number of citations = 0.2*MAX]
5. Limit the search to papers published in 2017-2018
   Does limiting the search to papers published in 2018 yield less than 100 documents?
   - Yes? Review documents
   - No? Continue to step 5.1.
   5.1. Limit the search to [number of citations = MAX] and [number of citations = 0.2*MAX]
6. Did the search yield more than 15 relevant papers?
   - Yes? Literature search is finished
   - No? Repeat process with new search operator

Given the limited time available, the choice has been made to limit the review of the summaries of the articles to a maximum of 100 summaries per detailed search operator. Though in most cases many more articles were available, limiting the search to 100 articles was deemed sufficient to identify the relevant articles, without spending too much precious time on the selection process. Similarly, the stopping rule for reviewing the most cited documents was defined as the papers that were cited less often than 20 percent of the most cited paper. This stopping rule was chosen because it led to a manageable number of papers, in this review between 16 and 59, for each of the search strings. This number is independent of the number of papers included in the search, as such a dependency could have been an issue had an absolute stopping rule been used. For example: for a search string yielding 2000 documents, reviewing 100 documents would mean that only the most relevant papers were included. For a search string yielding only 130 results, however, it would mean that not the most relevant papers, but nearly all papers were included. As the focus of this literature review is not on all literature, but on the most recent or relevant literature, such a stopping rule would defeat the purpose of the search.
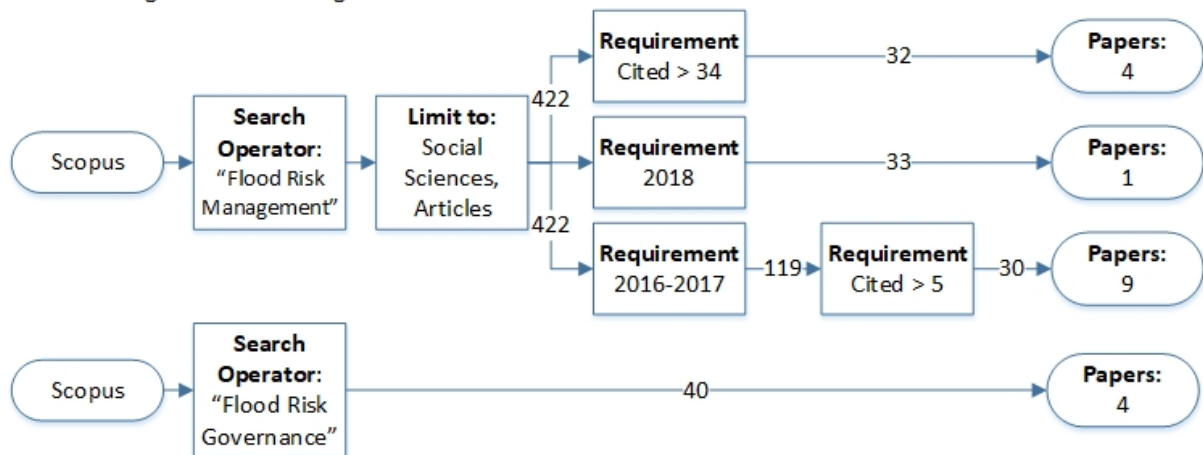
Figure A: Search Strings Literature Review. The numbers on the arrows indicate the number of documents found.
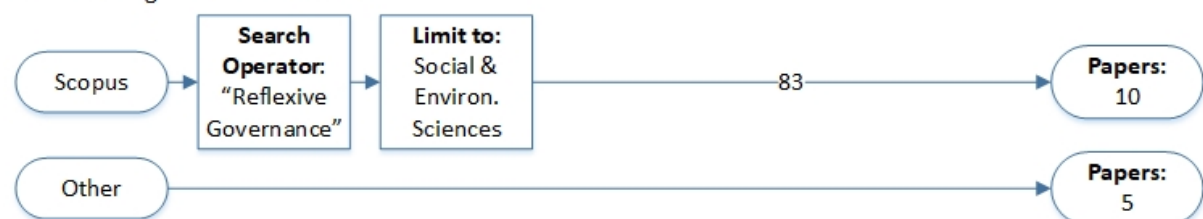
Appendix B: Case Study Protocol

B.1. Change record
No changes have been made to the interview protocol.

B.2. Background and research design
B.2.1. Mission and goals
These case studies are conducted as part of a master thesis research at the Delft University of Technology. The main goal of this research is therefore to prove a thorough command of the academic skills learned during the Complex Systems Engineering and Management master program. The main target audience for this research is therefore the graduation committee. However, a secondary goal is to contribute to the body of knowledge on the design of experiments for sustainability transitions. This means that the outcomes of this research, in the form of a scientific paper, should also be presented to the academic community. Although the research is conducted in collaboration with the Dutch Research Institute for Transitions, no sponsor interests influence the direction or desirable outcomes of the research.

B.2.2. Research questions
As described in paragraph 2.1, the goal of this research is to contribute to the body of knowledge on the design of transition experiments. This research goal is derived from the hypothesis that the governance of transition experiments plays an important role in the perceived success of such experiments. As the academic field of transition experiment research is new and rather underdeveloped, there is little guidance as to how to approach the validation of this hypothesis. The choice has therefore been made to narrow down the scope of the research to identify a method that can be used in future research to validate the hypothesis. The main research question is defined as follows:

How can the governance of transition experiments be evaluated in relation to the experiment's success, focusing on flood risk management?

As this research question is exploratory in nature, no formal hypothesis to be proven or disproven can be formulated. In order to answer the main research question, however, a number of sub-questions could be formulated:

1. *How can the evaluation of transition experiments be structured?*
2. *What are the characteristics of transition experiments to be included in the evaluation?*
3. *How can the characteristics be operationalized to allow for evaluation?*
4. *How does the governance of transition experiments contribute to their success?*

B.2.3. Research design
Based on the theoretical framework, described extensively in the main body of the report this research protocol is included in, a number of important features and indicators of transition experiments for flood risk management could be identified. Subsequently, a multiple-case, embedded research design has been selected as the most appropriate design for conducting this case study research. Two experiments, relatively similar yet taking place under highly different contexts have been selected for this research: the 'Green Corridor' experiment in Antwerp, Belgium and the 'Green Belt' experiment in Burgas, Bulgaria. The research will focus on the governance activities within the experiments and will link these to the context as well as the outcomes in order to gain a holistic understanding of the workings of these governance activities, while maintaining the research focus on governance.

B.2.4. Role of research protocol

The contents of chapter two provided a summary of the main text of the master thesis report related to this research. This main text will provide further elaboration on choices and methods. This research protocol, on the other hand, will provide practical guidance for conducting the case study research, including data collection procedures, questions and subsequent analysis of the data.

B.3. Data collection procedures

As the case study research will be conducted by a single research, as opposed to a team of researchers, there is less of a need for highly detailed, formalized data collection procedures to ensure comparable data is being collected. Some data collection procedures will be formalized nevertheless in order to ensure the quality of the data being collected.

B.3.1. Data sources

Two major types of data analysis methods have been identified: document analysis and stakeholder interviewing. The main benefit of using different sources of information is the possibility to triangulate the data. In doing so, different perspectives can be integrated and the validity, completeness and richness of the data can be enhanced.

For the *document analysis,* the most important document will be the experiment reports produced within the RESILIENT EUROPE framework for both cases. It is expected that these documents will provide a basic outline of the experiments, their design and their outcomes that can serve as a basis for further research. In addition to the experiment reports, various policy documents and communication documents will be analyzed to fill in any gaps and to provide more detail, especially on the experimentation process. The document analysis overall is expected to provide information for the case study as a whole as well as provide direction for the stakeholder interviews.

The *stakeholder interviews* will next be used to gain a deeper understanding of not the facts as described in the documents, but the experiences and perspectives of the stakeholders. This should help to, later in the research, prioritize governance mechanisms and identify relations between governance mechanisms and other aspects of the experiments. In order to gain different perspectives, different experiment stakeholders should be interviewed. The most relevant interviewee are the policymakers and professionals that designed and guided the experiment, as they will have ample knowledge of the entire experiment. In order to contradict or confirm the views of the policymakers, however, experiment participants should also be interviewed. As their participation in transition experiments is considered to be a key factor in the success of these experiments, their perspectives and experiences can be seen as a good indicator for the functioning of different governance mechanisms.

In addition to document analysis and stakeholder interviews, real-world observations and physical artefact analysis are often described as data sources for case studies. However, as the experiments have already ended any physical artefacts have been destroyed, these data sources were deemed irrelevant to the case study research.

B.3.2. Data needs

Now that the different data sources have been identified, these sources should be linked to the data that is needed for the case studies. For each of the features and their operationalized criteria, the type of information needed and the questions that need to be answered to generate that information have been defined. Be aware that these questions are not equal to the questions that will be asked during the interviews. The questions described below are what Yin (2014) defines as 'Level 2 Questions' and are aimed generating data about the case. So-

called Level 1 Questions to be answered by individual stakeholders during the interviews will be described in chapter 4 of this case study protocol.

B.3.2.1. Data needs input features

Awareness	
Questions	Did the experiment include stakeholders that were aware of the need for flood risk management? What was the impact of the level of awareness?
Data sources	Background documents: information on exposure to flooding and previous mitigation measures Interviews: information on awareness as a driver for collaboration

Commitment	
Questions	To what extent did the experiment include stakeholders that were committed to participating?
Data sources	Background documents: number of moments for participation Notes of meetings: names of participants, times they participated Or, if these notes are not accessible or do not exist: Interviews: estimate of average participation

Trust	
Questions	Did the experiment include stakeholders that trusted each other? What was the impact of the level of trust?
Data sources	Background documents: general democratic environment, relationship citizens and municipality Interviews: expressed trust stakeholders, impact trust as a driver for collaboration

Social Capital	
Questions	To what extent did the experiment include social capital? What was the impact of the inclusion of social capital?
Data sources	Websites: pre-existing networks Background documents: pre-existing networks and impact on collaboration Interviews: pre-existing networks and impact on collaboration

Expertise	
Questions	To what extent did the experiment include expert input? What type of expert input was included? What was the impact of expert input inclusion?
Data sources	Background document: what type of expert input Interviews: impact expert input on collaboration

Support	
Questions	To what extent did the experiment include support? What was the impact of this support?
Data sources	Background documents: quantity support Interviews: impact type and quantity of support on collaboration and implementation

Institutional alignment	
Questions	Did the experiment encounter any institutional barriers? What was the impact of these barriers?
Data sources	Background documents: barriers experienced Interviews: barriers experienced, impact of barriers on collaboration and implementation

Leadership	
Questions	Did the experiment include a leader? What was the impact of this leader?
Data sources	Background document: entity identified as leader Interviews: experience with leader entity, impact of leader on collaboration process

Table B.1: Data needs input features

B.3.2.2. Data needs activities features

Transparency	
Questions	Did the experiment include transparent information-sharing? What was the impact of the level of transparency
Data sources	Interviews: satisfaction of stakeholders with transparency, impact transparency on collaboration

Collaboration	
Questions	To what extent did the experiment include mechanisms for collaboration? What different mechanisms did the experiment include? To what extent were decisions made based on stakeholder agreement? Did the experiment include a variety of stakeholders? Did the experiment include less privileged stakeholders in the decision-making? Did the stakeholders accept the collaboration mechanisms?
Data sources	Background documents: occurrence and type of collaboration mechanisms Background documents: type of collaboration mechanisms Background document: number and type of decisions made per collaboration mechanism Background documents: stakeholders included in the making of each decision Background documents: type of stakeholders included in the making of each decision Interviews: type of stakeholders included in the making of each decision Stakeholder interviews: level of satisfaction with collaboration mechanisms
Reflexivity and learning	
Questions	Did the experiment include learning goals? What was the impact of these learning goals? Did the experiment include mechanisms for learning? What was the impact of these mechanisms
Data sources	Background documents: learning goals, role learning goals in experiment Interviews: learning goals, role learning goals in experiment Background documents: learning mechanisms, role mechanisms in experiment Interviews: learning mechanisms, role mechanisms in experiment
Sequence of actions	
Questions	How thoroughly was the experiment planned? How did the planning document impact the process of collaboration?
Data sources	Background documents: planning document Interviews: impact planning document on process of collaboration
Methodology	
Questions	How thoroughly was the experiment designed? How did the design impact the process of collaboration?
Data sources	Background documents: methods Interviews: impact design on process of collaboration

Table B.2: Data needs activities features

B.3.2.3. Data needs output features

Built capacities	
Questions	Did the experiment empower stakeholders? How did it achieve this?
Data sources	Interviews: sense of empowerment stakeholders, actions that led to this empowerment
Accountability	
Questions	Did the experiment foster a sense of problem ownership? How did it achieve this?
Data sources	Interviews: sense of problem ownership stakeholders, actions that led to this sense of problem ownership
Social change	
Questions	Did the experiment lead to the formation of new networks? Did the experiment lead to changed stakeholder perceptions about sustainability issues? How did it achieve this?
Data sources	Background documents: type of networks Interviews: type of networks Interviews: stakeholder perceptions, actions that led to these changed perceptions
Physical change	
Questions	Did the experiment lead to the implementation of a NBS? How did it achieve this?
Data sources	Background documents: type of NBS implemented, motivations for NBS Interviews: motivations for NBS
Actionable knowledge	
Questions	Did the experiment lead to the formulation of transition pathways? How did it achieve this?
Data sources	Background documents: formulation of transition pathways, processes of formulation

Transferability	
Questions	Were generalized lessons for transferring formulated?
Data sources	Background document: lessons learned
Scalability	
Questions	Were generalized lessons for upscaling formulated?
Data sources	Background document: lessons learned
Accounting for unintended consequences	
Questions	Were externalities identified?
Data sources	Background document: formalized identified externalities Interviews: non-formalized externalities

Table B.3: Data needs output features

B.3.2.4. Data needs outcome features

Socio-ecological integrity	
Questions	Did the experiment lead to a higher acceptance of NBS? What caused this higher acceptance?
Data sources	Background documents: plans for further implementation Interviews: plans for further implementation, why are they planned
Intergenerational equity	
Questions	Did the experiment lead to inclusion of less privileged stakeholder groups? Why did or didn't it lead to the inclusion of less privileged stakeholder groups?
Data sources	Background documents: plans for involvement less privileged stakeholder groups Interviews: plans for involvement less privileged stakeholder groups, events in experiment that inspired this inclusion
Socio-ecological stewardship and democratic governance	
Questions	Did the experiment lead to adjusted actions or behavior? What inspired these actions?
Data sources	Background documents: follow-up actions planned or undertaken, why are they planned or undertaken? Interviews: follow-up actions planned or undertaken, why are they planned or undertaken?
Livelihood opportunity	
Questions	Does the experiment ensure livelihood opportunity? How does it ensure this?
Data sources	Background documents: positive impact Interviews: positive impact, how is it ensured
Resource efficiency	
Questions	Was the climate impact of the materials used in the experiment considered?
Data sources	Background document: materials used Interview: materials used, why where these materials used
Precaution and adaption	
Questions	Did the experiment inform adapted flood risk mitigation strategies? How did it inform these strategies?
Data sources	Background document: strategies, how are they informed by the experiment Interviews: strategies, how are they informed by the experiment

Table B.4: Data needs outcome features

B.3.3. Data collection plan

The first source of information is Dr. Niki Frantzeskaki, one of the researchers involved in the RESLIENT EUROPE project. She can provide documentation relating to the project and experiments as well as the names and contact information of stakeholders that can possibly contribute to this research. Additionally, participation in the Resilience Matters Event in Rotterdam in March 2018 should allow for meeting stakeholders and approaching them for interviews. Email correspondence will then be used to plan interviews and discuss

possibilities for expanded document analysis. In order to ensure more, often busy, stakeholders will agree to being interviewed, the interviews are planned to take no longer than 1 hour (Weiss, 1994). Phone or Skype interviews are to be proposed to interviewees to allow them to select the medium they feel most comfortable with. In order to also gain insights from a group that is unlikely to (be able to) agree to an interview, namely the citizens from Burgas, a paper questionnaire will be designed and the interviewees from Burgas will be asked to distribute this questionnaire. All documents, artefacts and interview transcripts will be stored in an online cloud service of which a link can be provided if other researchers wish to have access to the data.

After the interviews have been planned, some preparations need to be made. In order to prioritize which data to collect, a quick analysis of the experiments documents should provide the main themes and gaps to be addressed through stakeholder interviews or through document requests. Next, interview questions and protocols will need to be designed to elicit the right information from the right interviewee. Ideally, pilot interviews would be held to refine the interview questions and techniques as well as ensure all relevant data can be gathered within the timeframe (Brenner, 1985). Given the stakeholders' limited availability to participate in this research, however, scheduling such interviews purely for testing the interview protocol is deemed unfeasible. Instead, two different approaches will be used to refine the interview protocol. First, experts will be asked to review the interview protocol and propose changes or improvements. Secondly, after each interview the interview protocol will be revisited and lessons learned during the interview will be applied to improve the protocol. In order to ensure the repeatability of the research, these changes will be documented and all versions of the interview protocol that have been used for data collection will be accessible on request.

B.4. Interview questions

The interview method that will be used to conduct the stakeholder interviews is the semi-structured method. After all, the topics of the interviews have already been predefined: first by the evaluation framework and second by the document analysis. There is some broad information available about the direction of the answers, but it is believed that open questions will provide deeper insights and will allow for more variation and unexpected answers than structured research methods. As a result, this type of research method is commonly used for evaluation research (Baarda, Van der Hulst, & De Goede, 2012).

A number of different interview protocols will have to be designed. One difference exists between the two experiments included in the case study research. As these experiments had different designs and challenges, as well as differences between the information available in the documents, different information gaps and lines of investigation will have to be addressed. Secondly, interview protocols will differ between stakeholders as they have different levels of knowledge and experiences. Overall, the interview respondents could be divided into two groups: process managers, who set up and guided the process of experimentation, and participants, who were involved in the design and implementation of the experiment. The former consists of policy makers and process counselors, whereas the latter consists of the citizens. By differentiating the interview protocols between the experiments and stakeholders, the information most relevant to the case can be collected.

The different interview protocols are included in Appendix C.

B.5. Guidelines for reporting

As identified before, the case study report will have two different audiences, both of them, however, academic. Both the master thesis report as the scientific paper will therefore follow the linear-analytic structure, including a descriptions of the problem introduction, theory

research and methodologies (Runeson & Höst, 2009). For the repeatability of the research, and for the reviewers to assess the quality of the report, it is therefore important that statements made can be supported, both in the case description and in the other parts of the report, with the relevant data sources. Special care will be taken when reporting on the outcomes of the case studies. Interview quotes should be used to not only support statements, but also add some flair to the reporting (Weiss, 1994). In order to protect the privacy of the interviewees, however, care should be taken to ensure that the quote is as close to the interview excerpt as possible, without providing clear insights into the speaker's identity. In addition to providing some background information and setting the context, which will help readers understand and interpret case observations, it is deemed relevant to clearly link the case observations to the evaluation scheme. This should allow readers to quickly discern how the case relates to the evaluation scheme.

Appendix C: Interview Protocols

C.1. Interview Protocols Burgas
C.1.1. Visual representation interview focus
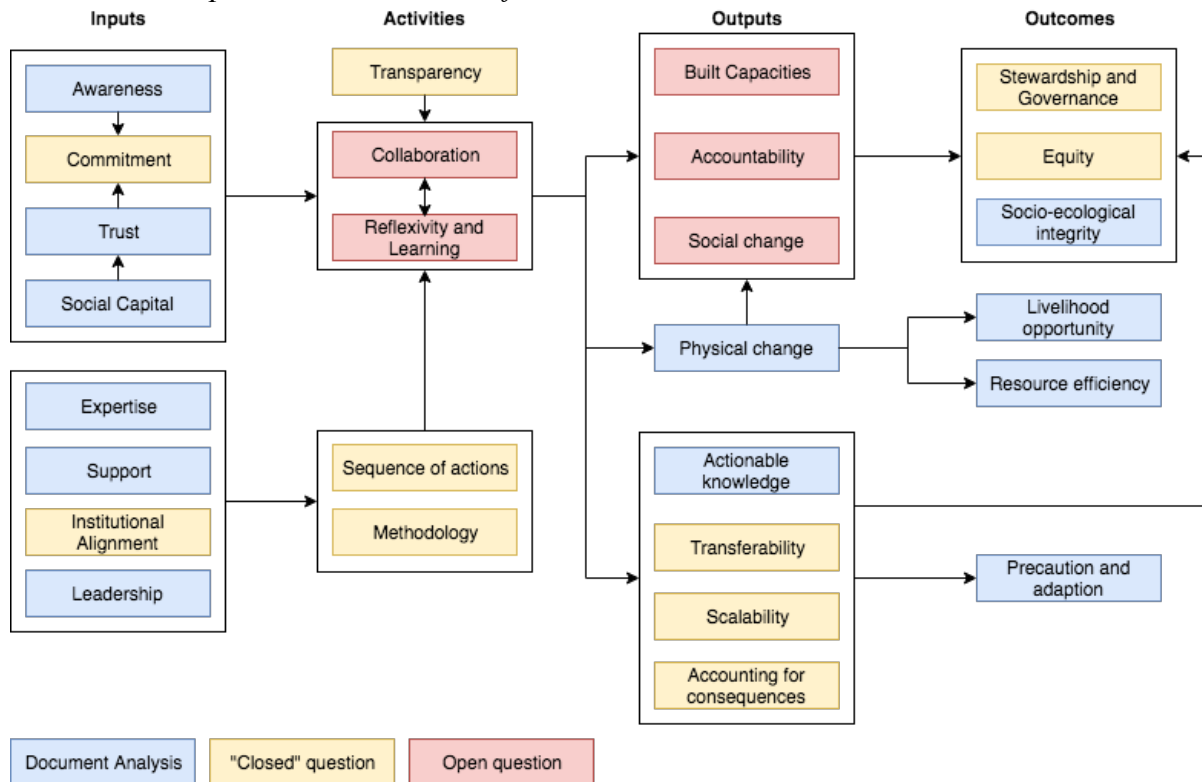
Figure C.1: Interview focus Burgas

The interview focus is based on a quick scan of the available documents. Features that have been colored in blue are sufficiently explained in the documents. Those indicated in orange are described in the documents, but require some additional questions to make sure I interpreted the information correctly or to fill in some gaps. Those indicated in red require the most attention and will be the focus of the interviews.

As for Antwerp, the information needed from the policymakers and citizens is not the same. The information from the citizens will be obtained making use of a survey and the depth and breadth of their responses is therefore limited. The policymakers will therefore be asked about the feedback they received from the citizens, so they can answer some of the questions for the citizens.

Topic: Outcomes

Feature	Subtopic
Built capacities	Experience Confidence
Social change	Perspective on co-creation Perspective on NBS
Stewardship	Citizen plans
Transferability	Formalized lessons
Scalability	Formalized lessons

Topic: process of collaboration

Feature	Subtopic
Commitment	Commitment ULG Commitment 'Willow Experiment'
Collaboration	Decision-making ULG and 'Willow Experiment' Number and type of collaboration 'Willow Experiment' Inclusion less privileged groups Stakeholder satisfaction
Equity	Less privileged groups future inclusion

Topic: planning and resolving issues

Feature	Subtopic
Sequence	Planning
Method	Methods for collaboration
Refl & Learning	Learning mechanisms Changes made to the experiment
Consequences	Negative impact Mitigation measures
Inst. Alignment	Institutional Barriers

C.1.3. Questions for Policymakers

Introduction
First off, I would like to start by thanking you for agreeing to this interview. I will first remind you about the focus of this interview and then I will briefly explain some things about the interview process. As I explained in my emails, I am a master student from the TU Delft writing my thesis on the evaluation of sustainability experiments. The goal of my research is to develop an evaluation method to capture the lessons learned from these experiments. I have selected the Burgas experiment to test the evaluation method, which is why I would like to learn from your experiences. As I described in the email, this conversation is completely confidential. I will record this conversation in order to transcribe the interview. After I have done that, I will delete the recordings. I might refer to this interview in my report, but will make sure you will remain anonymous. During the interview, I might take notes, which will serve as a reminder for myself to make sure the interview stays on track. The interview will last for about one hour. Do you have any questions so far?

"Warm-up" Questions

1. Can you tell me about your role in the experiment?
2. How did you experience participating in the experiment? *   *If possible: specified to any markers mentioned in the first question*

Topic 1: Outcomes
3. Do you feel you gained any experiences from the experiment?
Probe for: Concrete skills or built capacities   Why do they feel they gained that skill?   What helped them to gain that skill (causality)?   Do they feel confident in using that skill?   Are they planning on applying that skill?   Did they observe similar changes in citizens?
Likely, Q3 will already provide some markers about changed perspectives that can be used as an opening for Q4.   4. You mentioned [marker], did the experiment inspire you to look differently at climate adaptation policy?
Probe for: Concrete changed perspectives   What helped them to change their perspective (causality)?   Are they planning on applying their knowledge?   Is this knowledge formalized?   Did they observe similar changes in citizens?

Topic 2: Process of collaboration
Likely, Q3/4 will already provide some markers about changed perspectives that can be used as an opening for Q5.   5. You mentioned [marker], can you tell me a bit more about the process of collaboration?
Probe for: Commitment (ULG and Willow Experiment)   Collaboration Willow Experiment   Impact collaboration (causality)
6. You talked about [marker], can you tell me a bit more about the involvement of the citizens?
Probe for: Stakeholder satisfaction   Inclusion less privileged groups   Future inclusion less privileged groups   Impact stakeholder involvement on the experiment

Topic 3: Planning and resolving issues
7. Can you tell me a bit more about how you planned the process of experimentation?
Probe for: Sequence   Methods   Learning mechanisms   Impact on process
8. Were there moments where you had to deviate from the existing planning? Can you tell me a bit more about those?

Probe for:
Concrete examples
Forecasting
Mitigation measures
Institutional barriers

Conclusion
We have reached the end of the interview. Before I will ask you some practical questions, I would like to ask:
2.   is there anything you missed or would like to add?
3.   How did you experience participating in this interview?
[Thank interviewee for feedback]
If there is anything you would like to say or ask, you are welcome to send me an email. I will be using this interview to inform my analysis of the evaluation method and the experiment. I plan to have the final report finished before August.
4.   Would you like me to send you a copy of the finished report by then?
[Repeat agreement]
I would like to thank you again for agreeing to participate in this interview and would like to wish all the best with increasing the resilience in Burgas!

C.2. Interview Protocols Antwerp
C.2.1. Visual representation interview focus
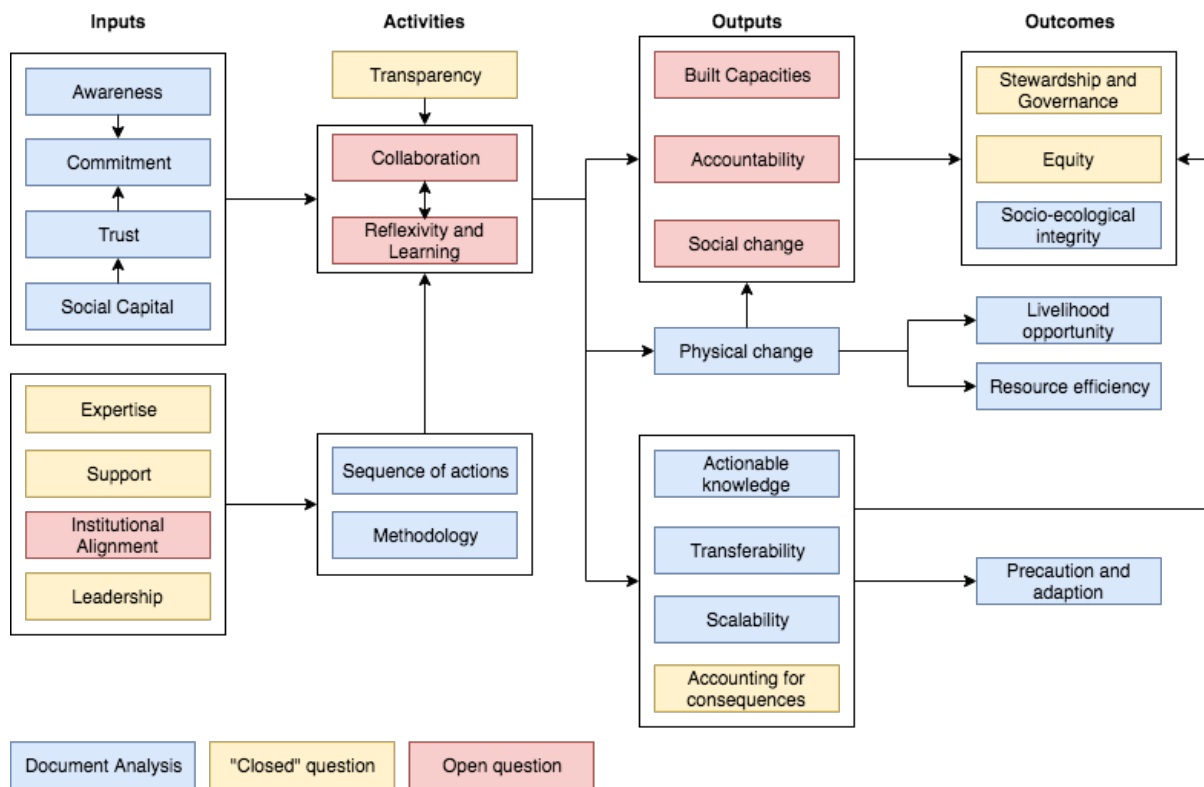
Figure C.2: Interview focus Antwerp

The interview focus is based on a quick scan of the available documents. Features that have been colored in blue are sufficiently explained in the documents. Those indicated in orange are described in the documents, but require some additional questions to make sure I interpreted the information correctly or to fill in some gaps. Those indicated in red require the most attention and will be the focus of the interviews.

The focus of the interviews will also depend on the interviewee: policymakers are asked about the experiment from the perspective of the municipality, citizens from the perspective of a participant.

Topic: outcomes

Feature	Subtopic
Built capacities	Experience
	Confidence
Social change	Perspective on NBS
	Perspective on co-creation

Topic: process of collaboration

Feature	Subtopic
Expertise	Expertise per task
Collaboration	Stakeholders
	Less privileged
Equity	Future involvement less privileged

Topic: resolving issues

Feature	Subtopic
Reflexivity & L.	Evaluation methods
	Process changes during process
	Outcome changes during process
Inst. Alignment	Institutional barriers
Consequences	Negative impact
	Mitigation measures

C.2.3. Questions for Policymakers

Introduction
First off, I would like to start by thanking you for agreeing to this interview. I will first remind you about the focus of this interview and then I will briefly explain some things about the interview process. As I explained in my emails, I am a master student from the TU Delft writing my thesis on the evaluation of sustainability experiments. The goal of my research is to develop an evaluation method to capture the lessons learned from these experiments. I have selected the 'Groene Ader' experiment to test the evaluation method, which is why I would like to learn from your experiences. As I described in the email, this conversation is completely confidential. I will record this conversation in order to transcribe the interview. After I have done that, I will delete the recordings. I might refer to this interview in my report, but will make sure you will remain anonymous. During the interview, I might take notes, which will serve as a reminder for myself to make sure the interview stays on track. The interview will last for about one hour. Do you have any questions so far?

"Warm-up" Questions
1. Can you tell me about your role in the experiment?
2. How did you experience participating in the experiment? *
If possible: specified to any markers mentioned in the first question

Topic 1: Outcomes
3. Do you feel you gained any experiences from the experiment?

Probe for:
Concrete skills or built capacities
Why do they feel they gained that skill?
What helped them to gain that skill (causality)?
Do they feel confident in using that skill?
Are they planning on applying that skill?

Likely, Q3 will already provide some markers about changed perspectives that can be used as an opening for Q4.
4. You mentioned [marker], did the experiment inspire you to look differently at climate adaptation policy?

Probe for:
Concrete changed perspectives
What helped them to change their perspective (causality)?
Are they planning on applying their knowledge?

Topic 2: Process of collaboration
Likely, Q3 and Q4 will already provide some markers about the process of collaboration that can be used as an opening for Q5.
5. A few minutes ago, you talked about [marker]. Can you tell me a bit more about the external experts that were involved in the experiment?

Probe for:
Roles of experts in experiment
Judgement of interviewee of experts involved/Impact of experts involved

6. You mentioned [marker] that helped to connect different stakeholders. Do you feel that the stakeholders in the experiment were representative of the Sint-Andries neighborhood?

Probe for:
Concrete examples of stakeholder groups involved
Involvement of less privileged stakeholder groups (social housing, elderly)
How stakeholder groups were involved
Impact of stakeholder involvement on experiment
Future plans for (less privileged) stakeholder involvement

Topic 3: Resolving Issues
Likely, evaluation activities have already been mentioned in Q3-6 that can be used as an opener for Q7
7. You also mentioned [marker], can you tell me a bit more about these activities?

Probe for:
Concrete examples of learning/evaluation activities
Forecasting activities
Outcomes of these activities
Changes made based on non-formalized learning
Forecasting activities

8. We talked about changing the experimentation process because of new insights, were there also instances where the process had to be adjusted because of unexpected barriers?

Probe for:
Role municipality
Concrete examples
Expected barriers

Conclusion
We have reached the end of the interview. Before I will ask you some practical questions, I would like to ask: is there anything you missed or would like to add?   How did you experience participating in this interview?   [Thank interviewee for feedback]   If there is anything you would like to say or ask, you are welcome to send me an email. I will be using this interview to inform my analysis of the evaluation method and the experiment. I plan to have the final report finished before August.   Would you like me to send you a copy of the finished report by then?   [Repeat agreement]   I would like to thank you again for agreeing to participate in this interview and would like to wish all the best with the continuation of the 'Groene Ader' project.

Topic: outcomes

Feature	Information needed
Built capacities	Experience
	Confidence
Accountability	Feeling of responsibility
Social change	Perspective on co-creation
	Perspective on NBS
	Perspective on own role
Stew & Gov	Plan to stay involved
	Plan to change individual behavior

Topic: process of collaboration

Feature	Information needed
Leadership	Steering group leadership
Collaboration	Frequency and form of meetings
	Satisfaction steering group meetings
	Satisfaction overall process
	Stakeholder inclusion
Equity	Plan to involve more stakeholders

Topic: relationship with municipality

Feature	Information needed
Support	Coach
	Municipal contact person
Inst. Alignment	Resistance
	Access
Transparency	Communication stadslab2050
	Communication municipality

C.2.5. Questions for Citizens

Introduction
First off, I would like to start by thanking you for agreeing to this interview. I will first remind you about the focus of this interview and then I will briefly explain some things about the interview process. As I explained in my emails, I am a master student from the TU Delft writing my thesis on the evaluation of sustainability experiments. The goal of my research is to develop an evaluation method to capture the lessons learned from these experiments. I have selected the 'Groene Ader' experiment to test the evaluation method, which is why I would like to learn from your experiences. As I described in the email, this conversation is completely confidential. I will record this conversation in order to transcribe the interview. After I have done that, I will delete the recordings. I might refer to this interview in my report, but will make sure you will remain anonymous. During the interview, I might take notes, which will serve as a reminder for myself to make sure the interview stays on track. The interview will last for about one hour. Do you have any questions so far?

"Warm-up" Questions
1. Can you tell me about your role in the experiment?

2. How did you experience participating in the experiment? * *If possible: specified to any markers mentioned in the first question*

Topic 1: Outcomes
3. Do you feel you gained any experiences from the experiment?
Probe for: Concrete skills or built capacities Why do they feel they gained that skill? What helped them to gain that skill (causality)? Do they feel confident in using that skill?
Likely, Q3 will already provide some markers about changed perspectives that can be used as an opening for Q4.   4. You mentioned [marker], did the experiment inspire you to look differently at climate adaptation?
Probe for: Concrete changed perspectives What helped them to change their perspective (causality)? Do they feel more responsible for mitigating flood risk?
Likely, Q3 and Q4 will already provide some markers about changed capacities and perspectives that can be used as an opening for Q4.   5. You mentioned [marker], do you plan on staying involved with climate adaptation?
Probe for: Plans to stay involved with public adaptation measures Plans to take individual measures Why they plan on staying involved (causality)

Topic 2: Collaboration
6. Can you tell me a bit more about the process of collaboration for the 'Groene Ader' experiment?
Probe for: Leadership Meetings and decision-making Satisfaction with meetings and decision-making Satisfaction with overall process Stakeholder inclusion Impact on the experiment

Topic 3: relationship with municipality
7. How did you experience the collaboration with the municipality?
Probe for: Relationship with coach and contact person Communication with stadslab2050 Access to resources Barriers Impact experiment

Conclusion
We have reached the end of the interview. Before I will ask you some practical questions, I would like to ask: is there anything you missed or would like to add? How did you experience participating in this interview? [Thank interviewee for feedback] If there is anything you would like to say or ask, you are welcome to send me an email. I will be using this interview to inform my analysis of the evaluation method and the experiment. I plan to have the final report finished before August. Would you like me to send you a copy of the finished report by then? [Repeat agreement] I would like to thank you again for agreeing to participate in this interview and would like to wish all the best with the continuation of the 'Groene Ader' project.

Appendix D: Survey citizens Burgas

Thank you for agreeing to take part in this survey about your experiences participating in the "Resilient Europe" project. Your answers will be part of a research into the "Resilient Europe" projects in different European cities and will help improve future projects. This survey should take about 10 minutes. Although you are asked to fill out some personal information, be assured that all answers you provide will be kept in the strictest confidentiality.

If you have any questions, please feel free to send an email to:
Laura van Buggenum: L.J.vanBuggenum@student.tudelft.nl

1. How did you participate in the "Resilient Europe" project?

2. How did you experience the communication with the Burgas municipality about the "Resilient Europe" project?

3. How did you experience the collaboration with the Burgas municipality on the "Resilient Europe" project?

4. What is your opinion on intensive collaboration between citizens and the municipality like in the "Resilient Europe" project?

5. Would you like to participate in future collaborations like the "Resilient Europe" project? Please explain your answer.

Appendix E: Evaluation Burgas

E.1. Input Features

E.1.1. Awareness

Between 2010 and 2017, the Dolno Ezerovo neighborhood experienced 5 major floods, each time causing significant damage to properties and infrastructure. The last severe flood in October 2017 caused 5 human deaths in different areas of the Burgas municipality. As a result, residents of the Dolno Ezerovo neighborhood are highly aware of the local flood risks and the need to take action. The civil servants interviewed attributed the high level of public participation in the experiment largely to this awareness. The municipality further enhances this awareness by employing a publicly accessible mobile application and website to give live updates of the neighborhood's water level surveillance every 15 minutes. Despite this easily accessible information, citizens in the neighborhood are ill-prepared to deal with future floods. They rely heavily on the government for protection and are not aware of bottom-up individual or communal actions they could take to reduce the risk of flooding in the area. As lamented in one of the municipality's policy documents: "The community fails to see itself as a powerful driving force that is able to achieve significant positive results with small efforts". As a result, awareness at the beginning of the experiment is defined as medium: stakeholders are aware of flood risks, but not of their own role in flood risk management. When looking at awareness as a driver for stakeholder commitment, however, it can be observed that an awareness of flood risk not the only driving force was for citizens to participate in the experiment. As the previous floods have damaged much of the local green spaces, citizens have very few local, public spaces they can go to meet up, play and relax. As the citizens do not know how to change this, participating in this experiment with a nature-based solution gave them a chance to address this issue.

E.1.2. Commitment

The experiment in Burgas was designed in two separate phases, each with a different actor constellation and goal. The first phase consisted of meetings to brainstorm and formulate an action plan to increase the neighborhood's resilience. The participants were all part of an existing consultative council and represented a broad range of public and private actors as well as some representatives from the neighborhood. Commitment was high and most participants attended all relevant meetings. The second phase of the experiment was the implementation phase, which was as much about generating awareness of the importance of citizen involvement in managing flood risks as it was about mobilizing people for the first action to be implemented. In this phase, part of the participants from the first phase, in particular the civil servants and citizens, collaborated with residents of the neighborhood who had not yet been involved in the experiment to reach the aforementioned goals. As only a core group of stakeholders participated in both phases of the experiment with other stakeholders only being active within one phase of the experiment, the commitment is defined as medium.

E.1.3. Trust

There was a rather large number of different stakeholders and stakeholder constellations involved within the different phases of the experiment. Especially within existing actor groups, such as the consultative council or the association of retired persons, relations had already been established and the level of trust between the different stakeholders was high. The innovative, or experimental, form of collaboration in this case, however, was the collaboration between the municipality and the citizens of Dolno Ezerovo. There was a big gap between these two parties at the beginning of the experiment. Although public consultation is mandatory for each project undertaken by the Burgas municipality, public participation can be quite low and the

municipality does not always succeed in translating the input from the citizens into solutions for their problems. As a result, citizens tend to believe that the municipality does not care about them and their needs. Additionally, citizens are used to top-down modes of government, so any initiatives encouraging bottom-up approaches to solve problems are met with resistance. Similarly, the municipality focuses on top-down solutions as well and there is limited experience with sharing more responsibilities with the citizens. As a result of the attitude of both parties, trust at the beginning of the project can be defined as low: stakeholders are not willing to rely on other participants' judgement and capacities.

E.1.4. Social capital
During the course of the experiment, the municipality utilized a number of existing networks to reach specific goals. In the first phase of the experiment, an existing municipal consultative council was asked to participate in the 5 sessions elaborating on the action plan and desired implementation. As the municipality had already established a good relationship with these stakeholders, convincing them to participate in the meetings was relatively easy and open discussion was possible. In the second phase, two more existing networks were mobilized. The key figures within the neighborhood's association of retired persons were asked to help the civil servants to talk to each other and the local residents in order to inform them about resilience and mobilize them for the implementation of the action plan. In a similar fashion, teachers from the local primary school were asked to share the message of urban resilience with their students, who, in turn, could share it with their parents. As actors within both networks were trusted by the neighborhood, they were instrumental in mobilizing more support for the action plan.

In the second phase, however, there was another important actor that acted as a mediator between the municipality and the networks: the head of the local center for administrative services. Her knowledge of the neighborhood was credited by the civil servants as being indispensable for completing the project. As one of the interviewed civil servants describes: "She knows how to talk to the people and how to involve them. It was really important to have her join the project as a mediator, because the people have confidence in her." Later in the interview, the civil servant continues: "We [the civil servants] just sit inside our buildings and we do not know what is going on outside. Therefore, it is very important for us to rely on mediators to make that connection between us and the local people." Another dimension of social capital was also credited during the interviews: the residents' knowledge of the neighborhood. The use of trees to mitigate flood risks, for example, was quite innovative for the municipality, but the plan was immediately met with approval by the citizens. Those who had been living in the neighborhood long enough to witness the local deforestation 20 years earlier, had already experienced firsthand that a lack of trees and greenery increased the number of minor and major floods in the area. This knowledge of the local situation could be utilized better in future projects and experiments.

E.1.5. Expertise
In addition to making use of social capital, a number of more formal experts were included in the experiment in order to guide the process. Each of these experts was included in the first phase of the experiment. Before the official start of the experiment, the project leaders of the Resilient Europe project spent 2 days in the city to provide a workshop on the topics of urban resilience and transition management and help refine the plans for the neighborhood. Their perspective was especially interesting for the municipality, as their participation is described in one of the interviews: "They [the project leaders] provide a different point of view and provide information that we, as the municipality, do not have access to." The interviewee refers here specifically to the academic knowledge of the project leaders that provided a framework for the entire experiment. The five discussion sessions were set-up and moderated by an expert from

within the municipality with ample experience with both the topic and consultative meetings. She helped guide the collaboration and helped define a specific action plan for the neighborhood. The last expert active within the first phase of the experiment was an external consultancy company that collected all of the input from the meetings and the public consultation and converted it into a coherent action plan for the municipality to use during the implementation phase.

E.1.6. Support
Next to making use of human capital, some other forms of support helped design and execute the experiment. The most important of which is the financial support of €45.500,-. Of this sum, 85% percent was funded by the EU and the other 15% was funded by the municipality. That funding is an important enabler is clearly illustrated by the planned further implementation of the action plan, which has come to a temporary halt because new financing opportunities have not been defined yet. In addition to funding, a risk assessment performed before the start of the experiment has been instrumental in communicating the challenges within the municipality and identifying the neighborhood and topic to focus on. It has been an important tool in specifying the actions to be undertaken by the municipality and its citizens.

E.1.7. Institutional Alignment
From an institutional perspective, the experiment experienced hardly any challenges from the institutional environment. The experiment did not encounter any issues with permits, laws or regulations and institutional alignment is therefore defined as high. The civil servants did admit, however, that much of this alignment had to do with the fact that the solution was both small-scale and designed with the local regulations in mind. They acknowledged that large-scale or bottom-up driven implementation of similar solutions could cause problems in the future. In addition to not being hindered by laws and regulations, the experiment enjoyed broad support within the municipality. Although such intensive collaboration with citizens is highly uncommon in Burgas, stakeholders within the municipality ware excited to try this new approach. Additionally, the goals of the experiment aligned nicely with the municipality's ongoing commitment to increasing the quality of life in the city. The presence of the deputy mayor of the European programs and policy department at the planting of the trees was a symbol of this municipal support of the experiment that was highly appreciated by the citizens. It is recounted by one of the participants: "[The presence of the deputy mayor] showed the attendants that an important person is interested in the event and their problems. It shows that he is willing to help them, to actually go outside and interact with them."

E.1.8. Leadership
As the experiment was still very much organized top-down, the Burgas municipality could be identified as the leading entity. A small number of civil servants were responsible for the entire experiment, from obtaining funding to organizing meetings, contacting stakeholders and providing the resources for the implementation of the solution. The role of the leaders, in this case, was to coordinate the different aspects of the experiment in order to reach a common goal. Without their guidance and efforts, the experiment would have never been completed. This is clearly visible, again, when looking at the continuation of the experiment which is still depends on the efforts of these civil servants.

E.2. Activity Features
E.2.1. Transparency

As all experiment participants were satisfied with the communication, transparency is defined as: high. This contributed to building trust between the experiment participants and increased the participants' satisfaction with the collaboration process.

E.2.2. Collaboration
Mechanisms for collaboration
In total, 6 meetings were organized to openly brainstorm about and discuss the topic of urban resilience with the all of the stakeholders under the supervision of the discussion leader. As a result of the first two meetings, the stakeholders reached a consensus about the contours of the action plan and the desired first step to design the plan. Having defined a solution to be implemented, a number of small or meetings were set up with local key stakeholders to inform them about the action plan and ask for their help in sharing this information with the other citizens. Consequently, the involved civil servants and local key stakeholders went door-to-door to discuss the action plan and solution with the neighborhoods' residents and to learn about their needs and wishes within this context. This was a passive form of collaboration, aimed at provoking citizens to think about their living environment and encouraging them to become more active. The teachers of the local primary school educated the students on urban resilience and flood risk management in order to change their perspectives and encourage them to participate in the experiment. These local efforts resulted in a neighborhood meeting where, under the supervision of the head of the local center for administrative services, the citizens reached a consensus on where to implement the solution. An event was organized around the implementation of the solution in order to draw more attention to the solution and the message of resilience it represented. It was meant to show citizens that they too could contribute to flood resilience through small solutions. Additionally, the event drew a large number of interested citizens and the civil servants made use of the informal atmosphere to, yet again, talk to the citizens to gather input and encourage them to become involved with the further implementation of the action plan. After the experiment, a meeting was organized to assess the experiment and adapt the action plan accordingly. After this meeting, an additional 2 meetings were organized in order to refine the action plan and a final meeting was organized to present the final action plan and inspire the stakeholders to continue to work on resilience in the neighborhood.

In addition to the collaboration within the experiment, the civil servants also collaborated with other cities within the Resilient Europe network. In total, they visited 4 sessions of two days where the civil servants from different cities came together to discuss their experiments and learn from each other. This collaboration was highly appreciated by them, as it provided them with inspiration for the experiment in Burgas. One of the civil servants describes: "[Learning from the other cities] helped us save a lot of time. If we saw something that worked elsewhere, we could apply it here without repeating their mistakes."

Decisions
Three main decision were taken through these mechanisms for collaboration. The first major decision was which neighborhood and related issue the action plan should be designed for, which was decided on during the discussion sessions. The action plan was then further designed during these meetings. The second major decision was also made during these discussion sessions and related to the type of solution that was to be implemented in the neighborhood as a starting point for further implementation of the action plan. This decision also included a focus on how this solution was to be implemented. The last major decision was taken by the residents of the neighborhood and detailed the location where the solution was to be implemented. Although each of these decisions was based on a consensus between the stakeholders, the residents of the neighborhood only had a major influence on the location of the solution. The first two decisions were based on a consensus between the consultative council

that did include residents, but was largely made up of stakeholders that did not represent the neighborhood.

Stakeholder representation
For the first meeting, before it was decided to focus the action plan on Dolno Ezerovo, a large number of stakeholders was included in the discussion. These consisted of stakeholders representing different municipal departments, safety authorities, businesses and industries, NGOs and citizens. After the decision was made to focus on flood issues in Dolno Ezerovo, only the parties with an interest in either the neighborhood or flood issues remained. They represented different municipal departments, local safety authorities and flood and safety related NGO's. In addition, citizens from the neighborhood were invited to participate as well. For the experimentation phase, the stakeholder representation shifted more towards the residents and the local and municipal authorities, although local business did stay involved, for example by providing food at the tree planting event.

Within the group of citizens, the Roma community was identified beforehand as a less privileged group that needed to be involved in the experiment. Although the interviewees could not provide any quantitative estimates of the involvement of this stakeholder group, there was a consensus that the entire neighborhood was well represented in the second phase of the experiment. Stakeholder inclusion is therefore defined as high: the experiment included a mixed group of stakeholders, including less privileged stakeholder groups.

In terms of stakeholder satisfaction, the experiment participants reported to be highly satisfied with the process of collaboration

The civil servants attributed the success of the process of collaboration mainly to the small size of the area involved in the experiment, which made it easier to reach the citizens, as well as the willingness of local key figures to cooperate and act as frontrunners. Throughout the interviews with the civil servants, these themes reoccurred when discussing possibilities for repeating the experiment in other areas of the city.

E.2.3. Reflexivity and Learning
The experiments itself took place within the timespan of about one year. The meetings had been designed beforehand, whereas the idea of the implementation experiment evolved as a result of the first few meetings. In both instances, outputs or learning goals had been identified. For the meetings, the main goal was to come to an action plan deemed robust by all stakeholders. Reaching a consensus on the action plan, although not explicitly described as such in the policy documents, can be interpreted as a form of social learning. In the experimentation phase, the goal was to implement a solution that met a couple of objectives in order to refine the action plan. Although not identified explicitly as such by the municipality, this goal can also be seen as a learning goal. As a result of the presence of goals that are not directly recognized as learning goals, the formulation of learning goals is defined as medium: abstract learning goals have been defined.

Evaluation of municipal projects is mandatory by Bulgarian law. Therefore, an evaluation usually takes place 6 tot 12 months after a project has ended. As the implemented solution was just a small part of the neighborhoods' action plan, no evaluation efforts have been planned or undertaken yet. During the experiment, no explicit reflection activities took place. However, the small-scale implementation in itself can be seen as an evaluation activity of the action plan, followed by an evaluation of the small-scale implementation by including the outcomes in the action plan. The existence of mechanisms for learning can therefore be defined as medium: abstract learning mechanisms have been defined. However, after the implementation of the small-scale solution, the experiences from this implementation were

discussed and the action plan was refined accordingly. Similarly, the final meeting was aimed at identifying the best practices and sharing them with the other municipal departments.

Additionally, through the methods of collaboration, civil servants learned about different mechanisms and their impact on the experiment. This is specifically visible during the implementation phase, when the civil servants tried different approaches if one approach did not work out. Furthermore, the civil servants stated that they learned a lot during the meetings with the other cities within the Resilient Europe network, inspiring them to adopt best practices. One of the interviewees describes: "Being in a network with the other cities was very beneficial. The principals of communication and direct contact were influenced by the other experiments.".

E.2.4. Sequence of Actions
The first meeting of the experiment treated a broad range of topics and issues and led to the identification of the final topic and neighborhood to focus on. The following meeting built on these outcomes by defining the contours of an action plan and designing a small-scale experiment to validate the action plan. Key stakeholders were asked to help with contacting the local residents by going door to door. This was the first time the citizens were actively encouraged to share their thoughts on the issue and to participate in the experiment. Next, three more meetings were planned to further refine the action plan and incorporate the lessons learned from the experiment. A final meeting was held to share the lessons from the entire process with the relevant municipal departments. Overall, the activities were sequenced in a way that they built on the input of the previous activities. The only exception to this finding, is the fact that the majority of the residents was not involved until a year after the start of the experiment, which meant they had limited input on the design of the experiment. Had they been actively involved from the beginning, the experiment could be even better tailored to their needs. The sequence of actions has therefore been defined as medium: actions are sequenced in a meaningful manner, with some exceptions.

E.2.5. Methodology
Because of the involvement of Resilient Europe's project leaders, the civil servants had access to different methods from the Transition Management theory. Stakeholder analyses were, for example, conducted, visioning activities took place and the identification of transition pathways helped define strategies to reach these visions. Additionally, each of the pathways was translated into measurable indicators so the implemented solutions can be monitored and evaluated in the future. Some parts of the experiment, however, were based on trial-and-error rather than predefined methods, such as the example of contacting the citizens. The methodology has therefore been defined as medium: the experiment was structured according to both predefined methods and ad hoc approaches.

E.3. Output Features
E.3.1. Built capacities
Designing and conducting the experiment has been a very useful experience for the civil servants. By adopting a learning-by-doing approach, they were able to gain the basic skills needed to collaborate more intensively with the neighborhood's residents. They indicate that they do not only feel more confident repeating the process elsewhere, but are sharing their methods with their colleagues from other departments. Additionally, they feel more comfortable now taking on integrated projects, combining tasks and knowledge from the different municipal departments. The citizens, on the other hand, indicated that they did not gain any new skills. Built capacities is therefore defined as medium: either the government or the citizens indicate they gained or enhanced capacities.

E.3.2. Accountability

The municipality has traditionally been the institution responsible for mitigating flood risks. Citizens are used to relying heavily on the municipality to protect them, without taking any responsibility for protecting themselves. The goal of the experiment was to make citizens aware of the tasks they can undertake to mitigate flood risks within their own neighborhood. In the survey, when asked if they would like to remain active in the field of flood risk management most citizens indicated that they would like to continue to collaborate with the municipality on the issue. This indicates that they do feel responsible for helping the municipality mitigate flood risks, but still expect the municipality to take the lead. Accountability is therefore defined as medium: stakeholders who had not yet been involved in flood risk management do feel responsible for managing floor risks, but want to remain reactive.

E.3.3. Social change

In terms of social change, it appears from the outside that little has changed. The experiment did not result in any formalized new networks or forms of sustainable collaboration. Beneath the surface, however, both the civil servants and the citizens report that the experiment did succeed in helping them change their perspective on certain things. For the civil servants, the experiment has helped them change their perspective on the municipality's traditional top-down approach. They now believe that intensive collaboration with the citizens will lead to better and more accepted results and think that the citizens themselves can actually fulfill a bigger role within policy design and implementation. As described in one of the interviews: "I think that the driving force should come from the citizens and not just, as it always has been, from the municipality." The citizens reported a similar change in the way they view the municipality, stating that their participation in the experiment helped them see the municipality in a more positive light. Perception is therefore defined as high: both the policymakers and the citizens indicate they changed their perspective. As one of the civil servants summarized: "We started to build a bridge of trust between the citizens and the municipality that was previously not there."

E.3.4. Physical change

The first solution to have been implemented, was the planting of 5 willow trees at the central square near the church, with another 35 planned to be planted at different vulnerable areas in the neighborhood. Because, at the moment of evaluation, the solution has only been implemented in one location, physical change is rather limited. Nevertheless, even such a small solution is considered to be important for achieving social change by the civil servants, who indicated that the trees were not just a practical solution, but also a symbol of a tangible result that could come from intensive collaboration between the citizens and the municipality. It emphasized that the citizens could contribute to actual change in the neighborhood.

E.3.5. Actionable knowledge

In the first phase of the experiment, highly detailed transition pathways to encourage different dimensions of urban resilience, such as infrastructure, social and institutional, have been defined. In addition to these pathways, detailed actions and long- and short-term indicators have been designed. These have been defined making use of a scientific risk assessment and stakeholder discussions. This led to two types of knowledge: about the problem and about the normative goals that need to be reached when solving the problem. A third type of knowledge that can be identified is empirically based knowledge on the best methods to achieve these goals and foster transformative change. These were included in the action plan based on the outcomes of the experiment. One of the civil servants recognizes for example: "The most important thing [for us] is to go to the people. In the past, not many people would come to public discussions,

because they are at work when we organized them, or they couldn't travel. They were simply not capable of participating.". Actionable knowledge has therefore been defined as high: knowledge was generated on the problem, goals and methods of sustainability transitions.

E.3.6. Transferability
In the policy document detailing the transition pathways, it is stated that the actions and pathways are applicable to other parts of the city as well, but no further argumentation is provided. The interviews with the civil servants and the policy documents did show that parts of the experiment, like the forms of collaboration, small-scale experiments and the integration of different dimensions of resilience are shared with different municipal departments and that the civil servants intend on applying these lessons in other projects as well. A challenge that was identified is involving citizens in neighborhoods where the awareness of flood risks is much lower than in Dolno Ezerovo, which might require a different approach to engage them. Similarly, the civil servants emphasize the need for actors that are willing to act as mediators or front runners, which might not be available in different neighborhoods. None of these statements have been assessed or tested yet. Therefore, transferability is defined as medium: lessons for transferring have been formulated, but have not been validated.

E.3.7. Scalability
Similar to transferability, lessons for the upscaling of the experiment have been formulated and shared with other municipal departments, but they have not been applied yet. Scalability has therefore been defined as medium: lessons for scalability have been formulated, but they have not been validated. For the scaling out of the experiment, that is: the continuation of planting trees, the municipality plans to apply the same methods as for the first solution implementation. Scaling up the experiment, however, by involving more neighborhoods is thought to be a challenge. Civil servants expressed the concern that the municipality would not have enough manpower to facilitate such intensive collaboration with a larger number of citizens.

E.3.8. Accounting for consequences
When looking at the action plan, only one unintended consequence is explicitly defined. The identified consequence is the issue of litter in refurbished areas that can now be used for leisure. This would offset some of the ecological gains that are a result of refurbishing these areas. An annual clean-up event is planned to increase citizen's awareness of the importance of keeping the public spaces clean and to mitigate this issue. Additionally, the maintenance of the nature-based solutions was already considered, by planning to outsource the maintenance to an external company, which is common practice in Bulgaria. No further externalities of implementing the action plan have been identified.

E.4. Outcome features
E.4.1. Socio-ecological integrity
The civil servants indicated that their newly found skills and changed perspectives did inspire them to continue to contribute to implementing the action plan, encouraging citizen collaboration in projects and expand the focus of future projects to include the different elements of society, ecosystems and institutions. They even inspired colleagues from different departments to adopt similar approaches in the future. As for the citizens, however, they did not yet plan or undertake any actions related to the experiment on their own. Socio-ecological stewardship has therefore been defined as medium: actions inspired by the experiment planned or undertaken by either policymakers or citizens.

E.4.2. Intergenerational equity

Although all stakeholders have been represented in the experiment, this was not the result of a focused attempt to include all citizen stakeholder groups. When asked about future plans to keep these stakeholder groups involved, civil servants state that no plans have been defined at the moment. As one of the civil servants describes: "We don't really separate people, saying they are a less privileged group or not. We just want to reach the people living in the area." Intergenerational equity is therefore defined as no: no planned involvement of less privileged stakeholder groups. It should be noted, however, that this does not mean that the experiment will lead to intergenerational inequality. In the experiment, less privileged stakeholder groups appear to have experienced no barriers for participation and were well-represented, despite the lack of a specific focus on these stakeholder groups.

E.4.3. Socio-ecological integrity and democratic governance
Although both the civil servants and the citizens expressed a willingness to continue to collaborate in such an intensive manner, the experiment did not lead to the establishment of any, formalized, forms of collaboration between the two stakeholder groups beyond the already existing public consultation practices.

E.4.4. Livelihood opportunity
Although nature-based solutions, in theory, could lead to economic benefits for citizens and local businesses, this was not a dimension that was considered in this experiment. As a result, no data on the livelihood opportunity could be obtained.

E.4.5. Resource efficiency
The trees planted during the second phase of the experiment were willows that were grown in local nurseries. As a result, the transport of the materials was resource efficient. Resource efficiency has therefore been defined as high: local materials used. In addition to being local, willows were chosen specifically because of their ability to absorb large quantities of groundwater. This means that a smaller number of trees is required to achieve the same level of water absorption as compared to other tree species.

E.4.6. Precaution and adaption
The action plan created in this experiment contains highly detailed actions to mitigate flood risks and included specific indicators and targets that need to be achieved in order to successfully do so. It does so by focusing on different solutions, including citizen preparedness, ecosystem restoration, monitoring and hard-engineered measures, thereby adopting a more risk-averse approach than focusing on only one solution. Precaution and adaption is therefore defined as high: detailed flood risk mitigation strategies are formulated.

Appendix F: Evaluation Antwerp

F.1. Input Features

F.1.1. Awareness

In the past years, the citizens of the Sint-Andries neighborhood have been involved with two large infrastructure projects that included climate adaptation measures. Additionally, some smaller-scale measures are being implemented on street-level in the coming years. As a result, there is some awareness of the need to climate-proof the city. However, as the citizens in the neighborhood have experienced few issues with water management themselves, the overall awareness of flood risks is low. Furthermore, the majority of the residents has little knowledge of possible measures and is unaware of their own role in implementing these measures. As identified in one of the policy documents: "People tend to have very little knowledge of their private situation in relation to drainage, water buffering and their connection to the sewage system." Awareness is therefore defined as low: stakeholders are not aware of neighborhood flood risks, nor of their own role in flood risk management. When looking at awareness as a driver for commitment, it appears that most citizens had a different motive for joining the experiment. As the neighborhood struggles with issues such as few public areas available for leisure, low air quality and vehicle nuisance, most citizens approached the experiment initially as a way to address these issues and improve their living environment. This has been an important topic for the neighborhood and a few frontrunners had already started promoting initiatives such as communal vegetable gardens and the depavement of apartment block courtyards. They approached this experiment as a way to continue these initiatives.

F.1.2. Commitment

The experiment in Antwerp consisted of two different phases. The first, the preparation phase consisted of a number of meetings with what is described as the Urbact Local Group (ULG), consisting of residents of the neighborhood. If possible, the residents attended all meetings and new residents continued to join the ULG over the course of the process. The second phase was the implementation phase of one of the projects designed during the first phase. The responsible members of the ULG were active throughout this phase, supported by residents of the neighborhood that were unwilling to become structurally involved in the process, but were happy to help out on one or two occasions to brainstorm about possible solutions and help implement them during the "do-day". The experiment consisted of a highly committed ULG and a group of less committed residents. Commitment is therefore defined as medium: the experiment consisted of a highly-committed core group as well as a group of stakeholders that participated in part of the activities.

F.1.3. Trust

Within the different actor constellations, the focus will be on newly built relationship between the municipality and the citizens. The citizens themselves indicated to have very little faith in the municipality when it came to adapting the neighborhood to climate change. They, for example, pointed out that the cay area, an area prone to regular flooding, was to be redeveloped, using dikes and greenery to create both an area for leisure and make the area more flood resistant. In a change of plans, the municipality decided to build paved parking spaces and maintain the existing, concrete flood barriers instead. In many more areas of the neighborhood, the citizens could show examples of similar reversals of decisions or missed opportunities. As a result, many citizens believed that the municipality prioritizes economic growth over sustainability and that they cannot be trusted to take ownership of the climate adaptation issues. The municipality, on the other hand, extends some trust to the citizens, mainly through the Participatory Budget. This annual budget, consisting of 1,1 million euros, allows citizens to

allocate money to self-chosen, and self-implemented, projects. This appears to be a symbolic gesture rather than an important policy, with the experiment's process counselor arguing that the politicians' interest in the projects is rather low. This is corroborated by a civil servant, stating: "I have the feeling that politicians use [the participatory budget] as window dressing and think of the citizens' projects as endearing rather than useful." If can therefore be summarized that trust between the citizens and the municipality is low: stakeholders indicate not to be willing to rely on other participants' judgement and capacities. The citizens, however, indicated that it is precisely this lack of trust that has driven them to take matters into their own hands and become active in this experiment, as well as other projects. For them, this lack of trust has become a driver rather than a barrier.

F.1.4. Social capital
Two networks were included to initiate and guide the experiment: Stadslab2050 and Antwerpen aan 't Woord ('Antwerp Speaking'). Stadslab2050 is a platform initiated by the Antwerp municipality, aiming to address sustainability issues through experimentation. Stadslab2050's capital consists of both know-how and a network of experts. Antwerpen aan 't Woord, on the other hand, is a bottom-up initiative led by citizens with the same goals, but a know-how and network focused on citizens and citizen-inclusion. These two organizations collaborated to initiate and facilitate the process of experimentation. The facilitation of the process of experimentation was mainly the task of the process councilor from Antwerpen aan 't Woord, who already had good relationships with the neighborhood's residents and served as a mediator between the citizens and the municipality. His knowledge of the local context as well as the institutional context proved highly valuable for the experiment. During the evaluation, such a mediator connecting the municipality and the citizens was unanimously identified as an important enabling factor. Aside from making use of Antwerpen aan 't Woord's existing network, local associations were invited to join. The Sint-Andries neighborhood has a very active community, with a large number of citizen-led associations and initiatives and a neighborhood newspaper. A number of those were very active in the different phases of the experiment, from the local knowledge of the 'green' neighborhood associations, to the help with the implementation from the youth associations. Key figures active in these networks, as well as those active in less actively involved networks, helped spread information about the project within their networks and encouraged citizens to join the experiment.

F.1.5. Expertise
The process councilor with extensive of collaboration and citizen empowerment described in the previous paragraph was the most visible expert involved in the process. He helped guide the process of interaction and experimentation and help connect the participants with relevant stakeholders and organizations. Before the start of the experiment, the Resilient Europe project leaders visited Antwerp and spent two days with the civil servants to share their knowledge on transition management and urban resilience as well as provide input on the experiment possibilities. Next, a consultation meeting was organized with 15 experts in the different scientific fields related to the experiment. These experts helped define future impact of climate change in the neighborhood and the issues related to this impact. The outcomes of this consultation meeting served as input for the meetings with the citizens. In order to learn from the process of experimentation, an external evaluator was asked to evaluate the process at the end of the experiment. The findings from this evaluation have been taken into consideration for the upscaling of the experiment.

F.1.6. Support

The biggest form of support was given to the experiment in the form of €46.000,-, of which roughly half was funded by the EU and half by Antwerpen municipality. Access to adequate funding appeared to be an issue for the experiment participants, as the actual implementation of the experiment was not included in the budget. Therefore, the experiment participants applied for, and were subsequently granted, another €3.000,- from the city's Participatory Budget to actually implement some of the solutions. In addition to being granted funding to implement the experiment, the citizens participating in the experiment also received assistance in the shape of coaching sessions. These sessions covered a broad range of topics, from how to integrate long- and short-term visions and results and how to work with the city's policies to how to make use of the existing opportunities. The goal of these sessions was to empower citizens to continue the process of experimentation on their own. It was a form of support highly appreciated by the citizens.

F.1.7. Institutional Alignment
Most of the experimental solutions had to be implemented in public spaces. As the public domain highly regulated in Antwerp, this caused some frictions with the local laws and regulations. As a result, plans had to be adapted and quite some effort went into getting the municipality to agree with the proposed solutions. As the process councilor recounts: "[Getting the municipality to agree] was a bit of a challenge. We had to search, push and pull." In the end, however, they did manage to implement a number of small solutions. Institutional alignment is therefore defined as medium: stakeholders experienced institutional barriers, but could execute the experiment. Both the citizens and the process councilor attribute the fact that solutions were implemented in the end to the small scale of the solutions. This made it easier to work around existing regulations. One of the civil servants adds: "Participating in the Participatory Budget program also helps to build this environment in which more exemptions can be made than usually possible.". In one of the evaluation documents written by the municipality, this sentiment is echoed, stating that the entire municipality needs to embrace a more flexible approach when it comes to experimentation.

F.1.8. Leadership
For the first part of the experiment, the process councilor could be identified as the leader entity. He was responsible for organizing the meetings, contacting the stakeholders and encouraging them to join the experiment. An important part of this approach, however, was to empower stakeholders to initiate their own experiments. Project owners were appointed and coached to take over the leadership role. This led to a number of project frontrunners that still leaned on the process councilor for expertise, but organized the process themselves. For the continuation of the project, a new organization, Klimaatrobuust Sint-Andries ('Climate resistant Sint-Andries') has been formed with a board to oversee all of these projects. Leadership responsibilities are shared now between the board, for the coordination and integration of the different projects, and the frontrunners, for initiating and executing experiments. Overall, leadership can be defined as high: stakeholders could identify a leadership entity and considered him to be fully effective.

F.2. Activity Features
F.2.1. Transparency
In terms of direct communication with each of the stakeholder groups involved in the experiment, civil servants, process councilor and citizens, there have not been any communication issues. Transparency is therefore defined as high. In terms of access to municipal information, however, there were some issues. During the first phase of the experiment, the citizens would not have been able to acquire funding without the help of the

process councilor. For the continuation of the experiment, citizens indicated that finding their way in the institutional environment is seen as a major challenge. From the evaluation: "It is difficult for citizens to find their way in the municipal maze of rules and opportunities. This is an important lesson to take into consideration when transferring or scaling up the experiment."

F.2.2. Collaboration
Mechanisms for collaboration
In total, 44 stakeholders participated in the first meeting of the experiment. The goal of this meeting was to identify the main challenges of climate adaptation in the neighborhood, according to its residents and organizations that are active in the neighborhood. These were compared to the findings from the expert consultation and an agreement was reached on the priorities. The civil servants and the process councilor translated the outcomes from the first meeting into a document with the neighborhood's main challenges and objectives. A second meeting was organized to validate this document, brainstorm on actions and to encourage the citizens to form teams and take ownership of one of the actions. As not all stakeholders could make it to this meeting, a second meeting was organized in a similar fashion. The project teams were linked to a relevant contact person within the municipality and were encouraged to start to meet outside of the plenary sessions. Because the Green Corridor action was identified as a high priority action, the project team received additional coaching sessions. By organizing the process this way, citizens were immediately encouraged to take ownership of the problem, with the municipality and the process councilor being available for assistance. To show the participants what they could achieve and to learn from other city's experiences, a visit to Rotterdam was organized. Local politicians were also invited in order to enthuse them about the experiment. Next, a "dream day" was organized in the local community center. This workshop allowed the project team and residents that were not yet involved in the experiment to work on different themes and design their ideal future neighborhood. Not only did this workshop provide input on the actions to be taken, it also helped the residents to build a collective vision and ensured the support from the community. Following up on the "dream day", a "do day" was organized where citizens collaborated to implement a number of solutions. This day was much appreciated by the project team and the neighborhood's residents, as it showed that plans could be successfully translated into actual solutions. It also helped the residents to see what their street could look like, which further increased local support for the experiment. Afterwards, an evaluation was conducted and a final meeting was organized. At this meeting, the results from the evaluation were discussed and the participants shared their knowledge and findings with each other.

Parallel to the collaboration within the experiment, there were four meetings organized by the Resilient Europe network to share knowledge with the different partner cities. The civil servants attended the first two meetings and invited the project teams to join them for the last two meetings. The citizens really appreciated this, as it was an opportunity for them to learn about experimentation and gather input on their own experiments. They were proud to be included in the meetings and it helped them to develop an even stronger sense of problem ownership. They also indicated that it was nice to participate in these meetings, because it forced them to reflect on their process and actions, something that did not occur in the experiment itself.

Decisions
Three big decisions that were made through collaboration could be identified. First, the action plan was designed by the residents themselves with the support of the civil servants and the process councilor. In doing so, the stakeholders agreed on a shared vision for the neighborhood as well as actions that help achieve this vision. A second decision by all stakeholders was made

to prioritize the Green Corridor experiment over the other actions identified. A third decision on how to make a start on implementing the Green Corridor was made by the project team in collaboration with the civil servants and the process councilor. In short, one can conclude that all major decisions, other than the decision to focus specifically on climate change mitigation, were made in collaboration with all stakeholders and were based on a consensus. The citizens were given a lot of freedom to influence the process of experimentation. For the civil servants, this was a new approach. As one of them recounts: "We gave the citizens a carte blanche. We noticed that this was highly appreciated by the citizens, but it was very intimidating for us: what will be the outcome?"

Stakeholder representation
At the beginning of the experiment, the civil servants and the process councilor had meetings with key figures of different stakeholder groups, like the social housing association, healthcare associations and local schools to define the project. A limited number of them, however, was actively involved in conducting the experiment. The stakeholder group most dominantly represented in the experiment were the neighborhood's residents. The second stakeholder group included in the experiment were the civil servants, who helped set-up the first few meetings and involve stakeholders as well as gave advice and assistance to the citizens. Although no explicit goal of stakeholder inclusion was defined at the beginning of the experiment, especially the process councilor was disappointed about the stakeholders represented in the experiment.

All of the participants were homeowners or private housing tenants and native-born citizens of Belgium. Retired residents were overrepresented. Less privileged groups in the neighborhood, such as the elderly and disabled or the, mainly immigrant, social housing tenants were not represented. Stakeholder inclusion is therefore defined as low: the stakeholder group was largely homogeneous. However, the majority of the stakeholders interviewed did not see this as an issue. They argued that everybody was invited to join the process and rather than invest time and effort in engaging reluctant stakeholders, they chose to focus on the frontrunners. As one of them describes: "It takes so much time and effort to engage these less privileged groups and one year is simply not enough time to engage them all." He then follows up with: "You want to be able to show results at the end of the process. Sometimes you have to make the decision to speed up the process, which results in the exclusion of some stakeholders."

The stakeholders that were involved in the experiment, were very satisfied with the mechanisms for collaboration. They allowed for open discussion and gave citizens the freedom to take ownership of the problem, while enjoying the support from the civil servants. The civil servants, on the other hand, learned about the wishes and needs of the citizens, which allows them to address their issues better. Clear indicators of the stakeholder satisfaction are a continuation of the collaboration between the citizens themselves and between the citizens and the civil servants in a similar fashion. Civil servants are eager to repeat the process in other parts of the city. Stakeholder satisfaction is therefore defined as high: stakeholders would not change the mechanisms of collaboration.

In the evaluation, it was concluded that the small scale of the experiment was very helpful for committing stakeholders to the experiment. Not only did it help the civil servants to directly approach the different stakeholders, the stakeholders themselves were more willing to join because the project seemed manageable and had an impact on their direct living environment. They argue that, although it is important to scale up the experiment to make significant change, starting small is a good practice for building the necessary momentum.

F.2.3. Reflexivity and Learning
The process of experimentation in the Sint-Andries neighborhood was highly focused on learning. At the beginning of the experiment, detailed learning questions were defined that

covered a range of topics and perspectives. Such questions included, for example, how to change the perspectives of actors, which is a methodological question and how to adjust the institutional environment to encourage more nature-based solutions, which focuses on the municipality. The questions were also linked to specific outcomes, such as a systems analysis and a stakeholder map. Formulation of learning goals is therefore defined as high: learning goals have been formulated in detail.

Additionally, these questions were linked to different collaboration mechanisms, which should help answer the questions. Apart from short evaluation moments after the meetings, a large evaluation was planned at the end of the experiment. This was not only a mechanism that helped the civil servants and the process councilor answer their learning questions, it also helped the different stakeholders to reflect on the process and share their best and worst practices with each other. The lessons learned were summarized in a report that has been shared with other stakeholders and have been used to adapt the action plan. Due to these predesigned evaluation sessions and links between the collaborations and learning questions, formulation of mechanisms for learning has been defined as high: detailed mechanisms for learning have been designed.

In addition to formalized learning goals and mechanisms, informal learning, mainly as a result of the process of collaboration could be observed. For example, one of the citizens brought up the idea of the Green Corridor at the beginning of the process and over time, a lot of the stakeholders involved in the experiment have come to embrace the idea as an important way to mitigate the impact of climate change, which is an example of social learning.

The citizens themselves indicated that participating in the Resilient Europe network meetings gave them the opportunity to reflect on their process and actions. They stated that they found reflecting on the experiment a very useful experience and that they would have liked to have done it more often in order to be able to ensure both the process and the outcomes were efficient and useful.

F.2.4. Sequence of Actions

Both the preparation and the experimentation phase of the experiment were planned in detail before the start of the experiment. The sequencing of the meetings was designed in such a way that it started with informing the participants and then moved on to empowering and inspiring the citizens, followed by helping the participants to start the process of experimentation on their own. Finally, the experiment was followed up with an evaluation. For the continuation of the experiment, residents were encouraged to participate in network meetings to broaden their knowledge for scaling up or out. Each of the actions built on the previous actions and activities were scheduled almost every month to keep the participants interested without overburdening them. The sequence of actions has therefore been defined as high: actions are sequenced in a meaningful manner.

Furthermore, the "do day" was planned to coincide with the city of Antwerp's "car free day", which meant it could make use of the events already going on in the neighborhood to draw residents to the area where the solutions were being implemented. This enhanced the efforts of the project team to encourage residents to join the experiment.

F.2.5. Methodology

During the process of experimentation, a number of methods has been applied, such as, for example, stakeholder analysis, problem identification and visioning activities. These were linked to specific learning goals or questions. Furthermore, specific tools and approaches had been identified to optimize the outcomes of the methods. Nevertheless, there was also an element of trial-and-error, as not all aspects of the experiment could be managed using fixed methods. A civil servant states: "We used the DRIFT methodology to identify stakeholders.

But reaching these stakeholders turned out to be the real challenge." The process councilor adds: "There is a gap between theory and practice. You need to figure out what makes people tick in order to engage them. That is not something that you can learn from a method, it requires patience, flexibility and experience more than anything.". Methodology has therefore been defined as medium: the experiment was structured according to both predefined methods and ad hoc approaches.

F.3. Output Features
F.3.1. Built capacities
The civil servants interviewed emphasized the importance of learning-by-doing. By being actively involved in the experiment, they gained skills needed to facilitate future collaborations with and between citizens and to engage citizens in discussions on abstract topics such as climate resilience. Citizens also reported to have gained or enhanced skills and capacities. For example, they have learned to consider the impact of different solutions on climate change mitigation and have become adept at sharing their message with others. They also learned how to design and execute experiments. Built capacities is therefore defined as high: both citizens and policymakers report gained or enhanced skills and capacities.

However, the civil servants attributed their willingness to apply those skills not so much to the skill itself, but to the confidence gained by successfully participating in the experiment. This is a sentiment echoed by the citizens: being forced to apply those skills throughout the experiment, with successful outcomes, helped them build the confidence to use the skills in future

F.3.2. Accountability
The citizens that were involved in the experiment indicated to have gained a better understanding of the possible threats of climate change and state to actively want to help mitigate these threats in their neighborhood. They have become very active in sharing this message with their neighbors and have already undertaken actions. Accountability has therefore been defined as high: stakeholders feel responsible for flood risk management and want to become more proactive.

Stakeholders were largely unaware of these threats at the beginning of the experiment and it took some time for them to connect climate change to their own neighborhood. Being informed about the future threats was only partially responsible for them changing their mind. One of them recounts: "I wasn't interested in climate change adaptation until my young neighbors started telling me how important the topic was to them. That's when I realized I needed to get involved.". Part of the reason citizens want to be more active in the field of flood risk mitigation is, again, the feeling that, if they don't do it, nothing will happen.

F.3.3. Social change
The experiment resulted in a new neighborhood association: Klimaatrobuust Sint-Andris ("climate resilient Sint-Andries"). This association intends to continue to implement the action plan in order to increase the neighborhood's resilience in the face of climate change. Additionally, the experiment helped change the perspectives of both the civil servants and the citizens involved. The civil servants stated that the experiment helped them see that citizens can actually be involved in complex issues such as climate change adaptation and that giving them the freedom to design their own solutions will not only lead to good outcomes, it will also help increase public support for the plan. One of the civil servants explains: "For a different project, we didn't present a plan, but asked the citizens to discuss solutions with us. And you see that they become more engaged, start to discuss related matters like water infiltration or the use of rain water. That is something that never happened before." It shows that citizens, too, changed

their perspective, from being less interested in climate change adaptation to believing it is an important issue to be addressed. They also learned that climate adaptation measures do not have to include bulky, hard-engineered solutions, but can actually be integrated with solutions to increase the quality of living in the neighborhood. Changed perception is therefore defined as high: both policymakers and the citizens indicate they changed their perspective.

F.3.4. Physical change
During the "do day", the participants and residents experimented with a number of nature-based solutions in a number of different areas in the neighborhood. They depaved 30 square meters of a square, installed 4,5 square meters of living pavement, planted plants around trees in the neighborhood, covering an area of 20 square meters and built a planter that doubles as a rain water reservoir, storing 120 liters of water. In addition to the structural changes, they made use of temporary solutions, such as planters, artificial grass and even a fake pond to showcase what could possibly be achieved in the future. This helped the residents to visualize possible solutions and increased the support for the experiment in the neighborhood. Showing that citizens can achieve change in a relatively short time period is considered to be very important for the continuation of the experiment by the participants. From the evaluation: "The participants were excited when they heard other cities used the same approach of learning by doing. Instead of just planning, we made sure we had a healthy mix [of thinking and doing]."

F.3.5. Actionable knowledge
The citizens and the civil servants collaborated to design different transition pathways with actions aimed at enhancing the different aspects of climate resilience, focusing not only on floods, but also on, for example, a creating pleasant living environment and building social capital. These pathways have been combined with the findings from the expert consultation to create a robust action plan. Last, indicators to measure the progress by have been defined. The action plan, or map of opportunities, as the municipality has dubbed it, displays actionable knowledge on the issue as well as the normative solutions to the issue. Based on the experiences with the experiment, empirical knowledge about the tools and methods that can be used to achieve the goals defined in the action plan has also been generated. This knowledge has not only been used to adapt the action plan and identify future opportunities and challenges for this specific experiment, but also for other projects and experiments. Actionable knowledge has therefore been defined as high: knowledge was generated on the problem, goals and methods of sustainability transitions.

F.3.6. Transferability
In the evaluation document, it is proposed to set up similar processes for other 'greenification' actions in the city. Especially the "dream day" has been identified as a method that could be applied elsewhere to inspire and engage citizens. Approaching the citizens with a blank slate instead of predefined plans has already been applied successfully in different projects. Similarly, the "do day" has been earmarked as a best practice for building problem ownership. Since the experiment, a number of different "do days" have been organized in the neighborhood with positive results, for example by the city's nature conservation foundation. The experiment will also serve as a blueprint for the city-wide project Tuinstraten ("Garden Streets"). However, the process councilor and civil servants identified a possible challenge for transferring the experiment to other topics or neighborhoods. They acknowledge that much of the success of such an experiment is dependent on the stakeholders. It requires both interested, and flexible, civil servants and proactive citizens that are willing to be frontrunners.

Although they know that these roles need to be fulfilled, it is not possible for them to define who those stakeholders should or could be in different neighborhoods. As the process

councilor warns: "There is no blueprint for involving citizens. You will need to start from scratch in every neighborhood. It requires flexibility and a lot of patience." This was not a new lesson, but rather one that has been validated by the experiment itself.

Transferability has been defined as high: lessons for transferring have been formulated and validated.

F.3.7. Scalability

In terms of scaling out, the experiment participants are working to implement more small-scale solutions in the neighborhood in order to, in the end, have made a significant impact on the neighborhood. They learned that the current working method has a good balance between planning and doing and are therefore continuing in a similar fashion, with regular meetings taking place since the end of the experiment. Additionally, they found that a coalition of citizens and civil servants could achieve better results than a homogeneous stakeholder group could. For scaling out, this collaboration is deemed even more important and a monthly meeting between the citizens and the civil servants has therefore been established. The experiment design has also been included on a larger scale, in the design of the city-wide water plan, which will include similar processes of collaboration, co-creation and small-scale experimentation. While the ability of the experiment to scale out is already being proven by the continuation of it, the ability to scale up has not yet been validated.

Some possible issues with scaling up or out have also been identified. It has been noticed that the maintenance of the nature-based solution is an issue, with uncertainty about the ownership and many of the solutions not surviving the winter. It is expected that citizens are even less willing to take responsibility for the upkeep of the green solutions if even more of them are implemented. The coalition is therefore looking for innovative collaborations with the parks department to take care of the upkeep of future volumes of greenery. Additionally, the current small-scale solutions already experienced issues with the local regulations. It is expected that it will become even more difficult to implement solutions once their scale starts to increase. The institutional framework will have to be adapted to allow for the scaling up of the experiment.

As most of these lessons have not yet been validated in a real-world setting, scalability is defined as medium: lessons for scalability have been formulated, but they have not yet been validated.

F.3.8. Accounting for consequences

As described in the previous paragraph, one of the main consequences of uptake is that the nature-based solutions will require continuous monitoring and maintenance, which currently depends on neighborhood volunteers. As these might not be able or willing to do so in the future, this has been identified as a possible future challenge. A second consequence of the uptake of these nature-based solutions is that they require large quantities of water to thrive, which means they will need to be watered during dry periods. It is, however, highly inefficient to use treated and delivered tap water for the watering of the solutions. The experiment therefore included a small-scale water storage solution and the topic of rainwater storage and use has been marked as important for the continuation of the experiment.

F.4. Outcome Features

F.4.1. Socio-ecological integrity

As a result of the process of collaboration and the information they received, the local residents have committed to realizing more nature-based solutions in the neighborhood. Not just together with the Klimaatrobuust Sint-Andries association, but also in their private lives, by, for example, installing (façade) gardens or considering rainwater drainage when planning home

renovations. The civil servants, on the other hand, have expressed their willingness to continue to collaborate with the citizens to realize a larger-scale implementation of nature-based solutions and commit to the goals as defined in the action plan. They have shared the lessons learned with other departments and projects aiming to implement nature-based solutions.

F.4.2. Intergenerational equity

When asked about future plans to include the less privileged stakeholder groups in the neighborhood, the civil servants and the process councilor stated that they did think that would be highly relevant, as they have the most to gain by becoming more resilient. The board of the Klimaatrobuust Sint-Andries association did not agree, however, stating that they will continue to try to involve all neighborhood residents in their projects, but would rather focus their efforts on supporting those that are interested than including those that are not. Intergenerational equity has therefore been defined as no: no planned involvement of less privileged stakeholder groups.

It should be noted, however, that these stakeholder groups, often concentrated in social or assisted housing blocks, are entirely forgotten. The interviewed citizens identified a number of projects aimed at increasing the climate resilience and living environment in and around these housing blocks or matching volunteers with residents that would require immediate assistance in case of flooding or extreme heat. Despite these less privileged stakeholder groups not being actively involved in the planned projects, the projects are aimed at ensuring that they too reap the benefits of the different projects.

F.4.3. Socio-ecological stewardship and democratic governance

Even though the experiment was aimed at empowering the citizens to mitigate the risks of climate change themselves, the experiment thought the participants that a coalition of both citizens and civil servants is capable of achieving bigger change than either of the parties could do alone, due to the different skills and mandates. A monthly meeting between the board of the Klimaatrobuust Sint-Andries association and the civil servants has been established to discuss progress and challenges that either party could help each other with. This wasn't an intended outcome of the experiment, but one of the civil servants explains: "If you start something as the municipality, the citizens are going to expect certain things from you. You cannot start such a project without meeting those expectations." The process councilor confirms: "People said to me: we're participating in this experiment now, but what about the future? Can we still call you? Ask questions? Will you still be there for us? For some ease of mind, it is important for them to know that there will still be a coalition after the experiment ends." This is a lesson the civil servants learned from participating in the experiment.

F.4.4. Livelihood opportunity

Although nature-based solutions, in theory, could lead to economic benefits for citizens and local businesses, this was not a dimension that was considered in this experiment. As a result, no data on the livelihood opportunity could be obtained.

F.4.5. Resource efficiency

The plants used for the implementation of a number of solutions came from local nurseries, which limited the resources needed for transport. In addition, some plants were re-used from other projects in the city, making the plant choice even more efficient. Resource efficiency has therefore been defined as high: local materials used. There were, however, some issues with the plants that were included in the experiment. Not all of them appeared to be as durable and quite of lot of them died as a result of drought, cold and being walked over. Merely half a year after they had been planted, they needed to be replaced again, which is not a sustainable practice.

F.4.6. Precaution and adaption
In the action plan, highly detailed actions and indicators have been defined to help reduce the flood risks in the neighborhood. These actions are related to different aspects of flood risk and climate resilience in the neighborhood, such as water infiltration and storage, but also citizen awareness and emergency training. Due to the inclusion of different aspects and many overlapping solutions, this is a highly risk-averse approach. Precaution and adaption is therefore defined as high: detailed flood risk mitigation strategies are formulated.

Governance of Experimentation
A framework for evaluating transition experiments for urban flood risk management

Academic Article

Governance of Experimentation: a framework for evaluating transition experiments for urban flood risk management

L.J. van Buggenum[1]

[1]Delft University of Technology, Faculty of Technology, Policy and Management
Jaffalaan 5, 2628BX, Delft, The Netherlands
L.J.vanBuggenum@student.tudelft.nl

Abstract. Transition experiments are thought to be important tools for achieving transformational change in the urban water system. The governance of these experiments is described as a key factor in the success of these experiments. This paper sets out to operationalize and validate an evaluative scheme for the appraisal of transition experiments. This paper employs a case study approach to evaluate the governance of two transition experiments through the application of the evaluative scheme. This research indicates that learning-by-doing and applying the principles of good governance are enabling factors for realizing stakeholder empowerment and social change. The large size of the evaluative scheme is a point of concern for the broad applicability and usability of the framework as a common method for evaluation.

Keywords. Governance, Transition Experiment, Evaluation, Flood risk management, Resilience

1. Introduction

1.1. Introduction

Flood risks are projected to increase in the coming years as a result of ongoing climate change (Feyen et al., 2012). Especially cities are expected to experience problems, as the high built-up density of cities hinders water absorption as well as increases the social and economic impact of flooding (Mentens et al., 2006). As a result, *Flood Risk Management* (FRM) has become a focal point of attention for both policymakers and researchers (Alexander et al., 2016). FRM Is defined here as a governance approach aiming to reduce the likelihood and impact of flooding by implementing structural and non-structural solutions, while taking into consideration the social, ecological and economic aspects related to flooding (Butler & Pidgeon, 2011; Evers et al., 2016; Hartmann & Driessen, 2017; Ward et al., 2013). Governance can be understood as the resources, rules and mechanisms that allow actors to manage flood risks in networks rather than hierarchies (Alexander et al., 2016).

The goal of FRM is to increase the resilience of urban water systems (Hegger et al., 2016). *Resilience* has been defined by Meerow, Newell and Stults (2016) as: "[…] the ability of an urban system, and all its constituent socio-ecological and socio-technical networks across temporal and spatial scales, to maintain or rapidly return to desired functions in the face of a disturbance, to adapt to change, and to quickly transform systems that limit current or future adaptive capacity." (Meerow, Newell and Stults, 2016, p.45). This focus on resilience is a consequence of the growing understanding that traditional flood defense measures are, in many cases, fallible, economically unfeasible or unsustainable in the long term (Hegger et al., 2014). Instead of promoting systems that are fail-safe, increasingly the focus of FRM is on designing systems that are safe-to-fail (Ahern, 2011).

1.2. Transition experiments for resilient flood risk management

Despite a growing understanding of the need to transition towards a more resilient water system, transformational change is not easily achieved (Farrelly & Brown, 2011). It requires technological innovation as well as radically adjusted practices, organizational structures and social values (Weiland et al., 2017). These changes are facilitated through a long-term process of change during which systems gradually shift towards more sustainable modes of operation (Rotmans et al., 2001). Small-scale experimentation can be a driver of this transformational change (Schot & Geels, 2008). Learning from as well the repeating and upscaling of experiments can provide the stepping stones towards a more resilient system (Collins et al., 2017). In this paper, the definition of such *transition experiments* by Van den Bosch and Rotmans (2008) is employed. They define these experiments as: "[…] innovation projects with a societal challenge as a starting point for learning aimed at contributing to a transition." (Van Den Bosch and Rotmans, 2008, p.13).

Three types of learning take place in transition experiments: broad, about the problem dimensions, reflexive, about the underlying assumptions and values and social, to reframe the perspective of the stakeholders (Van Den Bosch & Rotmans, 2008). These types of learning are the result of a process of stakeholder interaction (Buizer et al., 2016; Foley, Bernstein, & Wiek, 2016) Multi-stakeholder participation is therefore identified as an important characteristic of transition experiments (Kivimaa et al., 2017; Voytenko et al., 2016). *Reflexive governance* is the mode of governance concerned with facilitating these processes of collaborative learning (Hendriks & Grin, 2007; Schäpke et al., 2017; G. Walker & Shove, 2007). This type of governance comprises practices of governing that encourage actors to learn about innovations, evaluate and revise their underlying assumptions, values and frames and assess the process of

governance itself (Buizer et al., 2016; Dryzek & Pickering, 2017; Termeer et al., 2016; Voss & Kemp, 2005b).

1.3. Evaluating transition experiments

The governance of such experiments could play a key role in the contribution an experiment makes towards the transition process (Weiland et al., 2017). Yet despite this potential, the governance of transition experiments is still sorely under-researched (Hildén et al., 2017; Kivimaa et al., 2017). In-depth evaluation of real-world transition experiments is therefore highly desired to identify governance practices that amplify the transformative power of transition experiments (Hildén et al., 2017).

These experiments with their plurality of interdependent goals and related sub-governance arrangements, however, are highly complex (Alexander et al., 2016). A tool for structuring this complexity is the logic model framework (Cooksy et al., 2001; Julian, 1997; Savaya & Waysman, 2005). The basic premise of the framework is the understanding that an experiment reaches desired outcomes by performing activities and making use of resources, while operating within a specific context (McLaughlin & Jordan, 2015). It is therefore concerned with a holistic form of analysis by not just identifying the governance arrangements, but also the degree to which they contributed to the outcomes of the experiments (Alexander et al., 2016). The logic model framework consists of four categories: inputs, activities, outputs and outcomes (Cooksy et al., 2001; McLaughlin & Jordan, 2015; Savaya & Waysman, 2005). These should not be interpreted as linear, but rather as parallel and interdependent as experimentation takes place in iterative cycles rather than one linear process (Luederitz et al., 2017).
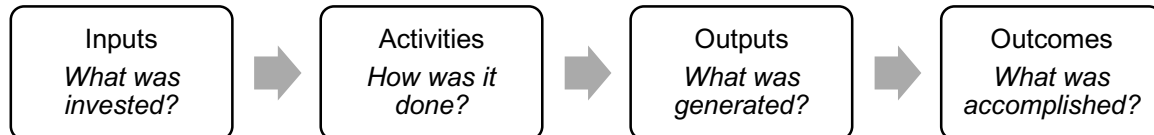

Figure 1: logic model of evaluation

The governance process takes place in the input and activity categories, while the output and outcome categories contain the short- and long-term goals of experiments. Theory on the relations between inputs, activities, outputs and outcomes, be it from literature or experiment-specific documentation and stakeholder interviews, is required to identify different criteria in each of the categories (Chen, 2012). Based on data from 61 real-world experiments, Luederitz et al. (2017) identified criteria for each of the four categories of the logic model of evaluation. They invite scholars and practitioners to apply the framework and to "critically reflect upon its potentials and limitations and take part in learning from and improving transition experiments" (Luederitz et al, p. 63). This research sets out to do so by operationalizing the scheme and applying it to two transition experiments focusing on urban flood resilience. The goal of this research is therefore to contribute to the body of knowledge on the governance of transition experiments by validating a tool for future research into the design and governance of these experiments.

This research aims to answer the following research question: How can the governance of transition experiments be evaluated in relation to the experiment's success, focusing on flood risk management? This paper is structured as follows: in section 2, a framework for the evaluation of transition experiments designed by Luederitz et al. (2017) will be assessed and operationalized. In section 3, the case study method used to validate the framework will be

introduced, followed by a description of the cases in section 4. In section 5, the results will be presented and discussed after which the conclusions will be formulated in section 6.

It should be noted that this article is based on the unpublished master's thesis report written by the author (Van Buggenum, 2018). This article therefore leans heavily on this report, with paragraph 4 having been directly copied and paragraph 5 having been adapted from the master's thesis report.

2. A framework for evaluation

In addition to making use of the criteria designed by Luederitz et al. (2017), literature on urban resilience, FRM and reflexive governance has been studied in order to operationalize the criteria and identify possible gaps in the framework.

2.1. Input features

Luederitz et al. (2017) identify five input criteria. *Awareness* of the need for flood risk management and the need to manage these risks in networks is thought to increase stakeholders' willingness to participate in transition experiments (Evers et al., 2016; Ward et al., 2013). *Commitment* of stakeholders throughout the experiment, the only criterion not reflected in the literature, is considered to be important for learning (Antikainen et al., 2017). *Expertise* is thought to provide valuable insights and direction to a decision-making process when used in combination with lay deliberation (Dryzek & Pickering, 2017). *Trust* is a key factor in ensuring an open and constructive design process (Thaler & Levin-Keitel, 2016). A high level of trust between stakeholders helps reach widely accepted agreements and builds networks for future collaboration (Benson et al., 2016). *Support,* such as funding, assistance and access to information, is an important tool for the design and implementation of FRM (Alexander et al., 2016). In addition to the five criteria identified by Luederitz et al., (2017) three more input criteria could be identified. *Institutional alignment* is needed to facilitate local autonomy and innovation that is uninhibited by conflicting laws and regulations (Krieger, 2013; Wilby & Keenan, 2012). *Leadership* is considered to be vital to initiate an experiment, guide interactions and resolve conflicts (Folke et al., 2010; Olsson et al., 2004). Additionally, leaders can give direction to the process and inspire and motivate stakeholders when the transition process hits a rough patch (Lebel et al., 2006). *Social capital*, such as networks and shared values, facilitates smooth collaboration during the experiment as well as continued collaboration after the experiment ends (Schäpke et al., 2017).

Feature	Criteria
Awareness	The experiment should include stakeholders that are aware of flood risks and their own role in FRM
Commitment	The experiment should include highly committed stakeholders
Trust	The experiment should include stakeholders that are willing to rely on each other's judgement and capacities
Expertise	The experiment should include participants with the skills to carry out the experiment
Support	The experiment should include the necessary support to carry out the experiment
Inst. Alignment	The experiment should not encounter any institutional barriers
Leadership	The experiment should include a leader with the skills to guide the experiment
Social Capital	The experiment should include pre-existing social capital

Table 1: Input criteria

2.2. Activity features

The *sequence of actions* It is concerned with the sequencing and timing of the different activities within the experiment to create opportunities and minimize threats (Luederitz et al, 2017). The

goal of *methodology* is to ensure the experiment reaches the desired outcome of facilitating transformational change. It includes the use of methods for problem analysis, solution design, monitoring and evaluating (Luederitz et al, 2017). *Collaboration* not only encompasses methods of collaboration, but is closely related to the concepts of good governance and legitimacy (Hartmann & Driessen, 2017; Johnson et al., 2007). Stakeholder inclusion, fairness and stakeholder satisfaction are therefore all relevant indicators for the quality of collaboration (Abrams et al., 2003). *Reflexivity and learning* can facilitate broad, reflexive and social learning (Van Den Bosch & Rotmans, 2008). It allows for generating knowledge as well as converging values and empowering stakeholders (Hendriks & Grin, 2007; Termeer et al., 2016). *Transparency* is closely related to the aforementioned legitimacy (Alexander et al., 2016). It is thought to increase the accountability of the experiment participants as well as increase the understanding and acceptance of choices made throughout the experimentation process (Luederitz et al., 2017).

Feature	Criteria
Sequence of Actions	The experiment should be conducted according to a meaningful sequence of actions
Methodology	The experiment should make use of a sound methodology
Collaboration	The experiment should include mechanisms for collaboration
	The experiment should facilitate collaborative decision-making
	The experiment should facilitate stakeholder inclusion
	The stakeholders should be satisfied with the process of collaboration
Reflexivity and Learning	The experiment should include learning goals and mechanisms for learning
Transparency	The stakeholders should be satisfied with the timeliness, relevancy, comprehensiveness, accuracy and reliability of information

Table 2: Activity criteria

2.3. Output features

Built capacities deals with the empowerment of stakeholders to address future sustainability challenges and continue to contribute to a transition towards a more resilient system (Butler & Pidgeon, 2011; Schäpke et al., 2017). *Accountability* is not to be interpreted as being able to be held accountable, but to feel accountable for facilitating sustainable change (Luederitz et al., 2017). *Social change* can be summarized as built social capital. It can consist of new networks, changed perspectives and shared values (Benson et al., 2016; Newig et al., 2016). Built capacities, accountability and social change are all thought to be important enabling factors for the continuation of the experiment and to respond to future disturbance to the system (Luederitz et al., 2017; Olsson et al., 2004). *Physical change* encompasses the implementation of solutions that increase the resilience of the water system and should be multifunctional, redundant and modular (Ahern, 2011). *Actionable knowledge* is evidence-based knowledge on sustainability solutions and their design and implementation (Luederitz et al., 2017). Similarly, *transferability* and *scalability* are, preferably validated, knowledge generated by the experiment on the possibilities for transferring or scaling up the experiment. Last, *accounting for consequences* encompasses knowledge generated on the possible externalities created by the experiment (Voss & Kemp, 2005b). The latter four features are not so much activities as they are knowledge generated by the experiment that can be used to advance the transition towards a resilient water system or can be generalized and applied to other sustainability transitions.

Feature	Criteria
Built capacities	Stakeholders should have gained or enhanced skills
Accountability	Stakeholders should have gained a feeling of responsibility for managing flood risks
Social change	Stakeholders should have formed new networks
	Stakeholder should have changed their perspectives

Physical change	Resilient, physical solutions should have been implemented
Actionable knowledge	Actionable knowledge on the sustainability problem and solutions should have been generated
Transferability	Knowledge on the transferability of the experiment should have been generated
Scalability	Knowledge on the scalability of the experiment should have been generated
Consequences	Knowledge on the possibility of externalities should have been generated

Table 3: Output criteria

2.4. Outcome features

Socio-ecological integrity recognizes the interdependencies between physical, social and ecological systems and aims to harmonize these interactions to make use of the natural capacity of ecosystems to regenerate and have these effects spill over into the other systems as well (Gibson, 2006). *Intra- and intergenerational equity* is another aspect of legitimacy, ensuring equal access to equal access to decision-making processes and the benefits from any implemented solutions (Alexander et al., 2016; Johnson et al., 2007). *Socio-ecological stewardship and democratic governance* deals with both stakeholder empowerment to continue to take care of system resilience as well as the inclusion of all relevant stakeholders in these processes (Folke et al., 2002; Pickett et al., 2004). *Livelihood opportunity* encompasses the access of citizens to commodities that are needed to live a decent life, including, for example, resources that are needed for sustaining life or performing economic activities (Luederitz et al., 2017). *Resource maintenance and efficiency* is defined as reducing the input that is needed for one unit of output, by, for example, recycling or more efficient production technologies (Bhamra et al., 2011). Lastly, *precaution and adaption* recognizes that a system is never in equilibrium and that risk-averse approaches, anticipating on changes, surprises and the need for constant adaptation, are desired (Carpenter et al., 2001; Meerow et al., 2016; Tyler & Moench, 2012).

Feature	Criteria
Socio-ecological Integrity	Further implementation of resilient solutions should have been planned or undertaken
Equity	Inclusion of less privileged stakeholders in facilitating the transition should have been planned or undertaken
Democratic governance	New or existing forms of participatory should have been established or improved
Livelihood opportunity	The experiment should have created economic opportunities
Resource efficiency	The experiment should have ensured resource efficiency
Precaution and adaption	The experiment should have facilitated a risk-averse approach towards flood risk management

Table 4: Outcome criteria

3. Material and Methods

3.1. Selection of the cases

In order to validate this framework, empirical research was conducted in Burgas, Bulgaria and Antwerp, Belgium. Both of the experiments took place within the framework of the RESILIENT EUROPE network, a program co-financed by the European Regional Development Fund and the member states of the European Union, Norway and Switzerland (URBACT, 2014). The choice to focus on the experiments within the RESILIENT EUROPE network was directly informed by the access of the researcher to information about these specific experiments.

This case selection method used has been labeled the 'diverse-case method' and aims to achieve variance in order to increase the representativeness of the cases selected (Gerring & Seawright, 2007). As the focus of the experiments on urban flood resilience was not to change, the experiments were selected to represent variance in their contexts. Bulgaria is a developing country and one of the weakest countries in the European Union in terms of democracy and governance scores (Sustainable Governance Indicators, 2018a, 2018b). Belgium, on the other hand, is a highly-developed country. Societal consultation is common and citizens are relatively well organized and equipped to participate in policy processes (Sustainable Governance Indicators, 2018a, 2018b). The experiments were therefore deemed likely to have encountered different barriers and opportunities and, as a result, have designed different governance mechanisms to respond to these barriers and opportunities.

3.2. Case analysis

In both cases, the logic model for evaluation was used to design the case study questions. A pragmatic research approach was employed, designing indicators based on the defined criteria as well as allowing for in-depth, qualitative evaluation. Policy documents and documentation from the RESILIENT EUROPE network were consulted to gain information. In addition, 8 semi-structured interviews were conducted with both civil servants, citizens and process counsellors involved in the experiments. As the citizens that participated in the experiment in Burgas were not available for interviews, six of them were asked to fill out a survey with open-ended questions to gain their input on the experiments. Interviewees were asked to reflect on the different framework criteria and to identify any observed relationships between the criteria. After transcribing the interviews, the data was systematically coded and subjected to thematic analysis (Braun & Clarke, 2006). The themes or patterns identified in the cases were then compared to each other and the logic model for evaluation in order to validate the findings. As limited literature on the governance of transition experiments is available, different fields of literature were analyzed to look for possible theoretical explanations for the research findings.

Due to practical constraints, only the most relevant outcomes of the case studies and the evaluation of the framework itself will be discussed in the paragraphs below.

4. Case Study Results

4.1. The 'Green Belt' experiment, Burgas, Bulgaria

The Dolno Ezerovo neighborhood is located at the outskirts of Burgas, bordering Lake Vaya. Most of the neighborhood is located below sea level. Since 2010, the neighborhood experienced 5 major floods, causing significant damage to properties and infrastructure and even human deaths (Burgas Municipality, 2018). Climate change is expected to increase the risk of flooding and it is therefore important to increase the neighborhood's ability to withstand these threats. The goal of the experiment was therefore related to increasing the neighborhood's resilience. Not by focusing on hard-engineered measures, as is common in Burgas, but by focusing on the relation between infrastructure, people and ecosystems. After receiving help from the Resilient Europe network project leaders, a collaboration process was designed, the main goal of which was to co-create a detailed action plan to increase the neighborhood's resilience. A risk assessment was performed by an external company which served as the main input for the design of the action plan. A budget of 45.500 euro was made available through EU funding and municipal funding.

Preparation Phase

In order to determine the focus of the experiment, a first meeting was organized with the municipality's consultative council. Stakeholders included representatives of the harbor and airport, local businesses, the Red Cross and the local police and fire departments. During this meeting, the contours of the action plan were defined. A second meeting was organized with the most relevant stakeholders from the consultative council, including different municipal departments and NGO's like the Red Cross and environmental organizations, as well as residents from the neighborhood. During this meeting, the idea of implementing a 'Green Belt' of nature-based solutions along the neighborhood's gullies was agreed upon. The idea of experimenting with the small-scale implementation of such a nature-based solution in collaboration with the residents was also conceptualized during this meeting.

Experimentation Phase
Initial attempts to engage citizens in the implementation of a small-scale solution had been unsuccessful. Overall, citizens felt let down by the municipality. Previous negative experiences with public consultations led them to believe that their input would not be appreciated by the municipality. The head of the local center for administrative services was therefore asked to act as a mediator between the civil servants and the citizens. With her help, the civil servants succeeded in engaging some local key stakeholders who engaged their networks. Additionally, the civil servants held door-to-door conversations with the citizens to discuss the experiment and action plan with them. This method of direct contact proved to be a useful tool in engaging all stakeholders, including less privileged stakeholder groups, in the experiment. A meeting was organized where the citizens could discuss where to implement the small-scale solution. A festive event, kicked off by the deputy mayor, was organized to plant a symbolic first five willow trees from a local nursery at this chosen location. A very high number of residents attended the event. The present civil servants made use of this high turn-out to inform residents about the initiative, receive input on the action plan and build public support. Residents expressed their satisfaction with the nature-based solution, which would not only help to protect against floods, but also improved the local living environment.

Conclusion phase
After the small-scale solution had been implemented, four more meetings with the stakeholders from the consultative council and some residents were organized. During these meetings, the stakeholders reflected on the lessons learned from the small-scale implementation and designed different pathways for achieving urban resilience. An external consultancy company was asked to design a detailed action plan based on these pathways, which was discussed and refined during the meetings. The last of the four meetings was aimed at identifying the next actions to take and to inspire all stakeholders involved to proactively work on increasing resilience in the neighborhood

Outcomes
Both the citizens and the civil servants reported to be very satisfied with the process of collaboration. The civil servants stated to have learned new methods to involve citizens in the process of policymaking and have started to see the potential of the residents becoming more proactive in climate change mitigation. The citizens, on the other hand, reported to have regained some trust in the municipality by actively being involved in the design of the action plan. They also have an increased awareness of the measures that can be taken to protect themselves and their neighborhood from floods. However, actual changes in behavior were only visible at a municipal level. Plans are being made to repeat the process of experimentation in other areas of the city. An increased awareness of the importance of collaboration between municipal departments inspired further collaboration. The citizens, however, have not yet

started to change their behavior, although indicate a high willingness to collaborate in a similar fashion with the municipality for implementation of the action plan.

4.2.The 'Green Corridor' experiment, Antwerp, Belgium

The Sint-Andries neighborhood is located in Antwerp city center, bordering the Scheldt river. It has a high building density, with few public and private green spaces available. Although the neighborhood has currently had limited experiences with floods, its location near the river and the high building and people density makes it vulnerable to the impact of climate change. In order to increase the neighborhood's resilience to these threats, the area had been selected for an experiment. The main goal of the experiment was to not only increase the resilience of the place, but also of the people. The experiment was therefore heavily focused on stakeholder empowerment. Civil servants, project leaders of the Resilient Europe network and an experienced process councilor collaborated to design the basic set up and goals of the experiment. Specific learning questions and methods were designed in advance and linked to different aspects of the experiment in order to increase the experiment's impact. A group of experts was asked to identify the main issues and related goals in the neighborhood. This information could be used to support the co-creation process. A budget of €46.000 euro was made available through EU funding and municipal funding.

Preparation Phase
In order to determine which topic to focus on, meetings were first held with different stakeholder groups in the neighborhood, such as the social housing association and youth organizations. This helped define the contours of the experiment. The first challenge then was to engage the neighborhood's residents. Although the process councilor already had an extensive network in the neighborhood, engaging the residents still required a lot of effort. By approaching interesting stakeholders or stakeholder groups individually, a group of 44 residents could be convinced to participate in the first meeting. Their motivations for joining this meeting varied, but many of them were more concerned about their direct living environment than about the threats of climate change. They reported they were driven to join because their lack of trust in the municipality caused them to believe that if they didn't address this issue, nobody else would. During this first meeting, the residents were informed about the experiment and were asked to identify the issues that were most important to them. Two follow-up meetings were then organized to allow the residents to identify actions and pathways to address these issues. These pathways were aimed at addressing different aspects of climate resilience, such as social and infrastructural resilience. The idea of the Green Corridor, a corridor connecting different nature-based solutions throughout the neighborhood, was designed.

Experimentation Phase
A group of residents volunteered to be responsible for the implementation of the Green Corridor. This project group met five times over the course of the experiment to discuss the project and plan the implementation of different solutions. The project group was coached by the process councilor and the civil servants. The process councilor applied for and received and additional €3.000 from the city's participatory budget to make a start implementing the Green Corridor. A meeting to Rotterdam was organized to inspire both the experiment participants and local politicians. Throughout the experiment, the citizens and process councilor tried to involve the local politicians, but there was limited interest for the project or the issue of climate change. This only enforced the citizens' belief that the municipality would be of little help to them when making their neighborhood climate resilient. In collaboration with the project group, civil servants and process councilor, a "dream day" was organized in the local community

center where residents were invited to design their ideal neighborhood and link these dreams to the Green Corridor project. This event was followed up with a "do day", where all residents were invited to help implement some small-scale solutions, such as depaving squares and building planters. In addition to permanent solutions, some temporary solutions, such as artificial grass and a fake pond were used to show what would be possible in the future. Plants from local nurseries were used and some plants from other areas of the city were 'recycled'. This event was organized at the same time as the city-wide "car free day", which drew a lot of visitors to the neighborhood. Both events were highly appreciated by the neighborhood. They helped people visualize what could be achieved in the neighborhood and how they could contribute to it. It helped to commit residents to the idea of a Green Corridor and created a sense of problem ownership. The experiment was concluded with an evaluation, conducted by an independent evaluator.

Outcomes
Both the citizens and the civil servants reported that they were very satisfied with the process of collaboration and the outcomes. Participating in the experiment was a useful experience for all stakeholders. The civil servants reported to have developed the skills to facilitate co-creation processes and indicated to have gained a new perspective on the capabilities of citizens to tackle such abstract issues as climate change. Citizens, on the other hand, learned how to organize themselves and to design and implement solutions that mitigate climate risks. They indicated that participating in the experiment made them aware of the threats of climate change and the issue is now high on their agendas. The citizens have established an association, Klimaatrobuust Sint-Andries ("Climate Resilient Sint-Andries"), in order to continue to implement climate risk mitigating solutions. In addition, stakeholders report that, inspired by the experiment, residents have also started to take private measures to mitigate the threat of climate change, for example by considering rainwater drainage when remodeling their house. A monthly meeting with the civil servants has been established to discuss progress and issues. The pathways and actions have been adjusted based on the outcomes of the experiment and both citizens and civil servants are committed to meeting the predefined goals.

5. Results and Discussion

5.1. Similarities between the cases

Both the experiment in Burgas and the experiment in Antwerp adhered to a number of good governance principles, the most dominant of which are consensus-based decision-making, transparency and responsiveness. In both cases, the open and collaborative process facilitated the formulation of a shared vision of the pathways towards resilience in the neighborhoods. Stakeholders highly appreciated the opportunity to ensure their interests were represented in the action plan. Additionally, equal access to information and equal opportunities to influence the decision-making increased the stakeholder satisfaction with the process and helped build trust and commitment. Although no other literature on transition experiments has been published yet to contradict or support this claim, the relationship between good governance and factors such as trust and satisfaction has been supported by literature on other topics (e.g. Bouckaert & van de Walle, 2003; Ott, 2011; Yousaf, Ihsan, & Ellahi, 2016).

Another dominant feature that could be observed is the importance of "learning-by-doing". In both cases, putting theory to practice helped the civil servants gain new skills, capacities and confidence. Similarly, by discussing the elements and goals of neighborhood resilience, the experiment participants reached a consensus on the vision for the neighborhood. This is a form of social learning that emerged as a result of actively working with the topic of

resilience. Additionally, by sharing their lessons with other civil servants within the Resilient Europe network and actively contributing to these other experiments, civil servants were forced to reflect on their own processes. These activities therefore encouraged learning about the process of collaboration itself. Again, although no research has been conducted on this phenomenon occurring within transition experiments, it seems to align with the so-called Experiential Learning Theory, which emphasizes the importance of learning-by-doing (Buri et al., 2012). Although experiential learning has been researched in many different fields, such as higher education and corporate learning, especially interesting for transition experiments is the focus on experiential learning and its impact on civic responsibility (e.g. Billig, 2000; Hatcher, 1996). The presumed relationship between learning-by-doing and building accountability can be highly relevant for the future design of transition experiments.

It should be noted that *equal opportunities to influence the decision-making* mentioned earlier does not imply that all stakeholders need to be involved in this process of decision-making. For both experiments, a small core group of stakeholder representatives was responsible for the majority of the decisions made, with a larger group of stakeholders being given the opportunity to provide input on these decisions at highly accessible events. By organizing the experiments this way, the process remained adaptive and manageable, while ensuring responsiveness to the populations needs. This led to quick results as well as broad public support for the action plans. One should be careful however that long-term commitment is considered to be a prerequisite for higher order learning in transition experiments, so the level and type of learning will likely differ depending on the stakeholders' commitment to the process (Antikainen et al., 2017).

5.2. Differences between the cases

The main difference between the two cases that kept reoccurring when comparing the two experiments, is the difference between a top-down and a bottom-up approach. Emphasizing the importance of learning-by-doing, most of the learning in the top-down focused experiment took place at the level of the civil servants. In the bottom-up focused experiment, on the other hands, citizens were given the opportunity to become actively engaged in the experiment, thus engaging in learning activities. Additionally, because the citizens played such a big role in the design and first implementation of the action plan, they reported to feel a strong sense of responsibility for the continued implementation of what they considered to be *their* action plan. The downside of such a bottom-up focus is the question of accountability, however. For a top-down approach, the accountability for the outcomes remains with the municipality. For a bottom-up approach, on the other hand, the boundaries become blurry. New, clearly-defined rules for sharing the responsibility between the municipality and the citizens will have to be designed in such cases.

In addition, an externality can be observed when comparing the citizen empowerment. In Burgas, the citizens involved in the experiment were not empowered or inspired to take individual actions to protect themselves and their neighborhood from flood risks. In Antwerp, on the other hand, both citizens directly involved in the experiment and citizens that were not directly involved in the experiment have started to take individual actions. Remarkably, the residents in Burgas have had more direct contact with the civil servants and should be better informed about the importance of neighborhood resilience. The citizens in Antwerp, however, had the opportunity to observe their neighbors take collective and individual actions to increase the resilience in the neighborhood. It can be hypothesized that awareness of possible individual actions to take is not enough to actually empower citizens, but that copycat behavior can be a driver for change. Psychologists refer to this phenomenon as social proof: people are more likely to change their perception or behavior if they observe others in their network doing so

(e.g. Cialdini, 2009; Contractor & DeChurch, 2014). Empowering local citizens can therefore be an important goal if one wants to scale up the experiment.

Another difference between the two cases is the difference in stakeholder inclusion. Where the experiment in Burgas succeeded in involving less privileged stakeholder groups, the experiment in Antwerp did not. It can be argued that the difference between a top-down and a bottom-up approach could have contributed to this difference. A bottom-up approach requires a high level of commitment from the citizens, which might act as a barrier for participation. An observation should be made, however, that, despite the exclusion of less privileged stakeholder groups in the experiment in Antwerp, efforts were made to ensure the experiment was responsive to their needs.

Some observations could be made about the input features. The most dominant of which is the difference in social capital between the two cases. Where the citizens in Antwerp had a lot of experience with and knowledge about bottom-up experiments, the citizens in Burgas had none. This raises the question to what degree the outcomes of the experiments can be contributed to the design and governance of the experiments and to what degree the outcomes can be contributed to this context feature.

Some remarkable differences between the input features and their impact on the experiment could also be observed. Where the lack of trust posed a barrier for collaboration in Burgas, it acted as a driver in Antwerp. Similarly, the citizens in Burgas felt encouraged to participate in the experiment by the high level of institutional alignment, whereas the institutional misalignment in Antwerp achieved the same goal. It could also be observed that a high level of awareness of flood risks could act as a driver for participation in the experiment, but that a lack of awareness does not need to be a barrier for participation, as long as the experiment can be linked to other issues that the neighborhood's residents are aware of. These are all interesting findings that warrant further research into the impact of input features on the activities and outputs.

5.3. Reflections on the research findings

The case study results are a first attempt at analyzing the governance of transition experiments in such a comprehensive manner. As a result of these pioneering efforts, limited literature is available to compare the results to. Nevertheless, theories and findings from different fields of research could be linked to some of the outcomes of the case studies. For the empowerment of stakeholders and the diffusion of innovation, the importance of learning-by-doing and social proof has already been extensively researched. Similarly, the importance of good governance has also been supported by a large body of literature. Furthermore, the replication of patterns and outcomes in both cases suggests some level of external validity that should allow for the generalization of the findings. One should emphasize, however, that the differences between the context of the findings in the literature and this research indicate that this is by no means definitive proof of a relation between the literature and the research outcomes. Due to the small number of cases, the replicated patterns observed could be a result of chance rather than causal relationships. More research into both the relationship between the experiments and the different bodies of literature is therefore required, as well as an increased number of case studies researching the role of governance in transition experiments.

Despite these limitations, the tentative conclusion drawn from this research is that governance features do indeed play a role in the outcomes of a transition experiment, especially influencing the social change resulting from participating in the experiment. Two hypotheses can be posed based on the findings that would warrant future research:

H1:   The more responsibility is transferred to the experiment participants, the larger the

social change that can be achieved.

H2:   *The more principles of good governance safeguarded in the experiment, the larger the social change that can be achieved.*

By researching these hypotheses, the 'black box' of transition experiments and their outcomes can be unpacked further. Both hypotheses, if proven, could influence the design and management of transition experiment and could support and accelerate transitions to more resilient water systems.

In addition to these hypotheses that warrant further research, some interesting gaps in the findings could be observed. Firstly, the influence of input features, but specifically social capital, on the outcomes of the experiment should be researched to determine their influence on the design and results of the governance mechanisms. Secondly, it could be observed that similar input features had completely different impacts on the experiments and the other way around. In order to understand the workings of the experiment, further research into why these differences occur is therefore desired.

5.4. Reflections on the role of the logic model for evaluation

Aside from studying the different mechanisms, processes and institutions and their role in the experiment, the case studies served a second goal of validating the evaluation framework and its operationalization. Based on the cross-case analysis, some statements regarding the validity of the framework can be made.

Firstly, three input features were added to the framework. This research shows that each of these features can indeed be observed and appear to play a role in the governance and outcomes of the experiment. Second, it can be concluded that the framework is sensitive to many different experiment outcomes, including some that are often ignored in policy evaluations. Outcomes such as trust, built capacities and changed perceptions are all hard to measure quantitatively, but can have a significant impact on the transition process. This was emphasized by the civil servants that participated in the experiment, with one of the civil servants in Antwerp stating: "Empowerment, local knowledge and built capacities are all impressive results. But politicians have a tendency to focus on the tangible results. The question is: how can we make these results more visible?". By using this framework for the evaluation of transition experiments, these less visible results are also included in the evaluation. Additionally, the holistic nature of the framework helps to structure an abundance of information and provides a research focus that takes into consideration the many possible relations, including some less obvious ones.

However, the application of the framework also exposed some weaknesses. The most important point of critique is the size of the framework, counting 27 features, each of which can be split up into smaller sub-features. Assessing each of these features, as prescribed by Luederitz et al. (2017, p. 72), took a considerable amount of time and effort. Furthermore, the evaluation of the experiments took place based on ex-post interviews with a few representative stakeholders. In order to conduct a truly reliable evaluation, especially considering outputs dealing with stakeholder satisfaction or social change, one would need to gain input from a much larger sample of the population as well as conduct baseline studies. Especially for small-scale experiments, such as the experiments evaluated in this research, one can question if the costs of conducting such a rigorous evaluation are proportional to the costs of the experiment itself. Similarly, one can question if the proposed application across a large number of experiments is feasible with such an intensive process of evaluation.

These findings cast doubts on the claims of Luederitz et al. (2017) with regards to the envisioned broad applicability and usability of the evaluative scheme. Researchers and

practitioners planning to apply the scheme should be warned of the trade-off between broad and in-depth evaluation that is inherent to such a comprehensive scheme. Future research into different framework application methods and the possibilities to reduce the number of features included in the framework could lead to the design of a more manageable use of the evaluative scheme.

6. Concluding Remarks

In this paper, the validation of the evaluative scheme for the evaluation of sustainability transition experiments designed by Luederitz et al. (2017) has been discussed. The main goal of this research was to identify an evaluation method for sustainability experiments focusing on urban flood risk management. A sub goal of identifying governance practices that contribute to the success of these types of experiments was embedded within this main goal. After identifying criteria to operationalize the framework, the framework was applied to two real-world experiments taking place in Belgium and Bulgaria. Both semi-structured interviews and document analysis were used to gain data, which was then subjected to thematic analysis. By comparing the patterns identified within the cases to each other and the evaluative scheme, some interesting observations could be made.

Firstly, in both cases, the importance of learning-by-doing was emphasized. Hands-on experience with the different aspects of the experiment helped the experiment participants gain skills, knowledge and confidence they can apply when transferring or scaling up the experiment. For the empowerment of stakeholders, it is therefore important to set up the experiment in such a way that power and responsibilities are equally shared between the different stakeholders. Secondly, in both cases the experiment participants reported that good governance practices, such as transparency and inclusion, helped build trust and commitment. For both observations, the findings align with existing theories in different scientific fields, but their role in transition experiments had not yet been analyzed. These findings therefore provide interesting directions for further research into the governance of transition experiments.

As for the validity of the framework, this research proved that it does assist in evaluating and analyzing transition experiments. However, it should be noted that the size of the framework negatively impacts its manageability. Its usefulness for large-scale cross-evaluation of different experiments or the evaluation of small-scale experiments can therefore be questioned. It would be advisable to apply the framework as an interpretivist rather than a positivist evaluation method.

Some evaluative scheme-specific recommendations for further research can therefore be made. Firstly, the use of the evaluative scheme for the ex-ante design of experiments and evaluations still needs to be validated. Secondly, the continued application of the scheme to real-world cases can help refine the framework's features and identify features that should be added or excluded from the scheme. This includes the validation of the three added input features. Furthermore, more research into the governance of transition experiments is needed. This research implicated that a relationship between learning-by-doing and the experiment outcomes exist. The strength of this relationship needs to be further explored. Similarly, the role of good governance in transition experiments is a second avenue for further research. Last, the observed differences between the impact of the input features between the two cases provide an interesting discrepancy that warrants further research. Overall, more research is needed into the necessary conditions for successful transition experiments and the role of experiment governance therein.