Article

Stimulating Learning through Policy Experimentation: A Multi-Case Analysis of How Design Influences Policy Learning Outcomes in Experiments for Climate Adaptation

Belinda McFadgen [1,*] and Dave Huitema [2,3] [ID]

[1]   Institute for Environmental Studies (IVM), VU University Amsterdam, De Boelelaan 1085,
      1081 HV Amsterdam, The Netherlands

[2]   Deputy Department Head Environmental Policy Analysis, Institute for Environmental Studies (IVM),
      VU University Amsterdam, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands; dave.huitema@vu.nl

[3]   Department of Science, Faculty of Management, Science, and Technology, Netherlands Open University,
      Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

*   Correspondence: belinda.mcfadgen@vu.nl

Abstract: Learning from policy experimentation is a promising way to approach the "wicked problem" of climate adaptation, which is characterised by knowledge gaps and contested understandings of future risk. However, although the role of learning in shaping public policy is well understood, and experiments are expected to facilitate learning, little is known about how experiments produce learning, what types of learning, and how they can be designed to enhance learning effects. Using quantitative research methods, we explore how design choices influence the learning experiences of 173 participants in 18 policy experiments conducted in the Netherlands between 1997 and 2016. The experiments are divided into three "ideal types" that are expected to produce different levels and types of learning. The findings show that policy experiments produce cognitive and relational learning effects, but less normative learning, and experiment design influenced three of six measured dimensions of learning, especially the cognitive learning dimensions. This reveals a trade-off between designing for knowledge development and designing for normative or relational changes; choices that experiment designers should make in the context of their adaptation problem. Our findings also show the role leadership plays in building trust.

Keywords: policy learning; policy experiments; climate adaptation; science-policy interface

1. Introduction

Many of the climate adaptation issues emerging to challenge modern society revolve around the threat and management of water. Sea level rise, flooding, water variability, and drought are all environmental problems exacerbated by climate change, and they require swift, innovative and effective solutions. To identify solutions that work, it is suggested that policy actors employ a "learning-by-doing" approach, where an idea is executed and evaluated to understand its impacts and reduce uncertainty [1]. Climate adaptation requires the production of new knowledge to understand the impacts of climate change, new knowledge about the impacts on the ecological system of society's response to those changes, and insights into how actors perceive and understand the changes that are happening [2]. Governance systems that enable learning will make better use of this knowledge and understanding and build adaptive capacity [3], improve their decision making, and potentially enable policy change [4,5].

Experimentation is also considered a key component of adaptive co-management [3,6]. There is considerable conceptual divergence for the notion of experimentation in environmental governance

(Ansell and Bartenberger [7]) and this paper focuses on the "policy experiment" [8]. As an ex-ante form of policy appraisal, policy experiments are used to test innovative climate policy solutions in the real world. There are several interpretations of the concept (that we discuss in section two), but relevant to us is Lee [1], who describes experiments for policy development as a "mode of learning" because they explicitly produce new knowledge for political decision making, particularly for environmental and social issues [6,8–12]. Although the characteristic flexibility of experimentation generally has the potential to assist in successful adaptation governance, where policy development has so far been defined by controversy, uncertainty, and long-time frames [13], experiments vary in purpose and design [7]. The aim of this paper is to explore whether policy experiments produce learning, and if so, how learning outcomes vary in different types of experiments, so we can draw conclusions regarding to what extent policy experiments can be designed to maximise learning for climate adaptation.

To date, conceptual and empirical work on policy learning and experimentation has been mostly limited to in-depth qualitative studies and theoretical discussion, which has most commonly found that while experiments produce new knowledge and understanding of how innovations affect the social, technical, and/or ecological systems, deeper normative learning changes are lacking in experiments [7,9,14–17]. Learning scholars have made concerted efforts to find out how learning is produced in broader collective settings and have demonstrated that both agent-based and process factors influence learning outcomes; for example, who is involved, the organiser's competence, the sort of information that is produced and how it is made available, the use of technology, and the extent of representativeness [5,18–21]. The insights from these studies into the factors that encourage learning are valuable, but multi-case comparisons are still needed to empirically test hypotheses on how learning is produced [22] and no work of this kind has been performed to evaluate learning from experiments [23].

Based on these knowledge gaps, we analyse a set of eighteen real-world policy experiments that were conducted in the Netherlands between 1997 and 2016 to answer the research question: In the context of climate adaptation, what is the relationship between a policy experiment's design and the types and levels of policy learning produced? This question can be broken down into a set of sub-questions:

1.   Do policy experiments produce policy learning, and if so, what types of policy learning?
2.   Do differently designed experiments produce different learning effects?
3.   To what extent does governance design explain the levels and types of learning produced?
4.   What are the implications of the findings for climate adaptation?

To answer these questions, we conducted a quantitative analysis to test hypotheses developed about the relationship between policy learning and the governance design of the experiments. We analysed extensive survey data from 173 people who participated in the set of policy experiments, which assessed new policy initiatives relevant to climate adaptation in the Netherlands (where climate change is a national priority). The analytical framework employs a multi-dimensional learning typology (cognitive, normative, and relational learning) and the experiments are grouped into three "ideal types" of policy experiment: the "technocratic experiment" which is populated and controlled by technocrats, the "advocacy experiment" which contains a broad set of like-minded actors controlled by a small group trying to push a certain idea, and the "boundary experiment", which is inclusive and egalitarian with a broad set of actors [23,24]. Based on theoretical assumptions from the learning literature, the hypotheses examine to what extent the experiment types produce varying amounts of cognitive, normative, and relational learning.

The next section of this paper sets out the definitions and typologies of policy learning and policy experiments used in the study. The hypotheses about the relationship between the two typologies are then explained, followed by a description of the 18 experiments and how the cases were assessed. The section following then provides an explanation of the data collection and survey methods. Results are presented, and form the basis of discussion on implications of the findings for learning and policy theory, as well as practical advice for organisers in the adaption field who are considering using experiments.

2. Theoretical Framework

This section first defines policy learning and describes the learning typology, followed by descriptions of the three-policy experiment ideal types.

2.1. Definition and Typology of Policy Learning

We start from Sabatier's definition of policy learning as: "relatively enduring alterations of thought or behavioural intentions that result from experience and that are concerned with the attainment (or revision) of public policy" [25]. This definition is applied at the level of an individual who has a bearing on public policy decision making but works in a collective setting [26]. In this study, it is the experiment participants who learn, and they can be one of five actor types: a policy actor, expert, business actor, NGO representative, or a private citizen. We draw on three types of learning: cognitive, normative, and relational (Table 1), which are cognitive or relational changes, as opposed to changes in behaviour or actions (e.g., new policies, strategies, etc.). As defined by Haug and others [27], cognitive learning can refer either to an individual's gain in knowledge or to greater structuring of existing knowledge. Cognitive learning in experiments includes changes in understanding about feedbacks and key relationships between humans and biophysical systems [15] and the discovery of previously unknown effects [12]. Normative learning is defined as a change in an individual's values, goals, or belief systems, such as a shift in a participant's perspective on the issues surrounding the experiment, or the development of converging goals among participants. Like second-order or conceptual learning, normative learning is considered vital to bring about systemic change [27]. Relational learning refers to the non-cognitive aspects of learning improvements in understanding of other participants' mindsets and an increase in trust and cooperation within the group, which gives a participant a sense of fairness and ownership over the process that in turn may increase acceptance of the new management approach [28,29]. A list of factors derived from the literature that are expected to have a positive influence on these learning effects is summarised in Table 1 and discussed in Section 2.4. The factors are drawn from several sources [3,5,18–21].

Table 1. A typology of learning with definitions and factors from the literature said to enhance the different learning types.

Learning Effect	Definition	Influencing Factors
Cognitive	knowledge acquisition; improved structuring of existing knowledge	Exchange of information open and sufficient; technical competency; diverse information from a range of participants.
Normative	change in perspectives; goal convergence	Diversity of actors to share perspectives; facilitation; discussion about participants' goals.
Relational	increase in understanding of others' mind-sets; increase trust and cooperation	Participants control the process/joint fact-finding; consensus decision making; buy-in to the process; facilitation; option to engage in process; open communication.

This typology has been used in several empirical studies to conceptualise and measure learning in collective settings relevant to environmental governance [3,27,30,31]. The first two learning types resonate strongly with the policy learning literature [27] whereas relational learning reflects the notions of understanding others' roles and capacities, which are developed in the social learning literature [28,32]. We use the typology here because it draws clear distinctions useful for empirical analysis and it separately categorises relational learning, which has previously been subsumed under normative or "higher forms" of learning [32].

2.2. Definition of a Policy Experiment

Historically, the notion of policy experimentation can be traced back to Dewey's classic 'The Public and Its Problems' which notes that policies could be "-experimental in the sense that they will be

entertained subject to constant and well-equipped observation of the consequences they entail . . . " [33]. This idea was later developed by Campbell, who challenged the taking of policy decisions without risk of criticism or failure [34]. He advocated policy evaluation, with fully experimental and quasi-experimental approaches, to gather evidence on the viability of proposed policy reform [8]. Experimental evaluation gained traction, and in the following decades randomised control trial (RCT) experiments were conducted to improve economic, health, development and education policy, particularly in (but not limited to) the US and UK [35].

However, the notion of an "experiment" does not always refer to the research methodology [6]. For example, from a planning perspective, experimentation is also understood as policy development in exploratory, incremental steps [36] and in the last couple of decades, adaptive (co)management and transition management approaches to environmental governance have developed and they have their own ideas of what it means to experiment [1,9,10]. Adaptive (co)management understands experimentation as a process that explores new ideas by testing them and using the results to refine the proposal under conditions of uncertainty, and transition management views experimentation as protected niche spaces where new innovations can emerge [37]. Transition management also informs the notion of experimentation in climate governance, which considers an experiment a radical invention, a novel improvement to existing policy action that exists outside the political status quo and seeks to change it [38,39].

As we require a set of policy experiments for systematic analysis, clear analytical categories to identify what is and what is not an experiment are needed. To pay regard to the different conceptualisations outlined above, it is posited here that policy experiments should be novel and innovative, but also able to play an evaluation role in environmental governance [23]. In line with this, and despite being wary that analytical divergence is all but a given, a definition of a policy experiment is proposed that we believe captures its important characteristics: "a temporary, controlled field-trial of a policy-relevant innovation that produces evidence for subsequent policy decisions".

By emphasising the role as producers of policy evidence, this definition characterises an experiment as a temporary science-policy interface. Connections between experiments and policy can be either direct (implementation requested by policy-makers) or indirect (results eventually inform decisions on policy options). Either way, the goal is for the experiment to create some form of policy learning through testing new policy instruments or concepts. The criterion requiring a policy experiment be "controlled" includes the attempts to form hypotheses and evaluate against expectations as a form of control (a "quasi-experiment" [8]). Enforcing the more stringent requirement of a "control group" would reduce the sample considerably, as they are very rare in environmental management [40].

2.3. Policy Experiment "Ideal Types" Typology

Theories in the policy science and science-technology studies (STS) literatures explain different forms of policy development and roles of science in policy making, which can be grouped into three models: the expert driven "technocratic" model, the participatory "boundary" model, and the political "-advocacy" model [24,41,42]. These models differ in their governance design and we use them as the basis of a neutral "ideal type" typology for policy experiments [26] (following German sociologist Max Weber's conceptualisation of an "ideal type" [43]). These choices include: what actors are involved; how authority is distributed; and what information is generated and shared. These design choices are based on the institutional rules set out in chapter seven of Ostrom's book "Understanding Institutional Diversity" [44]; the boundary, position, information, pay-off, and aggregation rules [42]. The focus on design provides a comprehensive and functional-analytical framework that relies on theoretical propositions about the production of learning [23]. Below we summarise the governance design of each ideal type, and Table 2 sets out their characteristics in more detail. These ideal type categorisations are theoretically derived and we note that no real-life case will ever perfectly match any one type—there will always be a degree of non-conformity when assigning cases to a type [23,40].

Table 2. Comparison of the characteristics of policy experiment ideal types.

Action Rules	Design Choice Indicator	Technocratic Experiment	Boundary Experiment	Advocacy Experiment
BOUNDARY	*Actor Constellation*	Expert actors	All actor types involved	Predominantly members of an advocacy coalition
	Access to experiment	Required	Those requesting involvement	Those invited by initiator
	Criteria for new participants	Expert actors	Those with local and/or expert knowledge	Those in support or who will build support for experiment
POSITION	*Initiator role*	Expert actors	Collaborators	Policy actors
	Use of facilitator	None	Yes (and neutral)	If yes, then with/for core members only
INFORMATION	*Contribution to goals*	None—already set by policy makers	By all actors	By actors who are in agreement
	Lay knowledge acknowledged/accepted	No	Yes, to a large degree	Yes, but not solely
	Scientific knowledge acknowledged/accepted	Exclusively	As one of many inputs	Only if from scientists within the coalition
	Information transmission	Information for majority	Information for all participants	Information for minority
	Opportunities for personal contact between scientists and policy makers	Few	Frequent	Very frequent but only within the group
	Outsiders informed of progress	Occasionally	Frequently	Rarely
CHOICE	*Authority at decision nodes*	Expert initiators	Participants share power	Policy initiators
	Variation in authority	Most actors have advisory role	Most actors have decision role	Most actors have no authority
PAY-OFF	*How costs distributed*	Minimal buy-in	Buy-in	No buy-in
AGGREGATION	*How decisions are made*	By experts in majority (in line with scientific methods)	Everyone by consensus (on basis of deliberation)	Policy actor by majority (on basis of shared principles)

2.3.1. Technocratic Ideal Type

The technocratic experiment represents an instrumental means to policy problem solving by generating (assumedly) objective knowledge for policy development, which is independent of its context or subjects. Organisers intend a separation of power between the experts who participate in, design, monitor, and evaluate the experiment, and the policy actors who make decisions based on the evidence produced [41]. Policy makers typically commission the experiment because they need evidence to support or refute a claim and end a political disagreement; they are absent from the experiment other than possibly providing funding and framing the issue to be studied. Construction of the policy problem and solution to be tested is determined by policy actors in advance, so the appropriateness of policy goals is not discussed or debated [14].

2.3.2. Boundary Ideal Type

In contrast, a boundary experiment represents a participatory and dialectical approach to policy appraisal that focuses not only on producing evidence but also on debating norms and developing shared values among participants. Organisers design a boundary experiment when they want to maximise the involvement of different actors in policy development. Participant diversity brings multiple knowledge types: scientific knowledge, practical knowledge, and traditional, lay knowledge, so both scientific and non-expert knowledge is utilised. Experiment results are verified by the range of actors who address both policy and local community needs [45].

2.3.3. The Advocacy Ideal Type Experiment

With this design, the organisers intend to push action in a policy direction by using the experiment as a "proof of principle" [46], for softening objections to a predefined decision [41], or as a tool to delay making final decisions (see the "stealth advocate" role in Pielke [24] for comparison). An experiment serves these tactics because a reversible temporary change provides a sense of security, and the change involved may meet with less resistance [47]. An advocacy experiment will mostly be initiated and dominated by policy actors, who may exclusively invite other actor types if they support the initiative, but non-state actors can initiate these sorts of experiments too. Those with authority retain control over design, monitoring, and evaluation procedures, reinforcing the existing structures of power in policy making.

2.4. Hypotheses on the Generation of Different Learning Effects by Different Ideal Types of Experiment

The literature on learning explores the institutional dimensions of a process to understand what sort of learning is produced and what factors are most important [19]. It is expected, for example, that a process where the participants have different backgrounds and perspectives (participant diversity) is more likely to produce a change in an individual's perspective than one with a homogenous set of actors [21]. Similarly, a process where participants share tasks and responsibilities and have an equal say over proceedings (joint decision making) may produce higher levels of relational learning [18] than processes where decision making is less inclusive. The degree of distribution and openness of information transmission and the extent of technical competence of individuals may also influence the extent of knowledge acquisition [5,20]. Opportunities to discuss opinions on goals and perspectives could lead to the development of common goals [20]. Based on this theory, our hypotheses predict relationships between the governance design of policy experiments and learning. We now discuss the anticipated learning effects per type of experiment in turn.

The *technocratic experiment* is hypothesised to generate high levels of cognitive learning, little normative, and some relational learning. The focus of this ideal type on instrumental rationalisation means the experiment is expected to produce large amounts of data regarding the impacts of the intervention which will be shared openly and regularly to all participants who will use it to increase their stocks of knowledge and restructure their existing understanding of the issues. However, the disconnection of the

experiment from the policy process and lack of an opportunity to decide on policy goals means the level of normative learning is predicted to be low or absent. The lack of participant diversity also means there will be no necessity to align interests and goals. Thus, some relational learning may be produced but it is likely to be limited to that between experts who build trust and understanding in a scientific capacity. Nevertheless, the open information distribution and partially shared authority improve the chances that participants will actively cooperate, possibly contributing to some relational learning.

For *boundary experiments*, we expect high relational and normative learning effects but only some cognitive learning. The design choices allow any actor who wants to be included to be involved. This ensures a mix of interests and knowledge, diverse views and perspectives, which are likely to encourage participants to reconsider their own priorities and views, and possibly to develop a common interest. Open communication, regular information distribution, and use of a facilitator to manage proceedings increase the chance that actors get to know one another and understand how others perceive the policy problem. Sharing authority and costs is likely to increase a sense of buy-in, and therefore a participant's trust and cooperation within the group. However, the diversity and focus on relationships means the uptake of knowledge may suffer somewhat; the consideration of different sorts of knowledge, however, may ensure a deeper understanding of system complexity.

In *advocacy experiments,* we hypothesise that some cognitive and normative learning will emerge, but little relational learning. Some knowledge acquisition is expected since expert and non-expert actors contribute their knowledge and skills; however, cognitive learning will be limited by the restricted information distribution and lack of open communication among participants. Some normative learning is expected, due to the sharing within the core group of views on the experiment goals; however, views are likely to be aligned due to the restriction of access to favoured participants. In contrast, restricted information distribution, the lack of buy-in, and rigid hierarchy of authority means trust and cooperation are likely to stagnate; the homogeneity of views is likely to offer little chance to understand alternative views, also contributing to low relational learning.

In sum, we posit that the ideal types will produce variant patterns of learning due to their different designs. The conceptual assumptions explained above are summarised in Table 3. We are particularly interested in whether the types produce significantly different scores for each learning type, allowing us to draw conclusions about the importance of experiment design. To test our hypotheses, 18 experiments were identified and analysed using the methods and results presented in the sections below.

Table 3. Expected learning effects for ideal type experiments.

Title	Cognitive Learning	Normative Learning	Relational Learning
H1: Technocratic type	High	Low	Medium
H2: Boundary type	Medium	High	High
H3: Advocacy type	Medium	Medium	Low

2.5. Role of Intervening Variables

While our hypotheses focus on how governance design facilitates learning, the literature also refers to non-institutional variables that may have a bearing on how participants in experiments learn. Five intervening variables are identified and analysed for this study. First, learning from an experiment may stem from the nature of experimentation itself, that is, the choice to test an innovation in practice. Here, it is expected that the new and uncertain environment, whether stemming from a policy crisis or the idea's novelty, will motivate participants to assess new information [48]. A second consideration, the charisma or competence of an organiser, is also considered notably influential in these situations. Charismatic leaders can produce a group environment that shares information openly and broadly, shapes shared values, and ensure new ideas are nurtured [19]. Three other intervening variables focus on the participants themselves: actor type, which generally indicates the competence of participants in

understanding the material; participant demographics [5]; and the extent to which participants knew each other previously.

3. Methods

This section sets out how the experiment cases were selected and what data collection methods were used. It also details how the sample cases were matched against the ideal type categories.

3.1. Case Selection

The search for adaptation-relevant policy experiments was conducted throughout various institutions of the Netherlands with a focus particularly on the country's water authorities ("water boards"). The water boards sit at the regional level between local and provincial government. Researchers conducted a semi-structured search and data collection through government websites and research programmes between February and November 2013. The search included phrases such as: test, pilot, innovation, and experiment, "*proef*", "*onderzoek*" (*test* and *research* respectively in Dutch). The search was national in scope, accessing research programme websites, ministry, province and water board websites, and projects mentioned in scoping interviews. The topic of water issues affected by climate change was selected to provide consistency of comparison of experiment subjects. Such issues are increasingly understood as a matter of urgency in the Netherlands, which is a lowland country particularly vulnerable to sea-level rise, flooding, salt-water intrusion, fresh water availability, and increased drought. The water boards' main responsibilities are the maintenance of dikes and dams, water quantity and water quality.

Our "policy experiment" definition was operationalised through the three criteria used to identify experiments: whether the project was testing for real-world effects in situ (temporary, controlled field trial); whether it was innovative with uncertain outcomes (innovation); whether its findings were intended to have relevance for policy (evidence for decisions). Projects deemed outside our definition included product testing, concept or scenario pilots, modelling projects, and reapplications of an initial experiment. In addition, for consistency, experiments were selected only where the intervention related to climate change adaptation, where there was state involvement, and where an ecosystem response was elicited (see Table 4). The initial search identified 147 innovative pilot cases (list available on request) with 18 cases meeting all six criteria. The cases have different spatial and temporal scales and deal with different problems related to the larger topic. However, they are comparable because they conform to the above stringent criteria.

Table 4. Criteria and associated indicators used to identify policy experiments in climate change adaptation in the Netherlands.

Criteria	Indicators	Relevance to Definition
Testing for real-world effects	In-situ intervention with monitoring and evaluation framework	Temporary "controlled" field trial
Innovation	Previously untried policy or management practice	Innovative intervention with uncertain outcomes
Policy relevance	Test of policy concept or approach	Produces evidence for policy decisions
State involvement	Organiser or other participatory role played by an actor employed by state or state agency	
Ecosystem response	Intervention extends across social-ecological system	
Climate change adaptation focus	Exploring new policy concepts to manage sea-level rise, flooding, fresh water availability, and increased drought	

Policy Experiment Cases

Adaptation involves searching for technical solutions, such as enhancing dikes or increasing water-storage capacity, as well as governance solutions, like reforming land-use planning, efficient water use, or agricultural transitions. We had a sample of eighteen policy experiments that tested the viability of proposed policy innovations relevant to Dutch climate change adaptation. There were five coastal management experiments, five water storage experiments, three freshwater experiments, three water variability experiments, and two dike management experiments in the sample. Ten experiments tested technical innovations (the application of a technical solution on the ecological system to measure its impacts); four experiments tested governance innovations (the application of a governance solution on the social and ecological system); and four experiments trialled both [23].

The experiments in our sample trialled and evaluated several new policy concepts. The "Building with Nature" concept (a design philosophy that utilises the forces of nature to meet water management goals) was tested in three experiments, the multi-functional land use concept tested five times, and three cases experimented with shared responsibility for water resources (i.e., passing responsibility for water management onto farmers). One experiment looked at pest management to minimise damage to inland dikes, and another tested the "Climate Buffers" concept (climate buffers are natural areas specially designed to reduce the consequences of climate change). In one experiment, De Kerf, examined the "Dynamic Coastal Management" concept, which is explained in more detail in [49]. Further information about the policy issue and intervention for each experiment case is provided in Appendix D.

All completed experiments were included in the sample, whereas ongoing cases were included if they had been implemented for at least two years or were subject to an interim evaluation [50]. Their start dates range between 1997 and January 2013; six were ongoing as of September 2016, but had reached intermediate conclusions that allowed assessment of their learning effects.

3.2. Data Collection

People who were actively involved in the experiments were considered experiment participants and were identified during interviews with project leaders and checked against project reports where possible. Participant numbers varied across cases, the lowest eight and the highest 40. Each participant was one of five different types of actor: policy actors ($n = 84$), experts ($n = 39$), business actors ($n = 16$); NGO representatives ($n = 16$); and private citizens ($n = 13$). 73 out of the total 173 participants claimed to be an initiator of the experiment they were involved in.

In April–June 2014, these experiment participants were sent (via email) an online survey that asked about their role in the experiment, their opinions on design aspects, and questions to gauge their learning experiences. Three reminder emails were sent at weekly intervals. From a total of 265 survey emails, 173 were completed, giving a 64% response rate. Each case had either a minimum of 6 responses or responses from at least half the participants (minimum 4). The respondents were asked a total of 63 survey questions. To determine the cases' design, 34 factual and attitudinal questions were asked; including specific questions about the role of the respondent, the information they contributed, the extent of their authority, and their role in financing the project.

We chose an ex post, self-reported learning approach to assess learning (following Baird and others [3], Muro and Jeffrey [21], Schulser and others [51], and Leach and others [5]). Two variables were measured for each of the three learning types, six variables in total. Two questions were asked for each learning variable (12 questions altogether) but only two variables had questions that reliably measured the construct (i.e., met the requirement of 0.7 Cronbach alpha score) so four questions were removed from the analysis. Therefore, our learning data consists of eight questions measuring six variables. The questions are listed in Appendix B and translated from Dutch. The questions were measured on a five-point scale, in line with previous learning studies [5,21,51]. Using 6–12 questions follows the Muro and Jeffrey [21] and Leach and others [5] learning assessments. Eleven questions were also asked to measure the intervening variables discussed in Section 2.5.

Although the survey data was collected from 173 respondents, the unit of analysis in this study is the experiment because we are comparing cases, so we averaged the respondents' scores for each case.

3.3. Matching Experiments to Ideal Types

To assign the cases to its ideal type, each case was individually assessed using fifteen indicators and subsequently assigned an ideal type. The indicators are based on the institutional rules described by Ostrom [44] and allow us to assess each case's institutional design, including actor constellation, variation in authority, extent of information distribution, and openness (see Appendix A for a detailed breakdown of indicators). Each indicator was assigned three "settings"—a description of measurable action related to each indicator for each ideal type. Using survey data, the experiments were assessed against the indicator settings and each case was labelled the type that its scores matched best (all but three experiments displayed characteristics of one, dominant type). A dominant type was assumed if one type had a majority of more than three indicators over the other types. For example, out of 15 indicators, experiment 2 scored one for technocratic, nine for boundary, and five for advocacy; thus, it was classified clearly as a boundary experiment. The assessment resulted in five technocratic, six boundary, and seven advocacy experiments.

4. Results

This section first presents the entire sample's learning results, then a breakdown of the scores for comparison by ideal type. This is followed by a Kruskall-Wallis H non-parametric statistical analysis to assess the relationship between structure and learning outcomes, and the role of intervening variables.

4.1. Overall Learning Results

Figure 1 shows the mean scores for the variables measuring the six dimensions of learning. Cognitive learning was highest, with "new knowledge" reaching a high score and "restructuring knowledge" a medium score. Normative learning scored noticeably lower; "goal convergence" scored medium but there was a definite lack of "priority change" across all cases. Relational learning displayed a similar pattern, with strong "trust building" but no recorded "understanding mindsets".
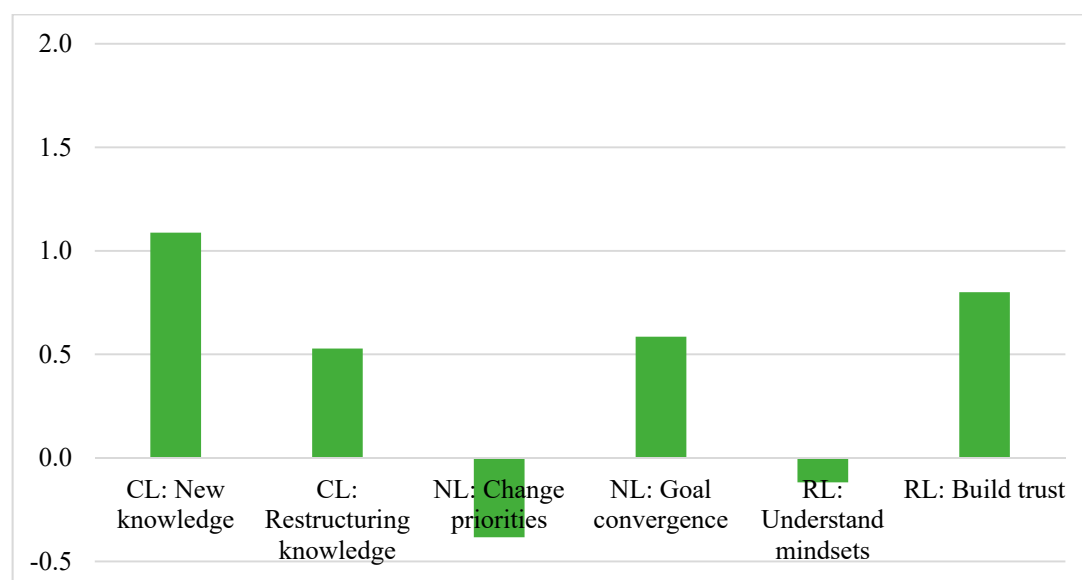

Figure 1. Mean scores for six measured learning variables across the entire experiment sample. High ≥ 1; medium = 0.5–1; low = 0–0.49; and none ≤ 0.

4.2. Learning Patterns in Types and Their Significance

To test the hypotheses that ideal types produce different levels of learning, the learning scores for the policy experiment cases in each type were compared and a Kruskal-Wallis H test (K-WH) was performed to assess the significance of the differences in scores. The Kruskal-Wallis H test is a non-parametric test that is used to determine if there are statistically significant differences in the distributions of an ordinal dependent variable (learning) between three or more groups of an independent nominal variable (experiment types). If the K-WH tests revealed significant differences in learning scores between the experiment types, then this provides evidence that design influences learning outcomes.

Starting with Figure 2, we find that, as expected, the four technocratic experiments produced more knowledge and knowledge restructuring than experiments in the other two types. Scores for the knowledge restructuring dimension were lower than for knowledge acquisition, with boundary experiments doing noticeably worse than the other types. When comparing the groups of cases, the KW tests were significant for both cognitive learning variables, confirming that technocratic experiments produce more cognitive learning than the other experiment types (see Appendix C for results).
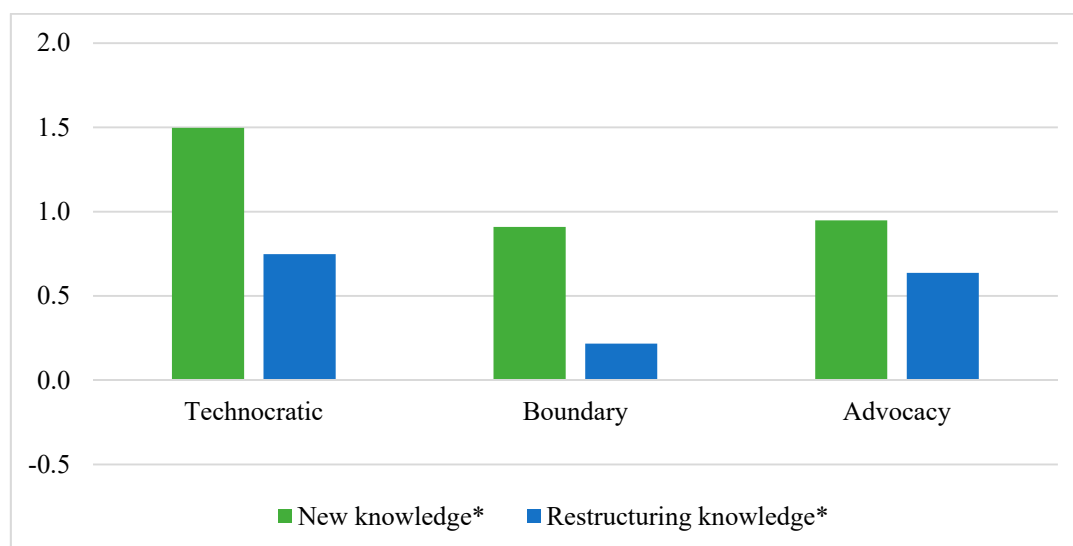

Figure 2. Comparing mean scores between ideal types, for both cognitive learning variables ($n = 18$). High ≥ 1; medium = 0.5–1; low = 0–0.49; and none ≤ 0. The asterisk (*) denotes that for this variable, there were statistically significant differences in scores between the cases, as revealed by the Kruskal-Wallis test.

Figure 3 sets out the normative learning scores for the experiment types. All experiment types record noticeably lower normative learning, with a "change in priorities" clearly not a product of any experiment type. Boundary experiments scored medium for "goal convergence"; higher than the other types, but lower than expected. Technocratic experiments also produced medium levels of "goal convergence", which was unexpected. When comparing the learning scores of each experiment type using the Kruskal-Wallis test, no significantly different scores were found (see Appendix C for results).
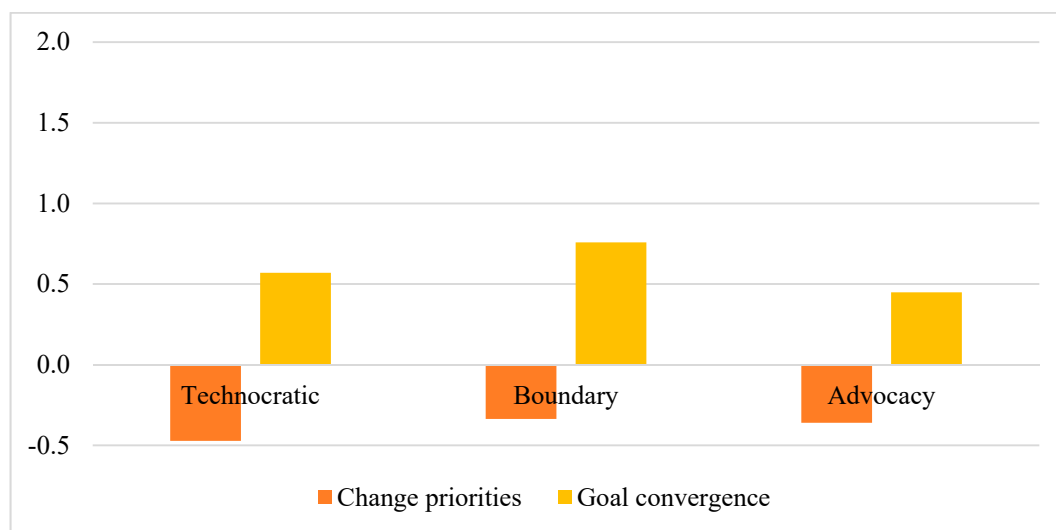
Figure 3. Comparing mean scores between ideal types, for both normative learning variables (*n* = 18). High ≥ 1; medium = 0.5–1; low = 0–0.49; and none ≤ 0.

Finally, Figure 4 sets out the results for the relational learning variables between experiment types. Again, one measured dimension of learning is stronger than the other, in this case it is "building trust" which scores well and the "understanding mindsets" which lags. Advocacy experiments recorded a surprisingly moderate amount of trust built among participants, with boundary experiments not scoring that much higher. All three types scored poorly for the understanding mindsets variable than expected, although the Kruskal-Wallis test reveals that despite the low score, participants in the boundary experiments learned to understand one others' mindsets significantly more than participants in the other experiment types (see Appendix C for results).
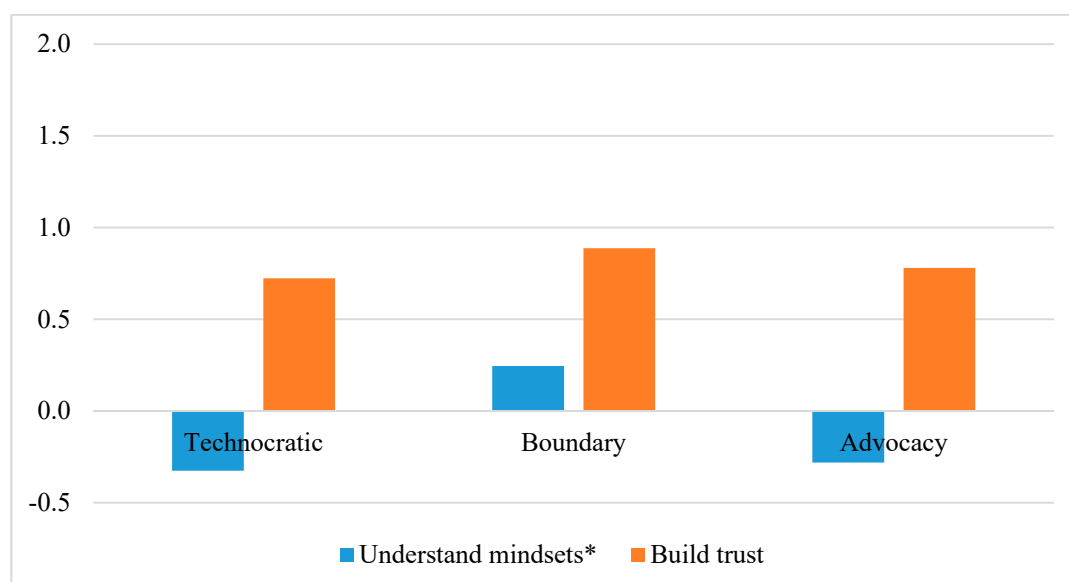

Figure 4. Comparing mean scores between ideal types, for both relational learning variables (*n* = 18). High ≥ 1; medium = 0.5–1; low = 0–0.49; and none ≤ 0. The asterisk (*) denotes that for this variable, there were statistically significant differences in scores between the cases, as revealed by the Kruskal-Wallis test.

As indicated by the asterisks in the Figures, the patterns we observed for three learning dimensions (both cognitive learning variables and one relational learning variable) appear to be statistically significant. This provides evidence that some of the variance in scores can be attributed to differences between types (see Appendix C for statistics). Reviewing the K-W H post-hoc tests (which statistically inform us which type scores significantly higher than the others) we observe that technocratic experiments produce significantly more knowledge than both other types, and significantly more knowledge restructuring than boundary experiments; and boundary experiments encourage the understanding of others' mindsets significantly more than advocacy types.

4.3. Influence of Intervening Variables

Data were collected from participants for the five intervening variables described in Section 2.5 that could also potentially influence the differences in how much and what type of learning was produced. There were three ordinal independent variables (extent the policy problem was urgent; leader competency; extent participants already knew each other) and two nominal independent variables (demographics-age and sex; actor type). The relationship between the ordinal independent variables and the six ordinal learning variables was measured using the Somers' d nonparametric test (SMD). Kruskal-Wallis H tests (KWH) were conducted to measure the relationship between the two nominal independent variables and the learning variables.

Appendix C sets out the statistics for the relationship between the five intervening variables and six learning variables. For the ordinal independent variables, we found that the competence of an organiser and the extent the participants knew each other, both had a positive impact on the amount of trust that was built in the experiment cases. We would expect that experiments would be bringing actors together in new constellations, but the results demonstrate this is not often the case, with 43% (77 out of 153 participants) that responded to the question-claiming they knew over half the other participants in their experiment. The extent an experiment addressed an urgent issue did not significantly correlate with any learning variables.

KWH tests assessed the relationship between actor type and learning. No significant relationship was found between actor type and cognitive or relational learning variables, but both normative learning variables had significant correlations, and it was individual citizens who experienced the most change for these two normative learning variables. Finally, no relationships were found between age and sex of participants and any learning variables.

4.4. Returning to the Hypotheses

Table 5 summarises the hypotheses (H1, H2, H3) and findings for predictions related to the six learning variables. Out of the 18 units (six variables x three hypotheses) 12 were rejected and six were not rejected. H1 (technocratic experiments) is partially correct for all three learning types, with more normative learning occurring than expected. The significance tests confirmed that technocratic experiments are strongest in cognitive learning but particularly weak in one dimension of relational learning, in comparison with the other types. In contrast, the H2 predictions were almost all incorrect, with boundary experiments producing only medium/low levels of normative and relational learning. It is worth noting, however, that boundary experiments scored higher than the other types for both these learning types, and for two of the four variables the differences were significant. For H3, advocacy experiments met expectations for cognitive learning, but normative learning was lower than expected and relational learning was higher.

Table 5. Results for the hypothesised levels of learning for each experiment type. Asterisks (*) denote the variables that have statistically significant differences in learning levels between the types.

Hypothesis (H) for Each Experimental Type	Cognitive Learning Variables		Normative Learning Variables		Relational Learning Variables	
	New knowledge *:	Restructure knowledge *:	Priority change:	Goal convergence *	Understand mindsets *:	Build trust:
H1: Technocratic	Failure to reject	Reject	Failure to reject	Reject	Reject	Failure to reject
H2: Boundary	Failure to reject	Reject	Reject	Reject	Reject	Reject
H3: Advocacy	Failure to reject	Failure to reject	Reject	Reject	Reject	Reject

5. Discussion

This research conducted a systematic quantitative analysis of learning effects from climate adaptation experiments in the Netherlands. Being quantitative, the findings allow for broad rather than deep analysis, but our research design ensured they are thorough and robust. Our analysis determined what types of learning were affected by differences in governance design and which were more influenced by non- institutional variables. The findings provide insight relevant to actors who want to initiate an experiment to test adaptation solutions, helping them to understand the potential effects of their design choices on learning experiences. Next, we discuss how there seems to be a trade-off between experiment types in terms of the learning they produce, which indicates that the choice in design may have to partially depend on context of the problem. We then look at how testing for intervening variables reveals some interesting relationships where design is not a factor, followed by discussion on the methods we used to measure learning.

Our findings show that technocratic experiments scored highest for the cognitive learning variables, significantly higher than boundary experiments. This provides illustrative evidence that design has a strong influence on reducing uncertainty of experiment impacts. As alluded to in our hypotheses, the findings imply a trade-off between cognitive, normative, and relational learning when making design choices. When they aim to increase knowledge and understanding of the relevant social-ecological system, organisers clearly have to choose between a technocratic design, which intends to improve scientific, objective knowledge; and a participatory design, which incorporates non-expert knowledge into the experiment, intends to debate norms, and develops shared values among participants. Although a boundary experiment produces cognitive learning, this design reduces the success of acquiring and restructuring knowledge about the changes being tested in the experiment. We also found that although not reaching particularly high scores, boundary experiments produced more goal convergence and understanding of mindsets than the other two types. Climate adaptation is a wicked policy issue that involves significant uncertainty and divergent framings, as well as deep uncertainty about the rate of change in the ecological system [2]. It is reasonable to assume that future solutions will require the input of many different actors and knowledges, and the combining of different understandings. On this basis, boundary experiments are arguably more appropriate for these contexts, where relational and normative learning would be a great benefit. However, as pointed out by Owens and others [41], deliberatively designed appraisals take a lot of resources and do not guarantee success. Therefore, the choice in design should be specific to the problem context, and for solutions that have low certainty of impacts but high consensus on values, a technocratic design may suffice [24].

Organisers can try to reduce the trade-offs we identified by tweaking their experiment designs. For example, organisers of boundary and advocacy experiments could ensure expert participants take a leading role in designing and evaluating the experiments to ensure knowledge is produced and disseminated throughout the group of participants to increase knowledge acquisition (cognitive learning). Following D.T. Campbell's suggestion that a true Experimenting Society includes opponents of an innovation in the experiment itself [8], organisers of technocratic and advocacy experiments could also increase relational learning by ensuring that participants who have different interests or

opposing views are included as participants, and that these views and understandings are shared within the group.

A second finding was that there was a moderate increase in trust recorded across the types, despite experiment design, and that this perceived increase in trust was strongly influenced by competent leadership. This resonates with the observation by Gerlak and Heikkila [19] that powerful and influential leaders have a key role in learning because they facilitate communication, bring together interests, and shape shared values. This possibly explains why advocacy experiments recorded moderate amounts of trust-building despite their restrictive and elitist design. Advocacy experiments scored better than expected, and were the most common experiment type used in Dutch water management. Strong leaders might alleviate some of the frustration participants feel about not being included in decision making or information dissemination, since initiators who advocate for a particular policy proposal tend to have social acuity and be good at team building as 'policy entrepreneurs' [52]. Another factor supporting the generally high levels of trust throughout the experiment was that participants tended to know one another. This could be explained by the Netherlands relatively tight knit water policy community, as well as highlights that experiments just do not involve groups that oppose the proposals being tested. The lack of perspective change and understanding different mindsets also indicates that experiments do not seem to be used to trial radical and abrupt changes, or if they are radical changes, they have been percolating a while in society and actors are not surprised or opposed to them by the time the experiments are organised.

Finally, we found that, although all experiments produced unexpectedly low normative learning (although in line with previous learning studies [3,7,27,31]) the perspectives change dimension registered very differently among actor types: individual "citizen" actors were significantly more likely to record favourably for both normative learning variables. This result did not come through in the initial analysis because there were so few individual citizens involved ($n = 9$), who were predominantly found in the boundary experiments. This finding demonstrates that opening policy processes to those not traditionally involved can lead to a considerable learning impact for those actors. Engaging a large number of citizens in policy experiments to increase normative learning may or may not be a suitable course of action, but it is noteworthy that experiments can facilitate a potential shift in perspectives on environmental policy issues.

Limits to the Research

The limitations to the validity of learning data gathered *ex post*, via self-reported learning methods are known and accepted for this type of research. The analysis of 18 cases (173 participants) counteracts this limitation to a degree, and this research design is a rare contribution to scholarship on learning [22]. The number of cases facilitates broad exploratory analysis rather than examination of the effects on learning of finer nuances of design choices. Similarly, although derived from published studies, the learning questions used in the research are necessarily general and partially abstract due to the differences in the thematic focus of the experiments. Surveying 173 participants as potential respondents compensates for the loss in precision, but the limitation of the generality of findings remains. The use of only one-two questions to explore the six learning variables possibly also contributed a loss of precision; however, it is not unheard of for studies involving many respondents to rely on only 1–2 questions for their findings [5,21]. Moreover, the questions are as standardised and thorough as possible, both closed and open questions; for example, asking for both a factual response about a participant's authority, and for an opinion on the openness of the experiment to outsiders. Time is also a limiting factor when interpreting the learning data because not all experiments are recent. Survey respondents who participated in the experiment conducted between 1997 and 2002 provide less reliable responses than a participant in a recent experiment. Since most experiments started within the last seven years we did not control for this intervening variable, but it is a limiting factor.

Despite the comprehensive set of experiments in the study, it would be unwise to extrapolate findings from water/climate policy experiments to policy experiments generally, due to the purposive,

snow-ball sampling strategy and a lack of international comparative or cross sector comparison. That said, the findings suggest the value of continuing research along such lines, since the framework could be applied to other policy areas. A methodological limitation exists in that although most experiments clearly match the characteristics of a single ideal type, a few form a hybrid of types. Some hybridity is to be expected since the ideal types are theoretical versions of reality made up of several points of view and phenomena, as Weber intended when he developed the concept (according to Weber, "an ideal type is formed by the one-sided accentuation of one or more points of view and by the synthesis of a great many diffuse, discrete, more or less present and occasionally absent concrete individual phenomena, which are arranged . . . into a unified analytical construct") [43]. Therefore, a case would never be expected to wholly meet a type and occasionally a case might fall between two types. Finally, we note that others perceive ideal types differently; for example, by constructing them using characteristics of the cases being examined [50]. In our use, however, the matching of real world cases to theoretical constructs facilitates comparison based on theoretically derived expectations, which has been lacking in research on learning.

6. Conclusions

A difficult problem such as how to adapt to climate impacts requires an approach that focuses on learning, and experiments are an increasingly favoured mode of learning that produce evidence of the effects of an intervention to improve decision-making. This paper is the first multi-case quantitative analysis conducted to explore the relationship between these two variables and it applies an explorative theoretical framework to test predictions about how an experiment's design affects learning. This focus brings an element of political analysis to the study of experimentation and by identifying the action of experiments at the science-policy interface, we could construct three ideal types based on design choices that capture the various ways knowledge is developed and used in water and climate policy making. The learning typology used in combination with the experiment typology provides a way to conceptualise and measure learning as changes in the learner. It also serves as an alternative to the loop-learning concept, which privileges learning that involves a change to underlying assumptions, and which is difficult to apply consistently [29].

The analysis confirmed that differently designed experiments produce different types of policy learning. Design clearly influences knowledge acquisition, restructuring of existing understanding, and a change in mind-sets. In contrast, trust and changes in participants' perspectives do not vary at across the experiment types; these learning types are influenced instead by the leader's abilities and what type of actor the learner is.

To perform the analysis, we assessed policy experiments related to climate change adaptation and water management in the Netherlands. Adaptation to climate change is an emerging policy field that requires new solutions to largely intractable issues and our findings shed light on how organisers can maximise different learning effects by carefully designing their experiments. Relational learning might be crucial with a set of participants who do not know each other or have a range of backgrounds, whereas boosting cognitive learning might be an aim where there is low certainty of issues but general societal consensus on the issue. Our results show that experiments can be used to build a common goal to some extent, but they will not help to harmonise conflicting views in a group by changing perspectives and views. Other, more explicitly deliberative processes would be necessary.

Our findings provide a greater understanding of the relationship between science and policy making; in particular, the sorts of choices that must be made when designing policy experiments and how these choices influence policy learning. Insight into these issues goes a long way towards improving political decision making for climate adaptation.

Author Contributions: Belinda McFadgen and Dave Huitema co-designed the study, Belinda McFadgen collected and analysed field data, Belinda McFadgen wrote the paper with the assistance of Dave Huitema.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Rule Group	Indicator	Explanation	Settings
1A	Actor constellation	Actor constellation calculates the extent of domination by one actor type. Dominance ≥ 50%	BIT = No dominance by one actor, 4 + actors. AIT = Dominant policy actors. TIT = Dominant expert actors.
1B	How participants gained access to the experiment	How participants entered the experiment. Participants can either be invited in, involved because they are part of the organizing team, or because they requested involvement from the organisers. The most common method is used for classification.	BIT—Rules allow requested involvement AIT—P. mostly invited in TIT—P. mostly obliged
1C	Criteria for new participants	The most common criteria given for allowing access of a new participant.	BIT—provides local knowledge. AIT—supports or builds support for project. TIT—subject or process expert.
2D	Initiator type	The actor type(s) which initiated the experiment.	BIT—collaboration between more than two actor types, or two of equal number. AIT—policy actor (dominant) TIT—expert actor (dominant).
2E	Use of facilitator	Recollection of facilitator involvement.	BIT—80% + recall facilitator. TIT—0% or less than two p. (indicating a mistake).
3F	Extent that the goals of the experiment were discussed among the group	Percentage and diversity of participants that contributed to the discussion on project goals.	BIT—>80% if most actor types AIT—only policy actors TIT—<30% or just expert actors.
3G	Whether lay knowledge was contributed	Percentage of participants that (exclusively) contributed lay knowledge.	BIT—>50% (at least one exclusively) AIT—30–50% TIT—<30%
3H	Whether scientific knowledge was contributed	Percentage of participants that contributed scientific knowledge.	TIT—50% < and solely BIT—30–50% AIT—30% < with most p. providing practical k.
3I	How satisfied participants were with the information they received.	To what extent participants were satisfied with how much information they received and the relevance of it.	BIT = everyone agree was sufficient (over 1). AIT = 50% < disagree (minority found sufficient).
3J	How satisfied participants were with the extent of personal contact during the experiment.	To what extent participants felt there was sufficient personal contact among the group.	BIT = everyone agree was sufficient (over 1). AIT = 50% < disagree (minority found sufficient).
3K	Whether outsiders were informed of progress.	How regularly non-participants were informed of the experiment's progress.	BIT = 75% < regularly informed. AIT = 50% < irregularly/not informed.
4L	Actor type with authority at decision nodes.	Actor type that makes decisions at design, monitoring, and evaluation stages (aggregated).	BIT—more than two parties have DM role; non-state actor has shared or dominant role. AIT—policy actors (dominant). TIT—expert actors (dominant).
4M	Extent of variation in authority over decisions in the experiment	Comparing how many participants had authority to how many didn't have authority.	BIT—majority of participants had decision making power AIT—majority of participants had no authority TIT—if experts had decision power
5N	Extent of buy-in	Look into how the costs of the experiment were paid; to what extent participants were expected to "buy-in" to the experiment.	BIT—costs were shared. AIT—a participant's organization paid all costs or no costs. TIT—no clear distinction; some paid, some shared.
6O	Decision making model	Percentage of participants with decision making power at each of the design, monitoring, and evaluation nodes (aggregated).	BIT—50% or more of participants—e.g., majority/consensus DM AIT—0–29% of participants—e.g., hierarchical DM TIT—30–49%—e.g., appointed steering group, narrow DM.

Appendix B

The questions are translated from Dutch.

Cognitive Learning	
Gain knowledge A	To what extent do you agree with the following statement: I gained new factual information from participating in the experiment.
Gain knowledge B	By participating in the experiment, did you improve your personal knowledge of the natural system in question?
Restructure knowledge	To what extent have the outcomes of the experiment been a surprise and compelled you to amend your initial expectations about the outcome of the intervention?
Normative learning	
Change perspective	To what extent do you agree with the following statement: Participating in the experiment has changed the importance I attach to environmental issues.
Goal convergence	To what extent do you agree with the following statement: The experiment ensured that participants discovered a common goal.
Relational learning	
Understand mind-sets	To what extent do you agree with the following statement: By participating in the experiment, I have developed a stronger bond with those with which I usually disagree.
Build trust 2A	To what extent do you agree with the following statement: As a result of the experiment, a mutual trust has grown between participants.
Build trust 2B	To what extent do you agree with the following statement: I would participate again in an experiment with these participants.

Appendix C

Appendix C.1. Statistics for the Individual Learning Question Scores

Scores: high ≥ 1; medium = 0.5–1; low = 0–0.49; and none ≤ 0.

	N	Min	Max	Mean	S.D.
Cognitive learning—gain knowledge	18	0.5	1.9	1.1	0.39
Cognitive learning—restructure knowledge	18	−0.3	2	0.5	0.55
Normative learning—change perspective	18	−1	0.3	−0.38	0.33
Normative learning—goal convergence	18	−0.1	1.1	0.6	0.33
Relational learning—understand mind-sets	18	−1.0	0.5	−0.1	0.41
Relational learning—build trust	18	0.5	1.2	0.8	0.24

Appendix C.2. Reliability and Correlations

Learning Dimension Questions	**Reliability Score**
Gain knowledge 1 & 2	0.7
Building trust 1 & 2	0.7

Appendix C.3. Extent of Statistical Significance for Differences in Learning per Ideal Type, as Calculated Using Kruskal-Wallis Test (p < 0.05). Asterix (*) Confirms Statistical Significance

Learning Type	Learning Variable	Significance between Experiment Types
Cognitive learning	Gain knowledge	0.025 *
	Restructure knowledge	0.04 *
Normative learning	Change in perspective	0.637
	Goal convergence	0.053
Relational learning	Understand mind-sets	0.033 *
	Build trust	0.394

Appendix C.4. Statistics Showing Relationship between Learning and the Intervening Variables, Using Somers' D and Kruskal Wallis Tests (p < 0.05). Asterix (*) Confirms Statistical Significance

Intervening Variables	Gain Knowledge	Restructure Knowledge	Change in Perspective	Goal Convergence	Understand Mind-Sets	Build Trust
Addressing urgent issue	$p = 0.23$	$p = 0.78$	$p = 0.53$	$p = 0.87$	$p = 0.61$	$p = 0.75$
Organiser competence	$p = 0.69$	$p = 0.89$	$p = 0.28$	$p = 0.93$	$p = 0.74$	$p = 0.018$ *
Demographics—age	$p = 0.63$	$p = 0.91$	$p = 0.39$	$p = 0.67$	$p = 0.5$	$p = 0.29$
Extent participants already knew each other	$p = 0.62$	$p = 0.24$	$p = 0.54$	$p = 0.016$ *	$p = 0.97$	$p = 0.009$ *
Demographics—sex	$p = 0.87$	$p = 0.29$	$p = 0.44$	$p = 0.88$	$p = 0.35$	$p = 0.19$
Actor type	$p = 0.23$	$p = 0.6$	$p = 0.013$ *	$p = 0.01$ *	$p = 0.27$	$p = 0.69$

Appendix D

Experiment (Exp.)	Policy Issue/Type of Problem	(New) Policy Concept/How Tested	Type of Solution
Exp 1	Coastal management/sea level rise	Building with nature. Technical measures	Technocratic Ideal Type
Exp 2	Coastal management/sea level rise	Building with nature. Technical measures	Boundary Ideal Type
Exp 3	Dike management/river level rise	Optimal spatial planning. Technical measure	Advocacy Ideal Type
Exp 4	Freshwater availability/decline in freshwater availability	Shared responsibility. Technical measure (control site); governance innovation	Advocacy Ideal Type
Exp 5	Water variability/increase in flooding or drought risk	Multi-functional land use/shared responsibility. Technical measure; governance innovation	Advocacy Ideal Type
Exp 6	Freshwater availability/decline in freshwater availability	Water husbandry/shared responsibility. Technical measure	Technocratic Ideal Type
Exp 7	Water variability/increase in flooding or drought risk	Shared responsibility. Technical measure; governance innovation	Boundary Ideal Type
Exp 8	Water variability	Saltwater-freshwater transitions. Management measure	Boundary Ideal Type
Exp 9	Water variability/increase in flooding or drought risk	Multi-functional land use. Technical measure	Advocacy Ideal Type
Exp 10	Coastal management/sea level rise	Climate buffers. Technical measure	Boundary Ideal Type
Exp 11	Coastal management/sea level rise	Dynamic coastal management. Technical measure	Technocratic Ideal Type
Exp 12	Dike management/river level rise	Pest management. Management measure (control site)	Boundary Ideal Type
Exp 13	Water variability/increase in flooding or drought risk	Dynamic level management. Management measure	Technocratic Ideal Type
Exp 14	Water variability/increase in flooding or drought risk	Flexible groundwater irrigation. Management measure	Boundary Ideal Type
Exp 15	Water variability/increase in flooding or drought risk	Multi-functional land use/shared responsibility. Technical measure; governance innovation	Advocacy Ideal Type

Experiment (Exp.)	Policy Issue/Type of Problem	(New) Policy Concept/How Tested	Type of Solution
Exp 16	Water variability/increase in flooding or drought risk	Multi-functional land use. Technical measure.	Advocacy Ideal Type
Exp 17	Coastal management/sea level rise	Building with nature. Technical measure.	Advocacy Ideal Type
Exp 18	Water variability/increase in flooding or drought risk	Multi-functional land use. Technical measure	Technocratic Ideal Type

References

1. Lee, K.N. Conservation Ecology: Appraising Adaptive Management. *Ecol. Soc.* **1999**, *3*. Available online: http://www.ecologyandsociety.org/vol3/iss2/art3/ (accessed on 8 February 2017).
2. Huitema, D.; Adger, W.N.; Berkhout, F.; Massey, E.; Mazmanian, D.; Munaretto, S.; Plummer, R.; Termeer, C.C.J.A.M. The governance of adaptation: Choices, reasons, and effects. Introduction to the Special Feature. *Ecol. Soc.* **2016**, *21*, 37. [CrossRef]
3. Baird, J.; Plummer, R.; Haug, C.; Huitema, D. Learning effects of interactive decision-making processes for climate change adaptation. *Glob. Environ. Chang.* **2014**, *27*, 51–63. [CrossRef]
4. Bennett, C.J.; Howlett, M. The lessons of learning—Reconciling theories of policy learning and policy change. *Policy Sci.* **1992**, *25*, 275–294. [CrossRef]
5. Leach, W.D.; Weible, C.M.; Vince, S.R.; Siddiki, S.N.; Calanni, J.C. Fostering Learning through Collaboration: Knowledge Acquisition and Belief Change in Marine Aquaculture Partnerships. *J. Public Adm. Res. Theory* **2013**, *24*, 591–622. [CrossRef]
6. Huitema, D.; Mostert, E.; Egas, W.; Moellenkamp, S.; Pahl-Wostl, C.; Yalcin, R. Adaptive water governance: Assessing the institutional prescriptions of adaptive (co-) management from a governance perspective and defining a research agenda. *Ecol. Soc.* **2009**, *14*, 26. [CrossRef]
7. Ansell, C.K.; Bartenberger, M. Varieties of experimentalism. *Ecol. Econ.* **2016**, *130*, 64–73. [CrossRef]
8. Dunn, W. *The Experimenting Society: Essays in Honor of Donald T. Campbell (Policy Studies Review Annual)*; Transaction Publishers: New Brunswick, NJ, USA, 1998.
9. Walters, C.J.; Holling, C.S. Large-scale management experiments and learning by doing. *Ecology* **1990**, *71*, 2060–2068. [CrossRef]
10. Kemp, R.; Schot, J.; Hoogma, R. Regime shifts to sustainability through processes of niche formation: The approach of strategic niche management. *Technol. Anal. Strateg. Manag.* **1998**, *10*, 175–198. [CrossRef]
11. Sanderson, I. Evaluation, policy learning and evidence-based policy making. *Public Adm.* **2002**, *80*, 1–22. [CrossRef]
12. Millo, Y.; Lezaun, J. Regulatory experiments: Genetically modified crops and financial derivatives on trial. *Sci. Public Policy* **2006**, *33*, 179–190. [CrossRef]
13. Massey, E.; Huitema, D. The emergence of climate change adaptation as a policy field: The case of England. *Reg. Environ. Chang.* **2012**. [CrossRef]
14. Fischer, F. *Evaluating Public Policy*; Nelson Hall: Chicago, IL, USA, 1995.
15. Armitage, D.; Marschke, M.; Plummer, R. Adaptive co-management and the paradox of learning. *Glob. Environ. Chang.* **2008**, *18*, 86–98. [CrossRef]
16. Farrelly, M.; Brown, R. Re-thinking urban water management: Experimentation as a way forward? *Glob. Environ. Chang.* **2011**, *21*, 721–732. [CrossRef]
17. Bos, J.J.; Brown, R.R. Governance experimentation and factors of success in socio-technical transitions in the urban water sector. *Technol. Forecast. Soc. Chang.* **2012**, *79*, 1340–1353. [CrossRef]
18. Mostert, E.; Pahl-Wostl, C.; Rees, Y.; Searle, B.; Tàbara, D.; Tippett, J. Social learning in European river-basin management: Barriers and fostering mechanisms from 10 river basins. *Ecol. Soc.* **2007**, *12*, 19. [CrossRef]
19. Gerlak, A.; Heikkila, T. Building a theory of learning in collaboratives: Evidence from the Everglades restoration program. *J. Public Adm. Res. Theory* **2011**, *21*, 619–644. [CrossRef]
20. Newig, J.; Günther, D.; Pahl-Wostl, C. Synapses in the network: Learning in governance networks in the context of environmental management. *Ecol. Soc.* **2010**, *15*, 24. [CrossRef]
21. Muro, M.; Jeffrey, P. Time to talk? How the structure of dialog processes shapes stakeholder learning in participatory water resources management. *Ecol. Soc.* **2012**, *17*, 3. [CrossRef]

22. Rodela, R.; Cundhill, G.; Wals, A.E.J. An analysis of the methodological underpinnings of social learning research in natural resource management. *Ecol. Econ.* **2012**, *77*, 16–26. [CrossRef]

23. McFadgen, B.; Huitema, D. Are all experiments created equal? A framework for analysis of the learning potential of policy experiments in environmental governance. *J. Environ. Plan. Manag.* **2016**. [CrossRef]

24. Pielke, R.A., Jr. *The Honest Broker: Making Sense of Science in Policy and Politics*; Cambridge University Press: Cambridge, UK, 2007.

25. Sabatier, P. Knowledge, policy oriented learning and policy change: An advocacy coalition framework. *Sci. Commun.* **1987**, *8*, 649–692. [CrossRef]

26. Heikkila, T.; Gerlak, A.K. Building a conceptual approach to collective learning: Lessons for public policy scholars. *Policy Stud. J.* **2013**, *41*, 484–512. [CrossRef]

27. Haug, C.; Huitema, D.; Wenzler, I. Learning through games? Evaluating the learning effect of a policy exercise on European climate policy. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 968–981. [CrossRef]

28. Webler, T.; Kastenholz, H.; Renn, O. Public participation in impact assessment: A social learning perspective. *Environ. Impact Assess. Rev.* **1995**, *15*, 443–463. [CrossRef]

29. Reed, M.S. Stakeholder participation for environmental management: A literature review. *Biol. Conserv.* **2008**, *141*, 2417–2431. [CrossRef]

30. Huitema, D.; Cornelisse, C.; Ottow, B. Is the jury still out? Toward greater insight in policy learning in participatory decision processes—The case of Dutch citizens' juries on water management in the Rhine Basin. *Ecol. Soc.* **2010**, *15*, 16. [CrossRef]

31. Munaretto, S.; Huitema, D. Adaptive Comanagement in the Venice Lagoon? An Analysis of Current Water and Environmental Management Practices and Prospects for Change. *Ecol. Soc.* **2012**, *17*, 19. [CrossRef]

32. Pahl-Wostl, C.; Mostert, E.; Tàbara, D. The growing importance of social learning in water resources management and sustainability science. *Ecol. Soc.* **2008**, *13*, 24. [CrossRef]

33. Caspary, W.R. *Dewey on Democracy*; Cornell University Press: New York, NY, USA, 2000.

34. Campbell, D.T. Reforms as experiments. *Am. Psychol.* **1969**, *24*, 409–429. [CrossRef]

35. Greenberg, D.H.; Linksz, D.; Mandell, M. *Social Experimentation and Public Policymaking*; The Urban Institute: Washington, DC, USA, 2003.

36. Lindblom, C.E. The science of muddling through. *Public Admin. Rev.* **1959**, *19*, 79–88. [CrossRef]

37. Berkhout, F.; Verbong, G.; Wieczorek, A.J.; Raven, R.; Lebel, L.; Bai, X. Sustainability experiments in Asia: Innovations shaping alternative development pathways? *Environ. Sci. Policy* **2010**, *13*, 261–271. [CrossRef]

38. Hoffman, M.J. *Climate Governance at the Crossroads: Experimenting with a Global Response*; Oxford University Press: New York, NY, USA, 2011.

39. Castán Broto, V.; Bulkeley, H. A survey of urban climate change experiments in 100 cities. *Glob. Environ. Chang.* **2013**, *23*, 92–102. [CrossRef] [PubMed]

40. Dryzek, J. *Rational Ecology: Environment and Political Economy*; Basil Blackwell: New York, NY, USA, 1987.

41. Owens, S.; Rayner, T.; Bina, O. New agendas for appraisal: Reflections on theory, practice, and research. *Environ. Plan. A* **2004**, *36*, 1943–1959. [CrossRef]

42. Huitema, D.; Meijerink, S. The politics of river basin organisations: Institutional design choices, coalitions, and consequences. In *The Politics of River Basin Organisations*; Huitema, D., Meijerink, S., Eds.; Edward Elgar: Cheltenham, UK, 2014; pp. 1–37.

43. Weber, M. *The Methodology of the Social Sciences*; Shils, E.A., Finch, H.A., Eds.; Free Press: New York, NY, USA, 1949.

44. Ostrom, E. *Understanding Institutional Diversity*; Princeton University: New Haven, CT, USA, 2005.

45. Funtowicz, S.O.; Ravetz, J.R. *Uncertainty and Quality in Science for Policy*; Kluwer Academic: Dordrecht, The Netherlands, 1990.

46. Voß, J.P.; Simons, A. Instrument constituencies and the supply side of policy innovation. *Environ. Politics* **2014**, *23*, 735–754. [CrossRef]

47. Vedung, E. *Public Policy and Program Evaluation*; Transaction Publishers: New Brunswick, NJ, USA, 1997.

48. Checkel, J.T. Why comply? Social learning and European identity change. *Int. Organ.* **2001**, *55*, 553–588. [CrossRef]

49. McFadgen, B.; Huitema, D. Experimentation at the Interface of Science and Policy: A Multi-Case Analysis of How Policy Experiments Influence Political Decision-Makers. *Policy Sci.* **2017**, 1–27. [CrossRef]

50. Van der Heijden, J. What 'Works' in Environmental Policy-Design? Lessons from Experiments in the Australian and Dutch Building Sectors. *J. Environ. Policy Plan.* **2014**, *17*. [CrossRef]

51. Schusler, T.M.; Decker, D.J.; Pfeffer, M.J. Social learning for collaborative natural resource management. *Soc. Nat. Resour.* **2003**, *15*, 309–326. [CrossRef]

52. Mintrom, M.; Norman, P. Policy entrepreneurship and policy change. *Policy Stud. J.* **2009**, *37*, 649–667. [CrossRef]