Routledge
Taylor & Francis Group

Journal of Environmental Planning and Management

ISSN: 0964-0568 (Print) 1360-0559 (Online) Journal homepage: https://www.tandfonline.com/loi/cjep20

# Are all experiments created equal? A framework for analysis of the learning potential of policy experiments in environmental governance

Belinda McFadgen & Dave Huitema

**To cite this article:** Belinda McFadgen & Dave Huitema (2017) Are all experiments created equal? A framework for analysis of the learning potential of policy experiments in environmental governance, Journal of Environmental Planning and Management, 60:10, 1765-1784, DOI: 10.1080/09640568.2016.1256808

**To link to this article:** https://doi.org/10.1080/09640568.2016.1256808

*Journal of Environmental Planning and Management*, 2017
Vol. 60, No. 10, 1765–1784, http://dx.doi.org/10.1080/09640568.2016.1256808

Routledge
Taylor & Francis Group

# Are all experiments created equal? A framework for analysis of the learning potential of policy experiments in environmental governance

Belinda McFadgen* and Dave Huitema

*Institute for Environmental Studies, VU University, Amsterdam, The Netherlands*

(*Received 28 June 2016; final version received 26 October 2016*)

Environmental changes are increasing the need to understand complex cross-scale feedbacks in social–ecological systems. However, consistent conceptualisation of learning associated with environmental governance is lacking, and research mainly centres on individual variables. This paper identifies a typology of such learning, and theorises about configurations of variables. Focusing on experimentation as an intervention geared towards learning, it proposes a definition of policy experiment. A theoretical framework is presented, summarising a typology of experiments based on learning-related variables embedded in design choices, and reflected in institutional rule aggregations. The framework facilitates systematic analysis of real-world cases and testing of hypotheses on the effects of different types of experiment on learning. A case study demonstrates application of the framework. Results suggest future research paths that include attention to additional relevant variables. The findings have relevance for scholars interested in experimentation and learning, and environmental policy-makers considering experimentation to assess policy innovations.

**Keywords:** Experimentation; policy learning; science–policy interface; institutional design

## 1. Introduction

Environmental problems increasingly require policy solutions that recognise the complexity of social–ecological systems and the inherently uncertain cross-scale feedbacks acting within them (Folke *et al*. 2005). Governance choices can influence how much is learned about these systems and therefore the responsiveness of policy to them: studies show strong links between particular governance factors (as independent variables) and learning outcomes (e.g. Schusler, Decker, and Pfeffer 2003; Muro and Jeffrey 2012; Leach *et al*. 2013). These studies, however, centre on the learning effects of individual factors rather than sets of combined factors. In this regard, the suggestion by Muro and Jeffrey (2012) that some organisational arrangements (which are sets of multiple factors and concrete aspects of governance processes) may be more effective in fostering learning than others is novel, and needs further exploration.

The research setting for this paper is policy experimentation: a process that generates learning through an explicit intention to test new ideas (Sanderson 2009). Political agents can indicate a serious commitment to improving policies by enacting experiments, generating evidence of what works, and thus quickening the pace of learning (Garaway and Arthur 2004; Sanderson 2009). Several studies demonstrate links between

*Corresponding author. Email: Belinda.mcfadgen@vu.nl

experimentation and learning. Armitage, Marschke, and Plummer (2008) describe experiments as learning mechanisms, and examine a case of adaptive experimentation with noted first- and second-order learning effects. Farrelly and Brown (2011) link transitions theory to social learning in an examination of urban water experiments and their potential as producers of second-order 'conceptual' learning (see also Bos, Brown, and Farrelly 2013). Van der Heijden (2014 p. 18) analyses learning as one output of experiments in the Dutch and Australian building sectors, finding experiments "need to be developed around the aim of drawing lessons." Finally, Greenberg, Linksz, and Mandell (2003) assess five experiments for their substantive and conceptual effects on policy decision-making.

In terms of what is learned, experiments are generally expected to provide instrumental, reliable knowledge about the effects of an intervention. However, experiments may also lead to other forms of learning, including changes in the understanding of interests and perspectives, in trust and understanding of relationships among participants, and in levels of awareness and experience of decision processes (Munaretto and Huitema 2012). Some consider them capable of producing reflexive learning processes (Armitage, Marschke, and Plummer 2008) but, others, Fischer (1995), for example, determine that experiments are unsuitable to enable changes in worldview or beliefs relevant to policy (see also Farrelly and Brown [2011]). These conclusions are theoretically informed but lack an empirical evidence base, in line with the lack of hypothesis testing and direct measurement in learning scholarship in general (Rodela 2011). Thus, while experiments are generally understood as vehicles for learning, the studies encompass varied notions of learning and a wide range of empirical examples of experiments without classification. The studies also do not go beyond investigation of the influence on learning of separate process factors or question whether some experiments generate different learning effects than others and, if so, under what conditions. Our research is intended as a focused first step to fill these knowledge gaps and to help fill the need for more structured assessments of learning effects.

We propose that differences in the governance design of experiments affect the types of learning they produce. This would build on the research into influences on learning, for example, by Armitage, Marschke, and Plummer (2008), who suggest that in order to understand how learning comes about, it is important to examine who participates in an experiment, how they participate, and the extent of representativeness (see also Mostert *et al*. [2007]). To facilitate a systematic testing of hypotheses about the learning effects of different experiments, an exploratory framework is developed that uses three policy experiment types based on the policy appraisal, social learning, and science–policy interface (SPI) literatures. The framework is demonstrated using a case study of an environmental policy experiment in the Netherlands. The results, further development of the framework, and limitations of, and future paths for this research are discussed.

## 2. Defining learning and experimentation

### *2.1. Policy learning*

It was noted over 15 years ago that the learning literature includes at least 50 theoretical definitions (Social Learning Group 2001), a number that can only have increased since then.[1] A conceptualisation of learning is chosen for this research that has proven to be highly applicable to environmental governance. As we are mainly interested in learning in policy settings, we use a slightly revised version of Sabatier's oft-cited definition of learning: "the relatively enduring alterations of thought or behavioural intentions that

result from experience, [and] which are concerned with the attainment or revision of [public policy]" (Sabatier 1988 p. 131).[2] For learning among the participants in an experiment, this concerns the gaining of new knowledge and improved structuring of existing knowledge (*cognitive learning*), as well as acquiring a deeper understanding of the policy process that requires reflection on, and changes in, perspective, goals, or priorities (*normative learning*). Following Haug, Huitema, and Wenzler (2011), we supplement this definition with an additional dimension: *relational learning* – a change in trust, the ability to cooperate, and understanding of other parties (Webler, Kastenholz, and Renn 1995; Pahl-Wostl 2006).

There is a marked difference between learning amongst participants in the experiment and the impact on actors within the wider policy network. The learning effects for the participants in an experiment are expected to be relatively direct, some available for assessment immediately after an experiment has ended. However, the effect of experiment results on political decision-makers is more likely to be indirect and protracted (Weiss 1977; Greenberg, Linksz, and Mandell 2003). For this reason, we analyse 'enlightenment' as the type of learning experienced by political decision-makers as an impact of an experiment (Weiss 1977) in addition to the types of learning experienced by experiment participants, as described above. Decision-makers may decline to use research findings in their work, but this does not necessarily mean the results have not had an effect. According to Weiss, enlightenment is a subset of research utilisation that focuses on the 'gradual sedimentation' of the knowledge and understanding produced by research into policy-making.

Section 3 explains the application of these learning typologies in more detail, but first, we define policy experiments for the purposes of our analysis.

### *2.2. Policy experiments*

Experimentation is currently a popular area of exploration in several fields, including the policy sciences, adaptive management, transitions management, and climate governance. Understanding of the concept in environmental governance has diverged considerably: experiments are analysed as *ex ante* policy appraisals that assess alternative management interventions (Lee 1999; Campbell 1998; Sanderson 2002; Armitage, Marschke, and Plummer 2008; Huitema *et al.* 2009); as 'niches' to test specific mechanisms and increase the likelihood of adoption into the wider 'regime' (Hoogma *et al*. 2002; Berkhout *et al*. 2010, Farrelly and Brown 2011; Bos and Brown 2012); or as 'purposive' initiatives existing outside normal processes of policy implementation (Hoffman 2011; Castán Broto and Bulkeley 2013). The different understandings, nevertheless, all consider experiments as temporary and reversible interventions without permanent policy consequences (Tassey 2014), and as venues to promote learning. This paper provides a specific definition of policy experiment for environmental governance that encompasses these common characteristics.

Reviewing the literature, Huitema *et al.* (2009) conclude that experimentation is understood in environmental governance as either as a *research methodology* or as an *approach to management*. In the former, experiments test hypotheses on the ecological or social system response to different interventions in order to improve our scientific knowledge of the system and identify better solutions for particular problems (e.g. Richter *et al*. 2003; Cumming *et al.* 2013). This understanding arises particularly in political science, where analysing alternative social interventions using an experimental design (randomising groups, including a control setting) is considered the most rigorous method (Campbell 1998; Vedung 1997; Haynes, Goldacre, and Torgerson 2012;

Druckman and Lupia 2012). The second approach sees management as a form of experimentation, based always on incomplete, uncertain information, and consequently as a kind of hypothesis testing (e.g. Walters and Hollings 1990; Pahl-Wostl 2006). Here, the need for rigorous scientific design is relaxed; a quasi-experimental design can be implemented instead (Bennett 1996). Most of the environmental governance references to experimentation fit this approach. Castán Broto and Bulkeley (2013), for example, study climate experiments they define as novel, purposive initiatives emerging outside formal policy processes. Experiments can also be alternative policy processes that enact new ideas with the aim of identifying ways to upscale them (Farrelly and Brown 2011; Bos and Brown 2012). Climate experiments and experiments intended for upscale do not require testing and evaluation trials, and focus on the novelty and diffusion mechanisms of the intervention.

Here, we focus on experimentation as a management approach, and treat it as a temporary SPI – a social process "which encompass[es] relations between scientists and other actors in the policy process, and which allow[s] for exchanges, co-evolution, and joint construction of knowledge with the aim of enriching decision-making" (van den Hove 2007 (p. 815); see also Hoppe [2010]). To address the current imprecise use of the concept, we propose two further requirements: (1) a focus on novelty–risk-taking with actions outside established practices (cf. Hoffmann 2011), release from established laws, and the creation of free spaces (niches, shadow networks) where innovations can emerge (Olsson *et al*. 2006); (2) explicit expectations and assessment. In our understanding, an experiment should determine whether an innovation works, not how it can be made to work (Sanderson 2002). Thus, an experimental initiative must contain an explicit action theory that includes the intended effects; organisers must be open to testing the theory. Evaluation leads to reliable information derived from management interventions in practice, potentially providing new evidence for policy-making (see Rondinelli 1993, Campbell 1998, Lee 1999, Millo and Lezaun 2006, Armitage, Marschke, and Plummer 2008).

Since an experiment should both enable and test an innovative concept, we propose the following definition: 'a temporary, controlled field-trial of a policy-relevant innovation that produces evidence for subsequent policy decisions'. Policy relevance can be explicit (e.g. testing a new policy instrument) or implicit (e.g. testing new policy concepts or policy-related management interventions). The 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 to create some form of policy learning.

Variation in the characteristics of a policy experiment, as defined above, may influence the learning generated, particularly in comparison with non-experimental interventions. However, we focus instead on the governance design of experiments because the literature on learning (e.g. Mostert *et al*. 2007; Muro and Jeffrey 2012; Leach *et al*. 2013) emphasises that process factors heavily influence the levels and types of learning generated (see Section 3.3). The next section sets out the development of an analytical framework to delve deeper into the question of how aggregations of factors, as embedded in governance design choices, could influence learning.

## 3. Analytical framework

The analytical framework brings together typologies for learning and experimentation to facilitate the testing of hypotheses about the nature of their relationship. For participant

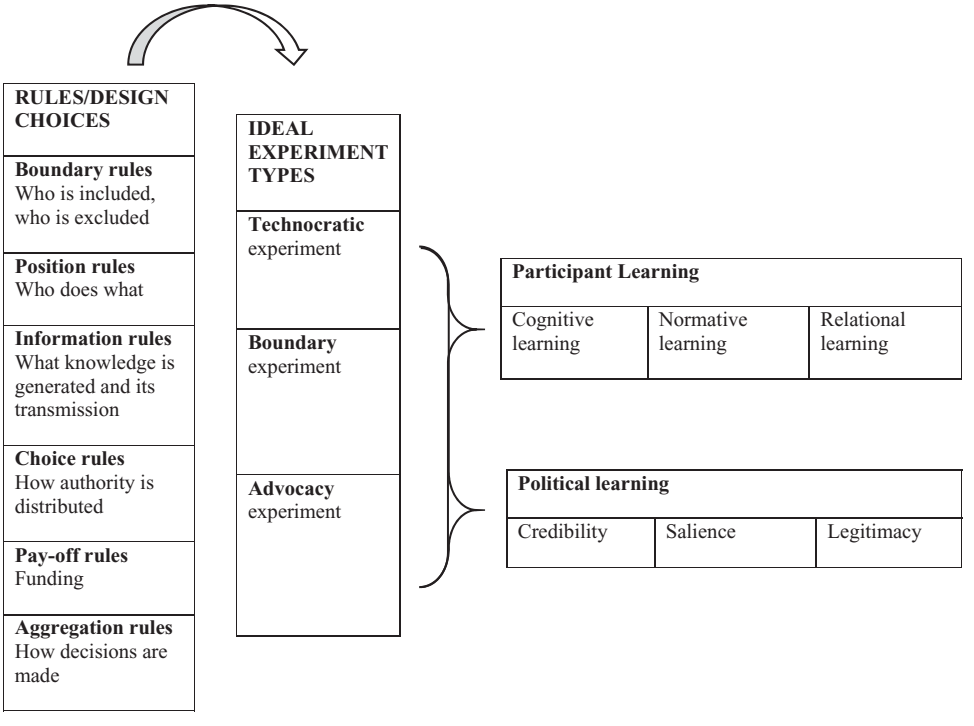


Figure 1. Main elements of a proposed analytical framework used to explore the relationship between different ideal types of policy experiment, as based on differences in governance design, and their learning effects on participants and policy decision-makers within the policy network.

learning, we use the typology developed by Haug, Huitema, and Wenzler (2011). The typology used to assess learning as an effect on the wider policy network (enlightenment) comes from science–policy evaluation (Cash *et al* 2003; Saarki *et al*. 2014). For experiments, the framework uses 'ideal types' (in the sense of Weber [1968]), to categorise variations in governance design choices. The analytical framework is set out in Figure 1.

### *3.1. Typologies for assessment of policy learning*

The selected unit of analysis is learning at the individual level, systematically measured via survey or interview, with results then aggregated to produce findings on group learning. The three learning types used for assessment, *cognitive learning*, *normative learning*, and *relational learning* (Haug, Huitema, and Wenzler (2011), contribute to environmental governance. For example, well-structured information as a form of cognitive learning brings advocacy and enlightenment functions (Sabatier 1978; Grin and Loeber 2007). Normative learning can enhance individual awareness; at the group level (social learning), it may be expressed as the development of common interests or goals, leading to political consensus and collective action (Leach *et al*. 2013). Relational learning mirrors the goal of moral development, enabling participants to consider alternative perspectives and improve cooperation (Webler, Kastenholz, and Renn 1995).

This typology has been applied several times in the study of learning from policy appraisal or collaboration in environmental governance (Huitema, Cornelisse, and Ottow 2010; Haug, Huitema, and Wenzler 2011; Baird *et al*. 2014). Although the forms of learning echo other definitions, certain characteristics set them apart. In particular, cognitive learning is similar to single-loop learning, where a "mismatch is detected and corrected without changing the underlying values and status quo that govern the behaviours" (Argyris 2003 p. 108). In contrast, normative learning compares to double-loop learning, where a "mismatch is detected and corrected by first changing the underlying values and other features of the status quo" (Argyris 2003, p. 108f). However, loop typology involves a hierarchy: double-loop learning is a supposed improvement on single-loop learning because the latter fails to change guiding assumptions. The aim only to implement projects that lead to a change in assumptions risks not getting "the job done and could result in endless cycles of reflection without implementation" (Fabricius and Cundill 2014 p. 4). Improving existing practices helps develop better policies, and is arguably as valuable as changing practices altogether (Owens, Rayner, and Bina 2004). We, therefore, delineate learning types without valuing one more than another. The typology is also innovative in that it distinguishes relational learning, which (if measured at all) tends to be subsumed under normative or 'higher' forms of learning (Pahl-Wostl 2006).

To assess the impact of an experiment on political decision-makers, we focus on enlightenment. Enlightenment, as the deepening understanding or illumination of research where actors form a positive view of the value of the new insights, is a valid measure of research usability (Weiss 1977). To gauge the extent of enlightenment, Weiss developed research characteristics for rating by decision-makers, including technical quality, political acceptability, and relevance to policy (Weiss 1977). Based on such characteristics, we operationalise dimensions regularly used to assess SPIs: credibility, salience, and legitimacy (Cash *et al*. 2003; Saarki *et al*. 2014). Policy decision-makers are expected to find the experiment *credible* if they consider it authoritative and of high quality, and if they trust the outcomes. They consider an experiment *salient* if they regard it as relevant to policy at a certain moment in time. Finally, decision-makers will think an experiment is *legitimate* if they view the experiment process as fair and fully incorporating the values, concerns, and perspectives of different actors (Cash *et al*. 2003). We select these dimensions because they are (1) well established in the literature; and (2) they capture key expectations of policy experiments, including the production of good-quality evidence, a meaningful influence on policy, and the ethical treatment of communities in which they are embedded (Greenberg, Linksz, and Mandell 2003).

### *3.2. A typology of experiments*

As producers of evidence for political decision-making, experiments connect the science and policy worlds, and can thus be understood as temporary SPIs. One way of categorising SPI arrangements is by differentiating the roles of science in policy-making (Pielke Jr. 2007): science used to arbitrate, advocate issues, or in honest brokering of policy options. These categories echo the technocratic and interpretive approaches to policy analysis (e.g. Owens, Rayner, and Bina 2004; Fischer 2007). Through these established concepts, three ideal types of experiments can be identified for empirical investigation: the *technocratic, boundary*, and *advocacy* experiments. Real-world examples can then be approximated against these theoretically informed types (Weber 1968; Dryzek 1987).

The three types are distinguishable through differences in their governance design; such as whether they are open, partially open, or closed to participants; whether they rely

on lay knowledge or just expert knowledge as evidence; or whether they tend towards more hierarchical or egalitarian authority distributions. In order to structure empirical analysis, as well to connect to process factors highlighted in the learning literature, the governance processes are translated into specific settings of institutional rules, as described in the Institutional Analysis and Development Framework (IADF) developed by Elinor Ostrom (2005). Using Ostrom's rules to structure an analysis provides robust conceptual clarity and analytical precision (Huitema and Meijerink 2014), and they are highly relevant: six of the seven sets of rules (governance process factors) in the IADF have been found independently to influence learning outcomes in mostly collaborative settings (Schusler, Decker, and Pfeffer 2003; Mostert *et al*. 2007; Muro and Jeffrey 2012; Leach *et al.* 2013) (see Section 3.3).[3]

The six rule sets cover: boundary (entry or exit) – who participates and how they are eligible; position – available roles and how they are assigned; information – information types, distribution, and generation; choice – assignment of action and prescription of power; aggregation – collective decision-making and decision weighting; payoff – management of costs and benefits, funding. The sections below describe the rule settings for each ideal type. (The relationships between governance design and rule setting are set out in Appendix 1 [online supplemental data])

### *3.2.1. Technocratic ideal type*

The technocratic experiment resembles the technical–rational model of policy decision-making, where an expert elite generates knowledge for policy decisions that is assumed to be universally applicable and thus independent of its context or subjects (Owens, Rayner, and Bina 2004; Fischer 2007). Typically, due to political disagreement over the effects of a policy proposal, policy actors commission such an experiment in order to obtain factual evidence. They fund the project, develop an action theory, and, following Churchill's assertion that 'scientists should be on tap, but not on top,[4]' set the policy goals for the experiment. The experiment produces scientific information regarding the effects of the new policy approach with connections to the policy process suspended until the end of the experiment, when the results are presented to decision-makers. Thus, the experiment is expected to play a neutral role in politics as a 'science arbiter' (Pielke Jr. 2007). Expert actors are the initiators of, and participants in, a technocratic experiment, and maintain control over its design, implementation, and evaluation. Scientific knowledge is the only type of information valued and generated by the experiment, and fact finding occurs within the parameters of the goals previously set. This arrangement helps reinforce the view that science is impartial to politics, which upholds the scientific integrity of the evidence but may limit its policy relevance.

### *3.2.2. Boundary ideal type*

A boundary experiment is one where the policy process is open to any actor, state, or non-state that has a desire to influence policy-making. The role of the experiment resembles that of 'honest broker of policy alternatives' (Pielke Jr. 2007) as it engages with the policy process and develops policy solutions in accordance with multiple value perspectives. A boundary experiment is initiated by a collaboration of actors, and generates diverse knowledge types, including ordinary and practical knowledge from non-experts, rendering it potentially responsive to policy needs. This policy-relevant knowledge is also subject to an extended societal peer review (Funtowicz and Ravetz 1990) since non-state actors have

influence and decision power over either some or all of the experiment's design, monitoring, and evaluation. Discussion over goal setting is strong in a boundary experiment, and reflection occurs on whether the experiment adheres to accepted societal aims. Deliberative practices are encouraged with transparent information transmission, open dialogue, and regular communication among participants. Ideally, this engagement allows different interpretations of the policy problem to emerge that build to a common consensus on the most appropriate course of action (Dryzek 1987).

#### *3.2.3. Advocacy ideal type*

An advocacy experiment is designed to produce evidence that steers policy towards a predefined position (Pielke Jr. 2007). Reflecting their own ideology, e.g. a commitment to market mechanisms and the framings of neoclassical economics, or to their beliefs about which interested parties take precedence, the organisers intend to push action in a particular policy direction. They use the experiment in the following ways: as a 'proof of principle' (Voß and Simons 2014); for softening objections to a predefined decision (Owens, Rayner, and Bina 2004); or as a tool to delay making final decisions (Greenberg, Linksz, and Mandell 2003). An advocacy experiment is organised by policy-makers, and populated by dominant, traditional actors (Hoogma *et al*. 2002). Although appearing neutral, the experiment is not open, as participants must be invited. Involvement is restricted by certain conditions, and excludes those with contrasting expectations. A steering group of dominant participants controls the design, monitoring, and evaluation procedures, reinforcing existing power structures. Closed to outsiders, the advocacy experiment is not transparent; within the experiment, information distribution channels are inconsistent as the organisers maintain control over what knowledge is available in order to suppress criticism. Within the group, only the dominant participants discuss and shape goals through the use of a facilitator; so prevailing norms are protected, which also limits the generation of new ideas.

To summarise, the three experiment types each represent an aggregate of different rule settings that create divergent configurations of participants, information, and power distribution (see Appendix 2 [online supplemental data] for more details on how the rules are set for each ideal type). Experiments could be examined in other ways; for instance, by looking at their purpose (Ettelt, Mays, and Allen 2015), their non-governance-related characteristics (learning mechanisms, biases, and direct outcomes) (van der Heijden 2014), or their different implications for the policy process (e.g. Greenberg, Linksz, and Mandell 2003). However, our research investigates instead the potential significance of governance design as captured in different aggregations of institutional rules; the aim is to build on work showing an influence on learning of individual process factors/ institutional rules.

### *3.3. Experiment design and learning*

The framework is based on the proposal that aggregates of factors in the governance of experiments influence cognitive, normative, and relational learning effects amongst the participants, and on the perceived credibility, salience, and legitimacy (elements of enlightenment) of the experiment findings as an impact on political decision-makers. To build hypotheses on how experiment governance factors are related to learning, use is made of the significant theoretical and empirical advances in research that explain both participant learning (e.g. Pahl-Wostl 2006; Mostert *et al*. 2007; Newig, Günther, and

Pahl-Wostl 2010; Rodela 2011; Muro and Jeffrey 2012; Leach *et al*. 2013) and political impact (e.g. Cash *et al*. 2003; Koetz, Farrell, and Bridgewater 2012; Sarkki *et al*. 2014). For example, Mostert *et al*. (2007) outline a list of 71 governance factors in eight themes that they found hinder or enable learning. The factors include independent facilitation, dissemination of information, joint planning and influence over the process, diverse but limited numbers of participants, common understanding, and frequent discussions. These factors have since been utilised as independent variables in other learning studies (e.g. Muro and Jeffrey 2012; Leach *et al*. 2013). In order to facilitate systematic comparison, they are considered to be embodied within rule settings in the governance design of experiments. Similarly, variables have been identified from the above texts as those that influence enlightenment as the impact on political decision-makers. Based on the experiment typology and these variables, we propose the following hypotheses and supporting explanations:

**Hypothesis 1:** *A technocratic experiment produces high levels of cognitive learning, no normative learning, and limited relational learning within the circle of participants. Political decision-makers will consider the experiment's findings very credible, but not salient or very legitimate*.

We argue that the emphasis of the technocratic experiment is on building scientific expertise. Given preset objectives, an experiment will be based on chosen points of debate, and results of theory testing in the experiment will conform to the paradigm of those points. As the boundary rules preclude the entry of actors with different ideas, the information rules emphasise expertise that fits the paradigm. As a result, governance of the experiment precludes normative learning, and produces mostly cognitive learning. Because the information rules allow open communication without suppression, some relational learning may occur in the process, but most will be internal scientific (informing those best at solving the puzzle at hand and at understanding the social–ecological system), and unlikely to create higher levels of trust within a policy network. With regard to pay-off rules, funding for technocratic experiments is likely to be from organisations concerned more with publishing scientific results than with policy relevance. Due to the length of time that passes between commissioning the experiment and the generation of results, policy relevance and salience of the outcomes will probably be low. The closed character of the experiment is expected to reduce its legitimacy as the formulation of the research question, the data gathering process and the report writing will not involve stakeholder groups or ordinary citizens, and may not address arguments they consider important (Milo and Lezaun 2006). A high level of credibility is expected; however, reflecting the likelihood that (1) scientific information will be developed according to the highest standards of reliability; and (2) scientific output is, in general, perceived to be credible.

**Hypothesis 2:** *A boundary experiment produces medium levels of cognitive learning and high levels of normative and relational learning within the circle of participants. Political decision-makers will consider the experiment's findings moderately or very salient and highly legitimate, but only somewhat credible.*

This hypothesis stems from the open design of a boundary experiment, where the boundary rules are set so broadly that they include all stakeholders who want to participate. This is expected to lead to participant diversity, which is considered a trigger of learning (Schusler, Decker, and Pfeffer 2003, Mostert *et al*. 2007, Gerlak and Heikkila 2011, Muro and Jeffrey 2012, Leach *et al*. 2013). Participants are exposed to a variety of ideas and understanding of the policy problem, which are shared amongst the group

through open and transparent information rules. Through the choice rules, non-state actors influence the setting of the problem definition and experiment goals, which increases the probability of public support for the intervention and the generation of shared norms. Since there is no dominant paradigm, a kaleidoscope of perspectives shapes the experiment, creating possibilities for normative learning. The focus on capturing different knowledge types (e.g. non-expert/lay knowledge about the system within which the experiment is embedded) can be expected to enhance the breadth of understanding about the experiment's effects. However, the focus on including the actors' various perspectives and values may be expensive and time consuming (Owens, Rayner, and Bina 2004), distracting from the singular focus of developing objective evidence, and reducing the amount of cognitive learning (further, Bos, Brown, and Farrelly [2013] demonstrate that the extent of learning can also be affected by what role the learner plays in the experiment). Participants have decision power through choice rules to influence the evaluation process, thereby capturing a wide variety of concerns, and generating trust in the political process – relational learning – as participants feel their needs are being met (Dryzek 1987; Webler, Kastenholz, and Renn 1995; Mostert *et al*. 2007; Muro and Jeffrey 2012). Within the policy network, boundary experiments are perceived as highly legitimate, because the focus on stakeholder interests and establishing ways to meet those interests ensures the findings meet societal expectations; however, non-expert knowledge may undermine the technical quality of the experiment findings, so reducing credibility. If the political decision-makers are 'in tune' with the current societal norms, then the findings will be considered very salient, but if the policy network is responding to cues other than societal norms (e.g. international, economic, or political influences), the findings will be moderately salient.

**Hypothesis 3:** *An advocacy experiment produces medium levels of cognitive learning, low normative learning, and low to medium levels of relational learning. Political decision-makers will consider the experiment's findings salient (under some circumstances), but not very credible or legitimate.*

This hypothesis reflects the intention of initiators to present a predefined policy solution as the most suitable course of action. Boundary rules allow entry to a potentially diverse set of participants, but access is limited to those chosen by the initiator with eligibility restricted to those who support the proposal. Participants contribute knowledge (Hegger *et al*. 2012; Muro and Jeffrey 2012), leading to the expectation of some cognitive learning, and slight normative learning triggered by persuasion tactics (Haug, Huitema, and Wenzler 2011), but less than could result from a breadth of viewpoints (Schusler, Decker, and Pfeffer 2003). Both types of learning are inhibited by the lack of open and regular lines of communication (Muro and Jeffrey 2012). The familiarity of participants with each other and their aligned views holds the potential for some relational learning, but no new actor networks emerge, and the suppression of certain information and lack of authority for most participants inhibit trust building. Credibility and legitimacy are questionable for an advocacy experiment: credibility is undermined by the inclusion of policy and non-state actors in the experiment along with expert actors, and by the production of practical knowledge alongside scientific knowledge, where a focus on the former can distract from the latter. The reliability of knowledge attributed to the experiment is diminished when favourable information is promoted and contrary results are suppressed. The initiator's aim to advocate a particular proposal blocks participation by actors critical of the proposal, and undermines their concerns, limiting fairness, and the perceived legitimacy of the project. However, the salience of the

Table 1. Expected learning outcome levels for ideal-type experiments.

| Learning effect | Technocratic experiment type | Boundary experiment type | Advocacy experiment type |
|---|---|---|---|
| Cognitive | High | Medium | Medium |
| Normative | Low/none | High | Medium |
| Relational | Medium | High | Low |
| Impact | | | |
| Credibility | High | Medium | Medium |
| Salience | Low/none | High | High |
| Legitimacy | Medium | High | Low |

findings may be perceived as high when the experiment serves to keep an idea alive (Greenberg, Linksz, and Mandell 2003), and the presentation of outcomes is timely.

Table 1 summarises the hypothetical expectations.

It is important to recognise that governance design factors are not the only factors related to learning; per the non-governance characteristics described above, and the discussion section below, our application of the framework has led us to include a set of control variables for assessment in parallel with the design choice/learning rules framework. The following section demonstrates basic use of the framework as applied to an experiment set in the Netherlands.

## 4. Methods

In 2013–2014, we conducted a case study to empirically test the framework using an exploratory research approach, with mixed methods so as to triangulate the data and improve the validity of our findings (Creswell 2013). The study assesses the institutional rule settings for a water management experiment conducted from 2008 to 2012 in the Netherlands, which involved sand nourishment and associated oyster reef structures. The case is relevant to climate adaptation, and has potentially large social and ecological effects. It was chosen as a good example of the Dutch government testing a policy innovation – the coupling together of water management issues within one solution. It puts into practice a change in approach to water management – combining coastal flood defence and ecological restoration – and was conducted in the Dutch Delta's Oosterschelde, a former estuary in South-West Netherlands Figure 3.1 in (a map of the case study area is provided as figure 3.1 in Appendix 3 [online supplemental data]).

Data were collected via interviews, surveys, and document analysis. We conducted seven semi-structured interviews with policy advisors, experts, and stakeholders, on the phone or in person, to ask about the reasons for experimenting and the participant's role. These respondents and other participants also completed an online survey. Survey links were sent to all 25 project participants, with 20 responses collected, giving an 80% response rate (and 2 open question responses). The survey questions asked participants about their position, the extent of their authority, what information they contributed, etc., so we could map the experiment's governance design.[5] The survey also asked questions to gauge participants' learning effects. Here, we measure the learning process, where new information or knowledge is acquired, processed, and transferred across individuals within a group (Gerlak and Heikkila 2011), and measured via each individual 'reporting' their experiences[6] (Appendix 4 [online supplemental data] lists the learning questions).

Secondary data collection was elicited from a range of documentary sources including project reports, media articles, scientific reports, and email correspondence (among others). This data improved our understanding of each experiment's context and policy relevance, and verified the interview data.

Questions used in the survey to measure the participants' learning effects mirrored (as much as possible) questions that have been published in the existing literature (e.g. Schusler, Decker, and Pfeffer 2003; Muro and Jeffrey 2012; Leach *et al*. 2013). These Likert-scale survey questions asked respondents to rate their experience on a four-point scale: 1 = 'not at all', 2 = 'slightly', 3 = 'a moderate change', 4 = 'a considerable change' (see Appendix 4 [online supplemental data]) and inform the scale used in Table 3.

For assessment of the three learning dimensions of impact on political decision-makers, we were unfortunately unable to gain access to those involved, and so were unable to fully assess enlightenment. We instead assessed by proxy, via interviews with the organisers, a document analysis of the perceived quality of the results, and a determination of whether action proceeded from them, and whether critical questions were raised on the policy side. The lack of survey data from decision-makers in the policy network represents a clear limitation to our research design.

The next section first reflects on the use and applicability of the framework, and then presents empirical evidence of the rule settings and learning and enlightenment effects.

## 5. Results

In order to determine the experiment's ideal type, we assessed its governance design against possible institutional rule settings, as set out in Appendix 1 (online supplemental data). The analysis reveals that the experiment most resembles the technocratic ideal type (see Table 2). The boundary rules, which determine who is eligible for entry and how

Table 2. Results of analysis of design choice rule settings in Netherlands Oosterschelde Delta water management experiment.

| Institutional rules | Case study findings |
|---|---|
| Boundary rules | Set broadly to include scientific experts, state actors, and industry actors. State actors involved from both national and regional government. Access open to joining. |
| Position rules | State took roles of initiator and primary funder. No facilitator role. |
| Information rules | Information produced was predominantly scientific. Some lay information used in design phase of experiment B. Emphasis on instrumental information and non-instrumental information extended to awareness of policy goals, but facilitation not used to communicate differing views. Communication channels reported as open and information regularly received. Knowledge transmitted rarely to general public through press releases. Participants had face-to-face contact through workshops. |
| Aggregation rules | State ultimately made decisions on project goals and design. State took decisions on execution of sand nourishment; expert collaborative on the oyster beds. Evaluation and conclusions decided upon consensually by state and scientific experts. |
| Choice rules | Majority of participants reported holding advisory or decision-making power roles at the monitoring and evaluation nodes. Least equitable power balance found at design node. |
| Payoff rules | State paid most costs with contribution from research collaboration. |

they access the experiment, are set heavily in favour of scientific experts, with minor representation of business and policy interests. Civil society actors were actually invited into the process, but decided to take an observational stance (although an NGO is involved in the steering committee that received the results, and advised decision-makers). One group raised concerns over impacts on recreational uses of the Oosterschelde if the management approach were formally adopted within policy; however, they did not feel the need to participate for the experimental phase (Telephone interview, April 2013). The information rules determine what information is considered valid, how it is distributed among the group, and who receives it. The exchange of information was the highest among scientific experts, and most open during the design of the experiment. This skews the categorisation towards the technocratic type because information was not shared evenly across the group of participants. The experiment investigated scientific hypotheses, so scientific information about the natural world dominated; however, non-scientific lay information from the fishermen involved was used to design and implement the oyster beds because the scientists had limited knowledge of oysters and useable substrate materials (Interview, December 2012). The experiment design did not encourage the development of reflexive knowledge, although some participants recalled discussion of policy objectives. Finally, the choice rules determine the distribution of authority between participants. Authority was measured at three decision nodes: the design, monitoring, and evaluation nodes, and participants were asked about the extent of their decision-making authority at each node. Results show that policy-makers were mainly responsible for the design phase, with few experts having any say at all (bringing some similarity to the advocacy type), whereas scientific experts held decision power over the monitoring and evaluation stages. Since the rules were set to include mostly scientific experts who were assigned authority at nodes of their expertise and who regularly and openly shared scientific and non-reflexive information, this experiment most closely matches the technocratic ideal type.

We expected that a technocratic ideal experiment type would produce high cognitive learning, low normative learning, and some relational learning with results that are considered credible and partially legitimate by the policy network. The survey results for the case study (Table 3) mostly meet expectations, showing that experiment participants recorded high cognitive learning and medium relational learning; however, there was also some normative learning.

In terms of political decision-maker learning, the experiment produced generalisable scientific results, with a focus on monitoring and the use of state-of-the-art equipment. An evaluation of the experiment had taken place after four years, with results presented to a workshop attended by policy-makers from the national government and the province, and scientists from the participating knowledge institutes. The initiator reported no criticism or questioning of results, and it was suggested that policy-makers recognised the calibre of the knowledge institution contracted to perform the monitoring work

Table 3. Learning measurements from the participant survey and desktop analysis for the case study of Netherlands Oosterschelde Delta water management experiment.

| | | | |
|---|---|---|---|
| Cognitive learning | High (3.3) | Credibility | High |
| Normative learning | Medium/low (2.5) | Salience | Medium |
| Relational learning | Medium (2.75) | Legitimacy | Medium |

Note: Out of 4: where >3.0 is high; 2.5–3.0 is medium, and <2.5 is low.

(Interview, March 2014). As a result of the experiment, larger sand supplementations are now planned for the Oosterschelde, showing confidence of decision-makers in the results of this experiment. To gauge saliency, we looked at the science–policy relationship and the relevance of the solutions to policy after five years since the beginning of the experiment. Two interviewees mentioned tensions between scientists and policy-makers where scientists expected results based on certainty gained over time. Policy-makers felt they needed to make decisions based on other considerations (Interviews, December 2012; January 2013). This suggests that the experiment produced results that had more relevance to science than policy, and that results could not be directly converted into policy. Despite this, Dutch water governance remains intent on pursuing solutions that solve nature and water safety policy problems concurrently, and the experiment remains relevant, e.g. through a 'climate proof' initiative, and integration into the south-western Delta programme. Finally, to assess the legitimacy of the results – the degree to which the process was seen to be equitable – we looked at the way stakeholder concerns were met and the extent of their involvement. The organisers developed a reporting system with industry stakeholders that would alert experiment participants to adverse effects caused by the sand supplementation, with a ceiling effect arranged so operations could be halted if the effects crossed a threshold. However, the extent of industry involvement beyond this contingency arrangement is almost negligible; industry actors were not involved in the final workshop, evaluation, or presentation of results.

Reflecting on usability, the use of Ostrom's rules provides a scope sufficiently broad to capture a range of structural and process variables relevant to learning. The neutrality of the rules means they can be 'set' in various ways (e.g. boundary rules can be very open, very closed, or somewhat open), allowing differentiation of the three ideal types. The simplicity of the rules means they can easily be made the subject of survey questions to determine experiment type based on respondent data. This research approach ensures a consistent and systematic analysis, which is novel for a learning study (Rodela 2011). The learning questions were designed to capture data in line with the definitions of each learning dimension (Appendix 4 [online supplemental data]). It is a weakness, however, that they do not assess the position of the respondents before the experiment. In terms of relational learning, for example, we cannot judge whether the participants already had a relationship that was strengthened, or whether new bonds grew as a result of the experiment. A question asking how many participants were known to a respondent prior to the experiment should clarify their responses.

## 6. Discussion and conclusion

This paper initiates inquiry beyond the effects of separate governance design factors on policy learning. It sets out principal learning types, and constructs a framework that identifies the relationship between policy learning and ideal types of policy experiment based on governance design. Here, we reflect on the novelty of the framework, findings from its application, and possible future paths for the research.

The framework draws on relatively established learning typologies and a new typology of policy experiments derived from the science–policy literature. Basing the experiment typology on differences in governance design is a new theoretical approach that allows unprecedented testing of hypotheses on the effects of governance factors on policy learning. The use of a single case study illustrates the details of the framework, and how a type is determined and analysis conducted.

As expected, given the nature of ideal types, a degree of non-conformity of the experiment to type is seen: experts were not the only actor types involved; some lay knowledge was incorporated into the implementation of the experiment; and access was open to all parties rather than by invitation only. Collaboration of government with different 'target groups' as seen in the case study, for example, is common in Dutch environmental policy-making (Pettenger 2007). Further, the legal requirements in the case demanded open access due to the siting of the experiment in a Natura 2000 area,[7] which obliged the initiators to invite all stakeholders into the process. The type, nevertheless, provides an accurate summary of key choices made for the governance of an experiment. Analysis of a sample of multiple experiments is needed to show the degree to which the factors, as described for each ideal type, prove sufficiently realistic.

A key finding is that structural and process variables are insufficient to explain the resultant learning. In our case study example, control variables may explain the increased normative learning and salience. The extent of conflict in the surrounding community as context, for example, could cause participants to reconsider their perspectives despite the lack of deliberative governance factors; or, in terms of agency variables, an information source outside the experiment in the wider policy world might influence the reputation of the experiment, and increase its saliency. This could be particularly so given that the policy issue is topical (climate change adaptation and water management). In support of these observations, the literature suggests that agency and context factors should be accounted for as controls in subsequent hypothesis testing. For example, agency factors include the quality of leadership of the experiment initiator, the demographics of participants, and the extent of their motivation (Gerlak and Heikkila 2011; Leach *et al*. 2013). In the wider policy network, relations may vary considerably; the parties involved (e.g. regulators and interest groups) may be either antagonistic or collaborative. This might be important in terms of governance design and therefore the type of experiment, but could also affect the perception of experiment in the network and therefore enlightenment.

In terms of the empirical approach, an additional limitation arises, at least partially, in the basis of analysis on self-reported insights into learning. Self-reported learning scores are recognised as likely to be biased to some extent (Haug, Huitema, and Wenzler 2011). We also note again that the conclusions about enlightenment are largely drawn from document and interview data; data recording the perceptions of political decision-makers about whether or not they gained enlightenment would have been preferred as a direct source for assessment.

Limitations and the need for parallel assessment of additional factors understood, the framework has strong implications for both academic scholarship and policy practice. The inherent political nature of experiments is an under-represented angle in experiment studies (one noted by scholars in adaptive management in particular; i.e. Voß and Bornemann 2011); the lens of an SPI helps make it clearer. Theoretically, the governance design aspects of experiments, assessed as institutional rules *per* Ostrom, provide a thorough and relevant set of variables for learning analysis; the framework includes three perspectives typically studied independently: participation, information transmission, and power sharing. Use of institutional rule settings thus helps in understanding the politics of experimentation through the examination of exclusion and inclusion of participants, and of controls over decision-making and information. A lot is expected of experiments, but claims have rarely been tested in practice (van der Heijden 2014). The framework enables more empirical analysis through the typologies, since they are easily operationalised, and can be used in large-*N* studies if sufficient data are available. For practitioners, the framework provides options by clearly and systematically setting out

the learning implications of the design choices they make (or allow others to make). With experimentation on the rise, and so much variation in quality (van der Heijden 2014), these insights may be welcome.

The analysis of a single case study serves as a demonstration. Further research could develop a methodology for determining the degree of closeness of an experiment to an ideal type, and a method to show such variations graphically to facilitate easy comparison within a field of cases; these additions would be supported by research into what underpins variation from type and breadth of difference in a set of cases. Future work could focus also on connections between learning within the experiment, and enlightenment effects in the wider policy environment, e.g. do strong learning effects for participants contribute to impact on political decision-makers? If not, why not? The relative weight of individual rules within the aggregations offers another potentially rich area for exploration; do some rules affect learning more than others (as explored by Leach *et al*. [2013] in aquaculture partnerships)? However, the most important next steps for research are to use the framework to test the above hypotheses on multiple experiments to see whether experiments vary in learning outcomes based on governance design, (and are indeed not 'equal'), and to include in the analysis the presence of predetermined intervening variables (context and agency) that are not incorporated in governance processes.

In conclusion, this explorative research clarifies the concepts of learning and experimental design, and provides a mechanism to determine relationships between them as they arise in experiments as a kind of management intervention widely considered an 'antidote' to policy-making by 'spurious certitude' (Gunderson 1999). The framework may contribute to greater understanding of experimentation and learning as two key coupled prescriptions in adaptive management and increasingly in environmental governance in general, helping to cope with social–ecological crises arising in modern times.

## Acknowledgements

Thank you to the Dutch Knowledge for Climate programme, which funded this research.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Supplemental data

Supplemental data for this article can be accessed here.

## Notes

1. Relevant studies on learning include: Lindblom's (1968) study on policy learning, where he delineates between instrumental and political learning; Argyris and Schon (1980) studies on single- and double-loop learning; Hall's policy learning (1993) with first-, second-, and third-order learning.
2. This definition is considered applicable also because Sabatier refers explicitly to experimentation as a tool that enables policy learning.
3. The 'scope rules' were omitted since there is little relevance of scope rules to factors found to generate learning. They are also difficult to operationalise.
4. Cited by Randolph Churchill (1965, 127).
5. The survey consisted of closed questions, but allowed respondents the opportunity to comment in an open section at the end. It was piloted on professionals known by the researchers.
6. A common method of learning assessment among many is to assess observed products of learning; for instance, policy changes, new projects, or new strategies (Bennett and Howlett

1992; Armitage, Marschke, and Plummer 2008; Gerlak and Heikkila 2011). An alternative method, and the one utilised in this research, is to measure the process of learning.

7. 92/43/EEC
8. Climate Proof Areas brochure, 2010. http://www.climateproofareas.com/project/sand-nourishment
9. RWS 2009. Harde werken met zachte trekken: voorbeelden van levende waterbouw. http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2009/10/01/harde-werken-met-zachte-trekken-voorbeelden-van-levende-waterbouw.html

## References


Argyris, C. 2003. "A Life Full of Learning." *Organisation Studies* 24 (7): 1178–1192.

Argyris, C., and D. A. Schon. 1980. *Organisational Learning*. Reading, MA: Addison-Wesley.

Armitage, D., M. Marschke, and R. Plummer. 2008. "Adaptive Co-Management and the Paradox of Learning." *Global Environmental Change* 18: 86–98.

Baird, Julia, Ryan Plummer, Constanze Haug, and Dave Huitema. 2014. "Learning Effects of Interactive Decision-Making Processes for Climate Change Adaptation." *Global Environmental Change* 27 (2014): 51–63.

Bennett, C. J., and M. Howlett. 1992. "The Lessons of Learning: Reconciling Theories of Policy Learning and Policy Change," *Policy Sciences* 25 (3): 275–294.

Bennett, Trevor. 1996. "What's New in Evaluation Research? A Note on the Pawson and Tilley Article." *British Journal of Criminology* 36 (4): 567–573.

Berkhout, F., G. Verbong, A.J. Wieczorek, R. Raven, L. Lebel, and X. Bai. 2010. "Sustainability Experiments in Asia: Innovations Shaping Alternative Development Pathways?" *Environmental Science and Policy* 13 (4): 261–271.

Bos, J. J., and R. R. Brown. 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.

Bos, J. J., R. R. Brown, and M. A. Farrelly. 2013. "A Design Framework for Creating Social Learning Situations." *Global Environmental Change* 23 (2): 398–412.

Campbell, D. T. 1998. "The Experimenting Society," In *The Experimenting Society: Essays in Honor of Donald T. Campbell. Policy Studies Review*, Annual Vol. 11, edited by W. Dunn, 35–69. New Brunswick, NJ: Transaction Publishers.

Cash, D. W., W. C. Clark, F. Alcock, N. M. Dickson, N. Eckley, D. H. Guston, J. Jäger, and R. B. Mitchell. 2003. "Science and Technology for Sustainable Development Special Feature: Knowledge Systems for Sustainable Development." *Proceedings of the National Academy of Sciences of the United States of America* 100: 8086–8091.

Castán Broto, V., and H. Bulkeley. 2013. "A Survey of Urban Climate Change Experiments in 100 Cities." *Global Environmental Change* 23 (1): 92–102.

Churchill, R. S. 1965. *Twenty-One Years*. London: Weidenfeld and Nicolson.

Creswell, John W. 2007. Research design: Qualitative, quantitative, and mixed methods approaches. London: Sage.

Cumming, Graeme S., Per Olsson, F. S. Chapin III, and C. S. Holling. 2013. "Resilience, Experimentation and Scale Mismatches in Social-Ecological Landscapes." *Landscape Ecology* 28 (6): 1139–1150.

Druckman, J., and A. Lupia. 2012. "Experimenting with Politics." *Science* 335: 1177–1179.

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

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

Fabricius, C., and G. Cundill. 2014. "Learning in Adaptive Management: Insights from Published Practice." *Ecology and Society* 19 (1): 29. http://dx.doi.org/10.5751/ES-06263-190129

Farrelly, M., and R. Brown. 2011. "Rethinking Urban Water Management: Experimentation as a Way Forward?" *Global Environmental Change* 21 (2): 721–732.

Fischer, F. 1995. *Evaluating Public Policy*. Chicago, IL: Nelson Hall.

Fischer, F. 2007. "Deliberative Policy Analysis as Practical Reason: Integrating Empirical and Normative Arguments." In *Handbook of Public Policy Analysis: Theory, Politics, and Methods*, edited by F. Fischer, G.J. Miller, and M.S. Sidney, 223–237. Boca Raton, FL: Taylor & Francis.

Folke, C., T. Hahn, P. Olsson, and J. Norberg. 2005. "Adaptive Governance of Social–Ecological Systems." *Annual Review of Environment and Resources* 30 (1): 441–473.

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

Garaway, C. J., and R. I. Arthur 2004. *Adaptive Learning: A Practical Framework for the Implementation of Adaptive Co-Management: Lessons from Selected Experiences in South and Southeast Asia*. London: MRAG.

Gerlak, A., and T. Heikkila. 2011. "Building a Theory of Learning in Collaboratives: Evidence from the Everglades Restoration Program." *Journal of Public Administration Research and Theory* 21: 619–644.

Greenberg, D., D. Linksz, and M. Mandell. 2003. *Social Experimentation and Public Policy Making*. Washington, DC: The Urban Institute Press.

Grin, J., and A. Loeber. 2007. "Theories of Policy Learning: Agency, Structure, and Change." In *Handbook of Public Policy Analysis: Theory, Politics, and Methods*, edited by F. Fischer, G. J. Miller, and M. S. Sidney, 201–219. Boca Raton, FL: Taylor & Francis.

Gunderson, L. 1999. "Resilience, Flexibility and Adaptive Management: Antidotes for Spurious Certitude?" *Conservation Ecology* 3 (1): 7. http://www.consecol.org/vol3/iss1/art7/

Hall, P. A. 1993. "Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain." *Comparative Politics* 25 (3): 275–296.

Haug, C., D. Huitema, and I. Wenzler. 2011. "Learning Through Games? Evaluating the Learning Effect of a Policy Exercise on European Climate Policy." *Technological Forecasting and Social Change* 78 (6): 968–981.

Haynes, L., B. Goldacre, and D. Torgerson. 2012. *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. London: Cabinet Office.

Hegger, D., M. Lamers, A. Van Zeijl-Rozema, and C. Dieperink. 2012. "Conceptualising Joint Knowledge Production in Regional Climate Change Adaptation Projects: Success Conditions and Levers for Action." *Environmental Science and Policy* 18: 52–65.

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

Hoogma, R., R. Kemp, J. Schot, and B. Truffer. 2002. *Experimenting for Sustainable Transport: The Approach of Strategic Niche Management*. London: Spon Press.

Hoppe, Rob. 2010. "From 'Knowledge Use' Towards 'Boundary Work'. Sketch of an Emerging New Agenda for Inquiry into Science–Policy Interaction." In *Knowledge Democracy: Consequences for Science, Politics and Media*, edited by Roeland Jaap in 't Veld, 169–186. Berlin: Springer.

Huitema, D., C. Cornelisse, and B. Ottow. 2010. "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." *Ecology and Society* 15 (1): 16. http://www.ecologyandsociety.org/vol15/iss1/art16/

Huitema, D., and S. Meijerink, eds. 2014. *The Politics of River Basin Organisations: Coalitions, Institutional Design Choices, and Consequences*. Cheltenham: Edward Elgar Publishing.

Huitema, D., E. Mostert, W. Egas, S. Moellenkamp, C. Pahl-Wostl, and R. Yalcin. 2009. "Adaptive Water Governance: Assessing the Institutional Prescriptions of Adaptive (Co-) Management from a Governance Perspective and Defining a Research Agenda." *Ecology and Society* 14 (1): 26. http://www.ecologyandsociety.org/vol14/iss1/art26/

Koetz, T., K. N. Farrell, and P. Bridgewater. 2012. "Building Better Science–Policy Interfaces for International Environmental Governance: Assessing Potential Within the Intergovernmental Platform for Biodiversity and Ecosystem Services." *International Environment Agreements*. 12: 1–21.

Leach, W. D., C. M. Weible, S. R. Vince, S. N. Siddiki, and J. C. Calanni. 2013. "Fostering Learning Through Collaboration: Knowledge Acquisition and Belief Change in Marine Aquaculture Partnerships." *Journal of Public Administration Research and Theory.* 24 (3): 591–622. doi:10.1093/jopart/mut011.

Lee, K. N. 1999. "Appraising Adaptive Management." *Conservation Ecology* 3: 3–16.

Lindblom, Charles Edward 1968. *The Policy-Making Process*. Englewood Cliffs, NJ: Prentice-Hall.

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

Mostert, E., C. Pahl-Wostl, Y. Rees, B. Searle, D. Tàbara, and J. Tippett. 2007. "Social Learning in European River-Basin Management: Barriers and Fostering Mechanisms from 10 River Basins." *Ecology and Society* 12 (1): 19. http://www.ecologyandsociety.org/vol12/iss1/art19/

Munaretto, S., and D. Huitema. 2012. "Adaptive Comanagement in the Venice Lagoon? An Analysis of Current Water and Environmental Management Practices and Prospects for Change." *Ecology and Society* 17 (2): 19. http://dx.doi.org/10.5751/ES-04772-170219

Muro, M., and P. Jeffrey. 2012. "Time to Talk? How the Structure of Dialog Processes Shapes Stakeholder Learning in Participatory Water Resources Management." *Ecology and Society* 17 (1): 3. http://dx.doi.org/10.5751/ES-04476-170103

Newig, J., D. Günther, and C. Pahl-Wostl. 2010. "Synapses in the Network: Learning in Governance Networks in the Context of Environmental Management." *Ecology and Society* 15 (4): 24. http://www.ecologyandsociety.org/vol15/iss4/art24/

Olsson, P., L. H. Gunderson, S. R. Carpenter, P. Ryan, L. Lebel, C. Folke, and C. S. Holling. 2006. "Shooting the Rapids: Navigating Transitions to Adaptive Governance of Social-Ecological Systems." *Ecology and Society* 11 (1): 18. http://www.ecologyandsociety.org/vol11/iss1/art18/

Ostrom, E. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.

Owens, Susan, Tim Rayner, and Olivia Bina. 2004. "New Agendas for Appraisal: Reflections on Theory, Practice, and Research." *Environment and Planning A* 36 11: 1943–1959.

Pahl-Wostl, C. 2006. "The Importance of Social Learning in Restoring the Multi-Functionality of Rivers and Floodplains." *Ecology and Society* 11 (1): 10. http://www.ecologyandsociety.org/vol11/iss1/art10/

Pettenger, M. 2007. "The Netherlands Climate Change Policy: Constructing Themselves/ Constructing Climate Change." In *The Social Construction of Climate Change. Power, Knowledge, Norms and Discourses*, edited by M. Pettenger, 51–73. Aldershot: Ashgate.

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

Richter, Brian D., Ruth Mathews, David L. Harrison, and Robert Wigington. 2003. "Ecologically Sustainable Water Management: Managing River Flows for Ecological Integrity." *Ecological Applications* 13, (1) : 206–224.

Rodela, R. 2011. "Social Learning and Natural Resource Management: The Emergence of Three Research Perspectives." *Ecology and Society* 16 (4): 30. http://dx.doi.org/10.5751/ES-04554-160430

Rondinelli, D. 1993. *Development Projects as Policy Experiments: An Adaptive Approach to Development Administration*. London: Routledge.

Saarki, S., J. Niemela, R. Tinch, S. van den Hove, A. Watt and J. Young. 2014. "Balancing Credibility, Relevance and Legitimacy: A Critical Assessment of Trade-Offs in Science–Policy Interfaces." *Science and Public Policy* 41: 194–206.

Sabatier, P. 1978. "The Acquisition and Utilization of Technical Information by Administrative Agencies." *Administrative Science Quarterly* 23: 286–411.

Sabatier, P. 1988. "An Advocacy Coalition Framework of Policy Change and the Role of Policy-Oriented Learning Therein." *Policy Sciences* 21: 129–168.

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

Sanderson, I. 2009. "Intelligent Policy Making for a Complex World: Pragmatism, Evidence, and Learning." *Political Studies* 57 (4): 699–719.

Schusler, T. M., D. J. Decker, and M. J. Pfeffer. 2003. "Social Learning for Collaborative Natural Resource Management." *Society and Natural Resources* 15: 309–326.

Social Learning Group. 2001. *Learning to Manage Global Environmental Risks*. Cambridge, MA: MIT Press.

Steffen, W. 2009. "Interdisciplinary Research for Managing Ecosystem Services." *Proceedings for the National Academy of Sciences* 106 (5): 1301–1302.

Tassey, G. 2014. "Innovation in Innovation Policy Management: The Experimental Technology Incentives Program and the Policy Experiment." *Science and Public Policy* 41 (4), 419–424.

Van den Hove, Sybille. 2007."A Rationale for Science-Policy Interfaces." *Futures* 39 (7): 807–826.

Van der Heijden, Jeroen. 2014. "Experimentation in Policy Design: Insights from the Building Sector." *Policy Sciences* 47 (3): 249–266.

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

Voß, J. P., and B. Bornemann. 2011. "The Politics of Reflexive Governance: Challenges for Designing Adaptive Management and Transition Management." *Ecology and Society* 16 (2): 9. http://www.ecologyandsociety.org/vol16/iss2/art9/

Voß, J. P., and A. Simons. 2014. "Instrument Constituencies and the Supply Side of Policy Innovation." *Environmental Politics* 23 (5): 735–754.

Walters, C. J., and C. S. Hollings. 1990. "Large-Scale Management Experiments and Learning by Doing." *Ecology* 71: 2060–2068.

Weber, M. 1968. *Economy and Society: An Outline of Interpretative Sociology*. New York: Bedminster Press.

Webler, T., H. Kastenholz, and O. Renn. 1995. "Public Participation in Impact Assessment: A Social Learning Perspective." *Environmental Impact Assessment Review* 15: 443–463.

Weiss, C. 1977. "Research for Policy's Sake: The Enlightenment Function of Social Research." *Policy Analysis* 3 (4): 531–545.