LEARNING THROUGH GAMES? EVALUATING THE LEARNING EFFECT OF A POLICY EXERCISE ON EUROPEAN CLIMATE POLICY

CONSTANZE HAUG[1] & DAVE HUITEMA

Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam

- ***DRAFT, PLEASE DO NOT CITE OR QUOTE WITHOUT AUTHORS' PERMISSION –***

ABSTRACT

One of the key arguments for the use of simulation-gaming approaches in policy appraisal has consistently been their potential to stimulate learning. Yet few studies seek to ascertain the learning effects of these methods for participants in a systematic manner; on the whole, learning through policy games remains both under-conceptualised and under-evaluated. This paper seeks to contribute to filling this gap by developing a typology of learning effects (cognitive, relational, and normative) that can be expected from policy games. We subsequently present a set of tools for measuring them and test our approach on the case of a policy exercise on burden sharing in future European climate policy involving policy-makers and experts.

On the basis of our measurements, we found limited evidence for learning from the policy exercise, most notably in the cognitive and the relational domain. An interesting methodological innovation is the use of concept maps. Employed as pre- and post-measurements, they proved a useful tool for tracing conceptual change through the exercise among participants. The paper concludes with a plea for more systematic assessment of the learning effects of simulation-gaming and participatory approaches to policy appraisal more generally, with a view to enabling a deeper discussion on the benefits and limitations of these methods.

---

[1] Corresponding author, Email: Constanze.haug@ivm.vu.nl

1.   INTRODUCTION

Most environmental policy deals with problems that are 'wicked'. They are complex, both scientifically and socially, and often also uncertain. These characteristics explain the long tradition of participatory, interactive approaches in environmental policy appraisal. The adaptive management literature seeks to embrace them by stipulating that policy-making and governance of complex environmental problems should be conceived as a collective and more or less permanent learning process [1]. Problem definitions should be examined from as many perspectives as possible, in broad participatory processes that are structurally evaluated and redesigned. Experimenting with different options is an inherent part of adaptive management, and any policy adopted should be flexible and adjustable [2] [3] [4].

Scenario-based methods fit well into this logic since, if adequately designed, they allow policy-makers and experts to develop and assess policy options in a 'safe' environment, potentially providing a means of tackling the challenges of long-term planning under deep uncertainty. Policy games add to this the value of emulating the interactive, social dimension of the problem at hand. Departing from a future qualitative plot or storyline, policy games assign roles to stakeholders or experts (which may be close to their own, but can also be very different), and confront them with a set of pre-defined tasks or decisions which they need to complete under time constraints in a fictitious scenario setting. Policy games have been frequently applied in the environmental domain, to topics ranging from municipal urban planning up to large-scale, long-term issues of global environmental change.

A key argument for the use of policy games has consistently been their potential to stimulate learning [5] [6]. The literature points to their benefits in terms of building and improving relationships among stakeholders and of fostering a holistic view of the problem, as well as to their suitability as a venue for problem-solving and for stimulating thinking about the medium- to long-term future. Yet while the learning effects of using games and simulations in an educational environment are relatively well studied in the literature [7] [8] [9] [10], learning through policy games in research or as aid for decision-making remains both under-conceptualised and under-evaluated. Published accounts of policy games often claim that they provided a valuable learning experience to participants. However, how and what participants have actually learned is rarely analysed in a systematic manner. The same applies to scenario planning; a recent review by the European Environment Agency on 'evaluative scenario literature' found only few studies covering this domain [11]. The present article aims to contribute to filling this gap, by presenting a framework for evaluating the learning effects of policy games and a set of instruments for measuring them. We subsequently test our approach on the case of a policy exercise on burden sharing in future European climate policy, which we designed and ran with policy-makers and experts.

Writings on learning in public policy often depart from three questions that Bennett and Howlett [12] used as signposts in a review of the literature in this field: "who learns?", "what is learned?" and "to what effect?". Among these, the question of "what is learned" through a policy exercise forms the focus of the present article. With regard to "who learns", while different groups of actors

(participants, game designers, the wider social network or the public) can potentially benefit from a policy game, the emphasis here is on the learning of participants through such an exercise. The question "to what end", i.e. whether policy learning ultimately leads to policy change, is very hard to detect in the case of a one-off intervention such as a policy game. It therefore falls outside of the scope of this article.

The next section provides a short review of the use of simulation-gaming approaches in public policy and of policy exercises as one specific sub-form of this method. Section 3 develops the concept of learning through policy games, introduces a typology of three types of learning effects that might expected from these and presents the methods that we used to measure them. Section 4 outlines the context and design of a policy exercise that we ran with stakeholders in European climate policy. Section 5 analyses the learning effects of this exercise. Section 6 discusses the implications of our findings and presents avenues for future research.

2.   POLICY GAMES AND EXERCISES IN PUBLIC POLICY

The origins of policy games, also referred to as 'simulation-gaming', 'free-form gaming', or 'politico-military gaming' or 'war gaming', lie in the military sphere. After first examples going back as far as the First and Second World Wars [13], and applications in related domains like international relations and crisis management from the 1950s onwards, policy gaming has increasingly been applied to environmental and social policy issues since the 1970s. Policy exercises, first advocated by Brewer [14] and further developed by Toth [15] [16], are one sub-form of policy games. They responded to the need for adapting the method of model-based gaming from a military or business context to the challenges of long-term policy development in complex socio-ecological systems. In Toth's words, "at the heart of the process [of the policy exercise] are scenario writing of 'future histories' and scenario analysis via the interactive formulation and testing of alternative policies that respond to challenges in the scenarios. These scenario-based activities take place in an organizational setting reflecting the institutional features of the problem at hand" [16: 237].

The structure of policy exercises is less rigid than for policy games. Their primary goal is to advance thinking about unstructured, messy problems for which the set of relevant choices, cause-effect relationships and outcomes is controversial or unclear. Thus, while participants in a policy exercise are also tasked to make plans and decisions under a fictitious scenario, the storyline is usually richer and more elaborate than for other types of policy games. Moreover, policy exercises rarely involve fixed scoring systems that reward or penalize actions by participants. Policy exercises have been applied to issues of global environmental change on several occasions, but also to health, education and regulatory policy reform (see [17] for examples). A number of applications have focused on various aspects of climate policy, from global negotiations to local adaptation options [18] [19] [20] [21] [22].

According to Parson [23], policy games draw their usefulness from combining two characteristics, representation and deliberation. The basis for any policy game is a simplified model of reality or 'system of reference', which, however, retains the main actors, relationships and cause-effect chains. Policy games, like game theory and modelling exercises, thus aim at reducing complexity

through structured abstraction. At the same time, by relying on interaction between real-life stakeholders instead of agents with predefined preferences, they incorporate a strong deliberative element. Participants jointly explore the future 'possibility space', building up a shared understanding of key concepts and searching creatively for solutions [6]. Therefore, in order to make full use of the potential of the method, in line with the participatory appraisal literature, participants in a policy exercise should ideally be diverse, both in terms of (disciplinary) backgrounds and viewpoints on the problem at hand [24].

As stated above, there are strong claims in the literature about the learning potential of policy games, which is often attributed to the intense, vivid experience that participants go through in the course of such an exercise. Yet, what types of learning can actually be expected from a policy game? And to what extent can we substantiate that this learning actually occurs among participants? This will be the subject of the next section.

3.   LEARNING THROUGH POLICY EXERCISES

3.1   POLICY LEARNING AND SOCIAL LEARNING

There is a rich literature on learning in public policy and management, which draws on a range of disciplines and perspectives, from organisation theory to policy analysis and cognitive psychology. At a general level, learning is mostly conceived as implying a change in thought, usually the result of new knowledge or past experience. We take Sabatier's definition as a starting point; he defines policy-oriented 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: 131].

Most approaches to learning further have in common that they distinguish between different degrees and types of learning [26]. In terms of types of learning, writings on policy learning tend to emphasise its cognitive dimension, which we define here as the acquisition, enhancement or revision of new or previously gained knowledge. The public participation literature, analysing group dynamics and social learning in stakeholder-based processes, offers a useful extension of this concept; Webler et al. [27] distinguish between two components of social learning, 'cognitive enhancement' of participants (i.e. the acquisition of knowledge) and 'moral development', focusing on the interactive, inter-personal dimension of appraisal.

Furthermore, many authors differentiate between a technical level and one or two 'higher', conceptual levels at which learning can take place [28] [29] [30] [31]. For Argyris, for instance, single-loop learning is "when a mismatch is corrected without changing the underlying values and status quo that govern the behaviours", whereas double-loop learning means that the mismatch is "corrected by first changing the underlying values and other features of the status quo" [32: 1087f]. With a view to developing a typology specifically suited to examine learning from policy games, we conceptualise three types of learning that participants may take away from them:

(1) *Cognitive learning*, which refers to the acquisition of new or the improved structuring of existing knowledge. Cognitive learning roughly corresponds to the concepts of first-order or single-loop learning developed in the policy learning literature.

(2) *Normative learning*, which implies changes in the viewpoints, norms and values of participants as a consequence of participating in the policy exercise. It is closely related to second-order or double-loop learning, which also incorporates a strong normative dimension. However, in our view, in the context of interactive appraisal exercises, normative learning is not in principle of a higher value than its cognitive counterpart, but rather stands on an equal footing next to it.

(3) *Relational learning*. This term is used for instance in the literature on 'collaborative management' [33]. Relational learning refers to questions such as increased trust, improved ability to cooperate and a better understanding of the mindsets and frames of other participants.

Inevitably, any assessment of learning begs the question of the 'baseline' – i.e. do certain lessons from a policy game that the evaluator considers undesirable constitute not 'learning', but rather the opposite – 'unlearning'? We seek to circumvent this by considering any shift in the minds of participants that can be attributed to the policy exercise as learning.

3.2 INSTRUMENTS FOR EVALUATING LEARNING

A key dilemma in measuring the learning effect of interactive appraisal exercises consists in the tension between the fact that learning occurs through intense interactions in a social setting, yet that measurements are in principle only possible at the level of the individual. It is probably impossible to resolve this tension completely, but a combination of group-based (such as the debriefing session) and individual assessments may go some way in addressing it. We follow Blackstock et al. [34: 732] in our approach, who recommend a combination of recorded and reported data for this purpose, "gathered at multiple points throughout the process". Thus, we have taken measurements prior to the policy exercise as well as immediately afterwards, complementing those with process observations during the workshop. Ideally, there should be a further evaluative moment at a later stage to account for the fact that some learning might occur only upon reflection, some time after the exercise. However, since at that stage, the specific learning effect of the exercise may be difficult to distinguish from the impact of other experiences when using recorded measurements, we relied on self-reported data through interviews with participants for this purpose. The lack of a control group represents a limitation to our methodology which prevented us from controlling for the learning effect that the evaluation instruments themselves may have on respondents (the 'Hawthorne effect').

Overall, we have used the following instruments:

- Written ex ante and ex post surveys eliciting participants' views on facts and normative statements about burden-sharing in EU climate policy, the topic of the policy exercise. A similar approach was used by Groves et al. [35], testing views of Californian water managers on climatic and management uncertainties, as well as by Chermack et al. [36] who sought to assess improvements in 'strategic conversation quality' as a consequence of participating in scenario workshops. Our questionnaires further inquired about participants' backgrounds, motivation for participating in the policy exercise and their perception of what they had

learned from it. Twelve persons filled in both pre and post questionnaires, resulting in a response rate of 54%.

- Participants were further asked to draw concept maps of 'burden sharing in European climate policy' both prior to the exercise and immediately afterwards. A concept map is a structural representation "consisting of nodes representing concepts and [...] lines denoting the relation between a pair of nodes." [37: 569] Concept maps, also known as mind maps, causal or cognitive maps, have been used for a variety of purposes, from creative brainstorming to project management and as planning tools, but also have a history as assessment tools, especially in higher education [37] [38]. Various methodologies have been developed for tracing conceptual change through concept maps, most prominently by Morine-Dershimer [39]. In our case, while the response rate for the ex ante maps was close to 100%, the rate afterwards was considerably lower as many participants needed to get home quickly after the exercise. Effectively, this means we received five sets of comparable ex ante and ex post concept maps. However, they covered all groups of participants (policy-makers, academics, consultants, private sector) and can thus be considered relatively representative of the participants of our policy exercise.

- Process observations and tape recordings. During the exercise, two members of the organising team took notes on the interactions between participants, focusing mainly on the style of interactions and on activity patterns of individual participants. The tape recordings were equally used to determine activity patterns by participants and to check the observations and notes taken during the workshop.

- A debriefing session, generally considered a crucial element of scenario and simulation-gaming exercises, allowed participants to share their experiences and to collectively discuss what they had learned from the exercise.

- Last but not least, in-depth telephone interviews with ten participants were conducted two to four weeks after the workshop, prompting them to reflect on their experience and on the strengths and limitations of the policy exercise method.

Having developed a typology of learning effects that can be expected from policy games and presented a set of tools for measuring them, we now proceed to introduce our empirical case, a policy exercise on a topic of European climate policy.

4. THE CASE: A POLICY EXERCISE ON BURDEN SHARING IN FUTURE EUROPEAN CLIMATE POLICY

The policy exercise that we designed and organised took place in the context of a large research project on options for future European climate change policy, ADAM.[2] Over the past years, climate change has become an increasingly high profile issue in the European Union (EU) and has given rise

---

to significant policy activity at various levels of governance. Our policy exercise focused on one key element of the European climate policy architecture, namely the challenge of distributing greenhouse gas emission reduction efforts among EU Member States. Under the Kyoto Protocol, the European Community signed up to a joint emission reduction objective, which subsequently needed to be translated into targets for individual Member States. This process is known as 'burden sharing'. More recently, it has been relabelled into 'effort sharing' in EU policy discussions addressing this question for the post-2012 period. Burden sharing requires striking a balance between reducing emissions where they are most cost-effective while taking into account the diversity of socio-economic circumstances across the EU. Increasing differences in wealth and capacity within an expanding Union on the one hand, and the need for ever more stringent emission reduction commitments on the other suggest that the basic underlying problem of how to allocate burdens is there to stay in European climate policy [40]. In the future, more sophisticated and robust arrangements are likely to be necessary to ensure that the EU can live up to its aspiration for leadership in the global climate arena. Crucial questions in the design of burden sharing arrangements relate to the base year from which emission reductions are accounted for, the criteria on which to base the targets, as well as a variety of issues with regard to implementation, monitoring and enforcement.

The main objective of the policy exercise was to appraise options for future EU burden sharing arrangements, to assess their feasibility and to identify associated challenges and trade-offs. At the time of the exercise, negotiations were still ongoing on the EU Climate and Energy Package, the intended centrepiece of the EU's efforts to cut emissions over the period 2012-2020, which also includes a decision on 'effort sharing'. Our goal, however, was to look even further ahead and explore how this policy domain could evolve after 2020. Thus, participants, playing the roles of senior decision-makers from EU Member States and the European Commission, were tasked to negotiate key features of a post-2020 burden-sharing agreement in a simplified EU policy-making cycle. Their discussions were structured through 'policy element cards' that contained design alternatives for the central features of such an agreement. Issues requiring a decision included the criteria for burden sharing (e.g. emissions or GDP per capita, available emission abatement potentials, sectoral emissions, etc.), enforcement mechanisms as well as the extent to which member states could substitute domestic emission cuts with purchases of external carbon credits.

In order to explore how different political and economic circumstances could impact on the results of the negotiations, participants were divided into two subgroups that went through the same steps of play yet were confronted with two different scenarios, both set in the year 2018. The scenario in the first group reported a high degree of international cooperation on climate change under a successful successor agreement to the Kyoto Protocol. In the other scenario, emissions abatement was occurring through a variety of uncoordinated bilateral and multilateral initiatives and efforts at various levels. Remarkably from a methodological perspective, while the format stimulated animated in-depth discussions in both groups, the significant differences between the two scenarios hardly had any impact on the negotiation results in the two groups. The policy exercise did generate interesting insights into crucial 'deal-makers' and 'deal-breakers' in the context of burden sharing, however. These substantive outcomes are not further discussed here, but are analysed in more detail in [41].

Twenty-three participants from all over Europe participated in the policy exercise. Most of them were mid-career professionals with long-standing experience in climate policy, coming from backgrounds ranging from academia to policy-making and the private sector. Based on the information they provided in the ex ante surveys, their level of expertise on European climate policy and burden-sharing varied from medium to very high, with one of the participants being the key resource person at the European Commission for the draft decision on 'effort sharing' in the EU Climate and Energy Package.

The participants played in five teams, representing senior policy-makers of four EU member states (Germany, Poland, Spain and Sweden) and the European Commission. We decided to include only a limited number of countries in the exercise since we wanted to allow for deliberation inside the country teams while still having key interests and conflict lines on EU burden sharing represented within the subset of countries we had chosen. Before the policy exercise, participants received short role descriptions summarising important developments in their country's climate policy and economic outlook over the last decade. This information allowed them to deduce the issues that had priority for them in the negotiations. In assigning the roles, care was taken to match participants' own backgrounds with the nationalities they were representing. Thus, Sweden was represented by two Swedes and the role of Poland played by two participants based in Central and Eastern Europe. Moreover, a professional journalist published regular 'news bulletins' on the progress in negotiations and on positions adopted by Parties, providing the organising team with an opportunity to influence the dynamics of the game where needed and to feed external events into the game. The policy exercise itself lasted for a full working day. It was preceded by a dinner on the eve of the workshop, during which the policy exercise approach and objectives were outlined. During the dinner, one of the architects of the previous EU burden sharing agreement - a former chief climate negotiator of the Netherlands - gave a talk on his experience at the time, and participants had the chance to get to know each other.

The following section will discuss the evidence for cognitive, normative and relational learning among participants as a consequence of the policy exercise that we described above.

5. LEARNING THROUGH THE ADAM POLICY EXERCISE

5.1 COGNITIVE LEARNING

As stated earlier, we conceptualise cognitive learning as the acquisition of new and the better structuring of existing knowledge. In order to evaluate this dimension, we made use of the concept maps drawn by participants before and after the policy exercise. Furthermore, we examined the responses to questions on the self-reported learning effect in the ex post survey and in follow-up interviews.

The evaluation of the concept maps is based on a method initially developed by Morine-Dershimer [39]. It relies on two principles, (1) centrality – the proximity of concepts to the core of the map, which can be taken as an indicator how important they are in the perception of the author and (2) specificity – the extent of detail with which a concept is worked out in sub-concepts, or subordinate

branches. A comparison of shifts in the centrality and specificity of concepts from the pre- to the post-measurements allows tracing changes in the structuring of knowledge or a new prioritisation of certain aspects. All concepts on the five available sets of concept maps were first coded based on twelve different answer categories, including for example the 'politics of burden sharing', 'state of the physical climate', 'implementation of burden sharing agreements' or 'adaptation to climate change'. Subsequently, scores for the average centrality and specificity of these categories for all ex ante and all ex post maps were calculated based on Morine-Dershimer's system. Figure 1 depicts the shifts that occurred on participants' concept maps between the pre-and post-measurement, limited to four out of the twelve response categories for the sake of graph clarity and readability (n=5). The more a concept moves to the left on the graph, the more central it has become from the ex ante to the ex post maps; the more it moves upward, the more specific its treatment.
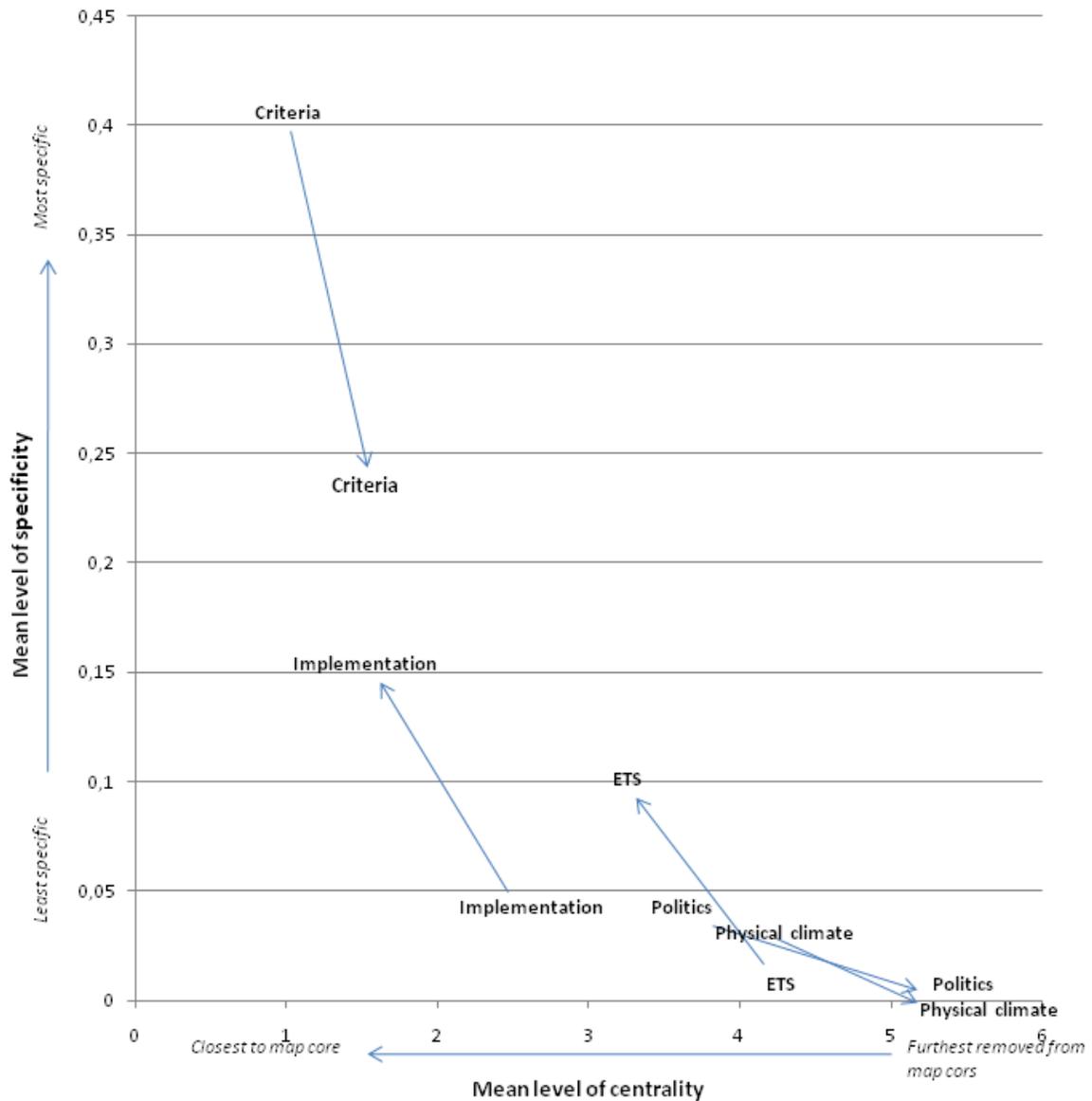
Figure 1. Pre- and post-measurements of centrality and specificity of four selected concepts on participants' concept maps.

Overall, conceptual shifts are clearly detectable on the concept maps. Especially issues related to the implementation of burden sharing arrangements and to the European emissions trading system (EU ETS) became both more central and specific on the ex post maps, while the 'physical climate' and 'politics' surrounding burden-sharing received relatively less attention afterwards. This is in line with the set-up of the policy exercise, where discussions circled extensively around the design of future burden sharing agreements and on interactions with the EU ETS.[3] Issues that stood central during

---

[3] From 2012 onwards, the European economy is divided into two sectors

the workshop thus moved up in both evaluation categories, testifying to some degree of cognitive learning. The size of the sample was too small to test for statistical significance, however.

In addition, we relied on information on the cognitive learning effect of the exercise as self-reported by participants in the ex post surveys and interviews. Several persons emphasised that the policy exercise had provided them with a better understanding of one issue that had previously received very little attention in the context of the negotiations on the 2008 EU Climate and Energy Package, namely the crucial importance of the criteria on which burden sharing agreements are based. Furthermore, some participants indicated in the interviews that the policy exercise had prompted them to reflect more about the requirements of long-term policy-making, and about dilemmas of compliance and enforcement. Beyond this, however, the majority agreed that the exercise had not been particularly successful in generating new policy options and in stimulating a truly future-oriented debate.

5.2 NORMATIVE LEARNING

Earlier, we defined normative learning as changes in viewpoints, norms and values. In order to evaluate this dimension, we relied on the pre- and post-surveys as well as the participants' own perceptions as reported in the ex post interviews. Both the pre- and post survey questionnaires included a list of twenty propositions related to burden sharing and to European climate policy more generally, prompting participants to indicate the extent to which they agreed or disagreed with the points made. The list contained normative statements like "the European Commission protects the interests of the less developed countries too much" or "the wealthier Member States care too little about economic development of the poorer countries." By including the same propositions in the ex ante and ex post surveys, we attempted to measure the change in participants' attitudes over the course of the exercise. We did so by comparing the standard deviations of the pre- and post-measurements. If the standard deviation diminishes from the pre- to the post-survey, arguably, normative convergence in the group increased. By contrast, where it increases, group opinion diverges even more following the exercise. Given our small sample size (n=12), we did not find variation that reached the level of statistical significance, except for a few likely coincidental comparisons. Furthermore, below the level of statistical significance, group opinion diverged on just as many statements after the exercise as it converged. Thus, overall the results from the survey do not provide evidence that normative learning has taken place. Groves et al. [35] also report rather small effects based on a similar assessment method.

The limited extent of normative learning visible from the pre- and post statements is supported by respondents' self-assessment in the survey. Asked whether their views on burden-sharing had changed as a result of the exercise, 44% stated that participating had "not really" changed their views on effort-sharing, 25% responded that it had altered it "a bit", with a further 25% responding that it had done so "to some extent". Similarly, in one of the ex post interviews, a respondent stated that the exercise had "reaffirmed [his] view on how EU policy-making operates". The self-reported normative learning effect was correlated with the reasons participants stated in the survey for their participation in the policy exercise. Those whose motivation was linked to the innovative format of the workshop and the opportunity for networking, rather than the interest in EU policy-making and climate policy options, were also more prone to report a change in viewpoint.

Some qualifications may apply to this finding. First, the survey statements may not be the best means for assessment as participants may tend to simply tick the same boxes as previously (given the short time interval between the pre- and the post-measurement) instead of genuinely reconsidering the propositions. Similarly, given the high level of expertise of the participants, almost all of whom considered themselves highly knowledgeable in this domain, there may be a certain reluctance to admit to normative learning in their area of specialisation. Last but not least, one might hypothesise that, other than cognitive learning, some normative learning effects are only realised much later or that they are simply not to be expected at all from a short one-off intervention like a policy exercise. Interestingly, however, in the case of one participant, a comparison between the pre- and post- concept map provided strong evidence for normative learning. Following an ex ante map that largely focussed on technical design issues related to burden sharing, one of the main branches on the post map read 'remove focus from effort sharing', indicating a rather fundamental shift in viewpoint, namely that no matter the details of its implementation, burden sharing was unlikely to ever become an environmentally effective climate policy strategy.

5.3   RELATIONAL LEARNING

Relational learning focuses on the interactive, social dimension of the policy exercise – whether participants gained a better understanding of their counterparts' mindsets and whether trust and ability to cooperate in the group increased as a consequence. However, since the participants in our policy exercise were not part of an existing network (the ex post survey indicated that the average participant only knew three other persons taking part), it was impossible to trace how the group as a whole changed through the exercise. Assessing this dimension would have been possible if we had made affiliation with an existing network a selection criterion for participation.

Thus, in our case, the learning benefits rather concerned other relational aspects, most notably increased understanding of the complexity and constraints of the EU negotiation process as well as the various ways in which the problem of burden sharing is perceived by different actors. 53% of participants felt that the policy exercise had increased their understanding of the dynamics of EU policy-making. Process-related answers were also the single most frequent response to the open survey question prompting the three 'key insights' that participants were taking away from the policy exercise. As one participant wrote, "Decision-making always has to take place with insufficient information. 'Fairness' is a social construct, interpretable in multiple ways. Grandstanding in negotiations may disguise basic willingness to cooperate."

For many participants, the biggest realisation concerned the diverging problem definitions and problem perceptions of the various actors. As one participant stated, "you get a real sense of how difficult these negotiations are and how you have to guess what kind of vested interest the other countries have. In my daily work I focus on environmental benefits and you don't try to understand so much why countries are against something but you just think the countries is [sic] not doing their job. It makes you realise what the agenda is of these countries and how these interests play together." Others stressed that the policy exercise had made them especially aware of the difficulty of the European Commission's mandate. As one person commented, "you can bring in little as the Commission since reaching any kind of agreement is your very first priority."

Overall, participants agreed that the policy exercise format had brought about a very different kind of interaction. They said it had created an atmosphere "full of fun and respect", which made it impossible "to stay out of the discussions or give very little input" and which meant that everyone was "forced to share their expertise". This in itself can be considered an important achievement, especially in a domain like climate policy where an effective interface between scientists and decision-makers is surely a *sine qua non* for developing policies that are effective and equitable also in the longer term.

6.   CONCLUSIONS

The purpose of this article was to introduce a framework for assessing learning through simulation gaming approaches in policy appraisal and to test it on the example of a policy exercise on European climate policy. Its main contribution should be seen in view of the need for a more systematic evaluation of these methods. While claims are often made about the significant learning potential of policy games for participants, the latter is rarely analysed in detail nor assessed in a systematic manner in the literature. This is hardly surprising; as Duke and Geurts [42: 211] remark, in most applications, "research efforts have to be interwoven with the uses of games in applied settings. As a consequence, most empirical research on policy gaming is a compromise and it is often quite limited from a research-design perspective." We feel that the evaluation of policy games, and their learning benefits more specifically, does warrant more attention than it has received – or than has been documented – to date. Our conceptualisation of the three different types of learning through policy games, based on the literatures of policy and social learning, may provide a first step towards more analytical rigour in this regard. Conceivably, this typology, as well as the evaluation instruments which we introduced, may be adapted to suit a wider ranger of applications in the domain of participatory appraisal. Surely, if social learning is proclaimed to be such a crucial goal for these kinds of activities, more care should be expended to provide evidence for their effectiveness in achieving it.

We also described a set of instruments by which learning through policy games can be measured, and we subsequently applied it to the case of a policy exercise that we developed and ran. Notwithstanding the practical difficulties, there is clearly promise in combining different sorts of methods and data for this purpose, allowing for cross-checking and substantiation of the evaluation results. An interesting methodological innovation in our view is the use of concept maps. Employed as pre- and post-measurements, they proved a useful tool for tracing conceptual change among participants through the exercise. Their strength lies in the fact that they can account also for subtler cognitive changes that participants may not necessarily be conscious of (and that would thus be missed by self-evaluation questions in surveys or interviews), as long as these can be detected at the group level. However, there remains a strong role for qualitative methods, especially ex post interviews, in the evaluation of learning effects. As Blackstock et al. [34: 732] emphasise, qualitative methods "'allow the study of a case in depth and detail, capturing the richness of people's perceptions and experiences in their own terms and developing an analytical understanding through the aggregation of these individual accounts." The main weakness of our research relates to the limited number of observations, especially with regard to the concept maps, on which our

measurements were based. Collecting more robust evaluation data will likely remain a challenge also in further assessments of this kind; it is always difficult to convince the rather high-powered participants of a policy game to partake in yet another add-on to an already strenuous workshop. This may equally call for further methodological innovations that enable data collection to be embedded better into the policy game itself, instead being perceived as a tiresome additional assignment. On a different note, we have yet to make use of a control group, which might strengthen our findings. The set-up of the policy exercise was also not suitable for more sophisticated observations regarding its relational learning effects since few of the participants knew each other beforehand.

How to interpret our findings in the broader context of discussions on the role of policy exercises in policy appraisal? In certain regards, our research gives reason for optimism. However short the policy exercise, we found proof for both cognitive and relational learning, even though the number of available observations was too small to yield really hard evidence for the former, while evidence for the latter was even weaker. The greatest benefits of the exercise were process-related, by providing insights into EU policy-making and the art of negotiation, as well as the constraints that the various actors are operating under. Given the urgent need for an effective interface between science and policy in the climate change domain, this in itself may be considered justification enough for the use of the method. According to the ex post survey, participants perceived the short duration of the policy exercise and the over-structuring of discussions through the 'policy element' cards (which, while considered useful in principle, were found to be too detailed and thus distractive by many) as the greatest obstacles to their personal learning experience. Therefore, a longer time frame and minor changes to the design of the exercise might have further increased its effectiveness.

Two other aspects are more worrisome. In terms of cognitive learning, participants found the exercise to be of limited value in projecting them into the future and in generating new policy options, both presumably key objectives when employing this method. Based on our observations, participants did make use of the future setting and the news bulletins from the year 2018 as a rhetorical device in interactions, but substantial reasoning on concrete options and strategies remained very much tied to the present-day situation. However, this is a problem that is encountered also in the context of other tools for strategic foresight, such as backcasting [24]. Turning to normative learning, we were not able to detect significant changes from the pre- to the post-survey measurements. This conclusion was largely supported by the self-evaluation of participants. If these tentative findings are confirmed by future research, one may conclude that, depending on their design, the strengths of policy exercises relate more to the cognitive and relational than the normative dimension. This would be regrettable given the importance of 'reflexivity', i.e. the willingness and ability to reconsider established goals, policy priorities and existing knowledge in a domain like climate policy where complexity and uncertainty abound. This also begs the question of the intended role of the designer of a policy exercise.

Future research could be directed towards a more refined assessment of relational learning. Moreover, it may be interesting to investigate in more depth the longer-term effects of policy games, in spite of the difficulties of distinguishing their impact from the variety of other stimuli that participants are subjected to in the meantime. The use of a control group may allow attenuating the

difficulties associated with such a longitudinal set-up. Last but not least, comparing the impact of different design options on the learning effects could be interesting, as well as comparison between groups with greater and lesser expertise. A control test is currently under way as we are preparing to run the same policy exercise with master students in environmental resource management; we expect the cognitive and normative learning effects from the exercise to be greater in this second case.

References

1. C. Pahl-Wostl, M. Craps, A. Dewulf, E. Mostert, D. Tabara, and T. Taillieu, social learning and water resources management, Ecology and Society 12(2) (2007) 5.

2. K. N. Lee, appraising adaptive management, Conservation ecology 3(2) (1999) [online] http://www.ecologyandsociety.org/vol3/iss2/art3/.

3. D.R. Armitage, F. Berkes, and N. Doubleday, adaptive co-management. collaboration, learning, and mult-level governance, UBC Press, Vancouver, 2007.

4. D. Huitema, E. Mostert, W. Egas, S. Moellenkamp, C. Pahl-Wostl, and R. Yalcin, 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) (2009) 26.

5. I. Wenzler and D. Chartier, why do we bother with games and simulations: an organizational learning perspective, Simulation Gaming 30(3) (1-9-1999) 375-384.

6. F. Joldersma, J.L.A. Geurts, and W.J.H. van't Spijker, spelsimulatie. theorie, definitie en plaatsbepaling in de bestuurskunde, Bestuurskunde 4(4) (1995) 148-156.

7. B. D. Ruben, simulations, games, and experience-based learning: the quest for a new paradigm for teaching and learning, Simulation Gaming 30(4) (1-12-1999) 498-505.

8. W. C. Kriz, creating effective learning environments and learning organizations through gaming simulation design, Simulation Gaming 34(4) (1-12-2003) 495-511.

9. R. Fortmuller, learning through business games: acquiring competences within virtual realities, Simulation Gaming (22-10-2007) 1046878107308075.

10. J. Gosen and J. Washbush, a review of scholarship on assessing experiential learning effectiveness, Simulation Gaming 35(2) (2004) 270-293.

11. A. Volkery and T. Ribeiro, scenario planning in public policy: understanding use, impacts and the role of institutional context factors, Technological Forecasting and Social Change In Press, Corrected Proof.

12. C. J. Bennett and M. Howlett, the lessons of learning - reconciling theories of policy learning and policy change, Policy sciences 25(3) (1992) 275-294.

13. G.D. Brewer and M. Shubik, the war game. a critique of military problem solving, Harvard University Press, Cambridge, Mass, 1979.

14. G.D. Brewer, Methods for Synthesis: Policy Exercises, in: W.C. Clark and R.E. Munn (Eds.), Sustainable Development for the Biosphere, Cambridge University Press, Cambridge, 1986.

15. F. L. Toth, policy exercises: objectives and design elements, Simulation Gaming 19(3) (1-9-1988) 235-255.

16. F. L. Toth, policy exercises: procedures and implementation, Simulation Gaming 19(3) (1-9-1988) 256-276.

17. D. Crookall and K. Arai, simulation and gaming across disciplines and cultures, Sage Publications, Thousand Oaks, 1995.

18. j. jäger, n. sonntag, d. bernard, and w. kurz, the challenge of sustainable development in a greenhouse world: some visions of the future. report of a policy exercise at bad bleiberg, austria, september 2-7, 1990, Stockholm Environment Institute, Stockholm, 1990.

19. J. Klabbers, R. Swart, A. van Ulden, and P. Vellinga, Management of Organized Complexity Through Gaming, in: D. Crookall and K. Arai (Eds.), Simulation and Gaming Across Disciplines: ISAGA at a Watershed, Sage Publications, Thousand Oaks, 1995.

20. E.A. Parson, global climate change policy exercise: results of a test run1995.

21. B. Kasemir, F. Toth, and V. Masing, Venture Capital and Climate Policy, in: B. Kasemir, j. jäger, C.C. Jaeger, and M.T. Gardner (Eds.), Public Participation in Sustainability Science: A Handbook, Cambridge University Press, Cambridge, 2003.

22. K. G. Lonsdale, T.E. Downing, R.J. Nicholls, D. Parker, A.T. Vafeidis, R. Dawson, and J. Hall, plausible responses to the threat of rapid sea-level rise in the thames estuary, Climatic Change 91(1) (2008) 145-169.

23. E. A. Parson, informing global environmental policy-making: a plea for new methods of assessment and synthesis, Environmental modelling and assessment 2(4) (1997) 267-279.

24. M. van de Kerkhof, debating climate change, Lemma, Utrecht, 2004.

25. P. A. Sabatier, an advocacy coalition framework of policy change and the role of policy-oriented learning therein, Policy sciences 21(2-3) (1988) 129-168.

26. A. Swartling and M. Nilsson, social learning and epi: communicative governance in swedish climate policy formation, Ecologic, Berlin, 2007.

27. A. Webler, H. Kastenholz, and O. Renn, public participation in impact assessment: a social learning perspective, Environmental impact assessment review 15(5) (1995) 443-463.

28. C. M. Fiol and M.A. Lyles, organizational learning, The Academy of Management Review 10(4) (1985) 803-813.

29. P. Glasbergen, Learning to Manage the Environment, in: J. Meadowcraft (Ed.), Democracy and the Environment: Problems and Prospects, Edward Elgar, Cheltenham, UK, 1996.

30. P. A. Hall, policy paradigms, social learning, and the state: the case of economic policymaking in britain, Comparative Politics 25(3) (1993) 275-296.

31. C. Argyris and D.A. Schön, organizational learning. 2. theory, method, and practice, Addison-Wesley, Reading, MA, 1996.

32. C. Argyris, a life full of learning, Organization Studies 24(7) (1-9-2003) 1178-1192.

33. M.T. Imperial and T. Hennessey, environmental governance in watersheds. collaboration, public value and accountability, Paper presented at the Twenty-First Annual Research Conference of the Association for Public Policy Analysis and Management, 4-6 November, 1999, Washington, D.C., USA, 1999.

34. K. L. Blackstock, G.J. Kelly, and B.L. Horsey, developing and applying a framework to evaluate participatory research for sustainability, Ecological Economics 60(4) (1-2-2007) 726-742.

35. D.G. Groves, D. Knopman, R. Lempert, S.H. Berry, and L. Wainfan, presenting uncertainty about climate change to water-resource managers. a summary of workshops with the inland empire utilities agency, RAND Corporation, Santa Monica, CA, 2008.

36. T. J. Chermack, L. van der Merwe, and S.A. Lynham, exploring the relationship between scenario planning and perceptions of strategic conversation quality, Technological Forecasting and Social Change 74(3) (2007) 379-390.

37. M. A. Ruiz-Primo and R.J. Shavelson, problems and issues in the use of concept maps in science assessment, Journal of Research in Science Teaching 33(6) (1996).

38. J.D. Novak, learning, creating, and using knowledge. concept maps as facilitative tools in schools and corporations, L. Erlbaum Associates, Mahwah, N.J, 1998.

39. G. Morine-Dershimer, tracing conceptual change in preservice teachers, Teaching and Teacher Education 9(1) (1993) 15-26.

40. C. Haug and A. Jordan, Burden Sharing: Distributing Burdens or Sharing Efforts?, in: A. Jordan, D. Huitema, H. van Asselt, T. Rayner, and F. Berkhout (Eds.), Climate Change Policy in the European Union: Confronting the Dilemmas of Adaptation and Mitigation?, Cambridge University Press, Cambridge, UK, 2010.

41. J. Stripple, R. Hildingsson, A. Jordan, T. Rayner, and C. Haug, Governance Choices and Dilemmas in a Warmer Europe: Exploring the Future, in: A. Jordan, D. Huitema, H. van Asselt, T. Rayner, and F. Berkhout (Eds.), Climate Change Policy in the European Union: Confronting the Dilemmas of Adaptation and Mitigation?, Cambridge University Press, Cambridge, UK, 2010.

42. R.D. Duke and J. Geurts, policy games for strategic management, Dutch University Press, Amsterdam, 2004.