
34. Using experiments to study democratic innovations

Kimmo Grönlund and Kaisa Herne

1 INTRODUCTION

This chapter focuses on the use of the experimental method to study democratic innovations. Needless to say, we see experimentation as a fruitful way to examine the specific mechanisms of attitude change and processes of discussion related to democratic innovations. However, we will also pay attention to some difficulties and challenges of experimental research. Our focus is in a specific case of democratic innovations, mini-publics (somewhat representative samples of citizen deliberating on political issues), mostly because experimental research has mainly concerned mini-publics (see Chapter 3 in this Handbook). In the beginning of the chapter we characterise what experiments are, then provide some examples of experiments and finish with some thoughts about the challenges of experimental research and some suggestions about how to handle these challenges.

The use of experiments in the social sciences has increased rapidly over the past few decades, in particular in economics and political science. In their recent book on political science experiments, Morton and Williams (2010) pay attention to a huge increase in experimental articles in the top political science journals (APSR, AJPS, JP). However, they also remind us about the long history of experimentation in political science (e.g. Gosnell, 1926). In political science, the topics of investigation range from tests of formal, often game theoretic models (e.g. Plott and Levine, 1978) to media influence (e.g. Ansolabehere and Valentino, 1994) and studies of democratic innovations (e.g. Grönlund et al, 2015). Following the general trend, experimental research on democratic innovations has also seen a rapid growth over the past few years. As noted above, experiments on democratic innovations have mainly concerned deliberative mini-publics, i.e. more or less representative samples of the public deliberating on specific political issues. Other types of democratic innovations have been rarely tested with the experimental method. The reason for the focus on mini-publics is likely to be based on their usefulness to test issues, such as opinion change due to deliberation, discussion characteristics, as well as various ‘side effects’ of deliberation, for example, political trust (see Chapter 8 in this Handbook), social trust and political efficacy (Chapter 11). Moreover, other types of democratic innovations, such as participatory budgeting (Chapter 5), citizens’ initiatives or referenda (Chapter 6), are harder to organise without a connection to the actual political decision-making process. The involvement of actors, such as states or municipalities is, in turn, likely to decrease the possibilities for scholars to design experimental manipulations freely. The aim of most deliberative mini-publics is to study the consequences of participation in democratic discussion. Experiments with mini-publics typically have an interest in the influence of different types of organisational arrangements on participants’ opinions and attitudes (e.g. Grönlund, 2016).

An experiment is a controlled study of a causal relationship. An experiment investigates a dependent variable, which is influenced by a certain number of independent variables. In

an ideal case, we know what those variables are, usually informed by theory and/or previous research. An ideal experiment holds $n-1$ independent variables constant and manipulates only one. When other variables are held constant, we are able to say that any variation in the dependent variable is due to variation in the manipulated independent variable. Controlled variation, which is the key to testing cause-effect relations, is possible only in the conditions of a laboratory. In the social sciences the laboratory is, however, unlike laboratories in the natural sciences. It is usually simply a classroom, often equipped with computers, but not necessarily even that.

The reason for using experiments is that they provide control of variables not available through naturally occurring data, and that control of variables makes it possible to detect causal mechanisms. For example, if we want to know how participation in a mini-public influences individual opinions, organising an experiment where some of the subjects participate in a mini-public, whereas others do not, allows for a comparison between these two groups. By this design, we are able to measure whether and how taking part in a mini-public influences opinions, political trust, political efficacy, political knowledge, and readiness for political participation. Of course, real-world cases where some people participate and others do not can also be studied, but only in an experiment can we hold everything else between the two groups constant, and thereby conclude that observed differences between the groups are due to the experimental manipulation – that is, participation in the mini-public.

In addition to control, a clear advantage of experiments is that they can be repeated. An exact replication of research is not easily available through naturally occurring data. However, replication of research is crucial because it can increase the reliability of observed results. Usually the motivation for experimentation is to test a given theory. However, in the particular case of democratic innovations, the motivation for experimentation can be a study of innovations not yet used in real-world democratic systems. Via experiments we can study the potential consequences of certain ways to organise democratic participation, and on the basis of results decide whether these ways should be put to use to improve the quality of democracy.

An experiment typically has three characteristics: experimental treatment or manipulation, assignment of subjects into a treatment and a control group, and randomisation. The experimental treatment corresponds to the main research interest, e.g. the impact of participation in a mini-public. A treatment group consists of those who get certain treatment, whereas the control group consist of those not under the experimental treatment. The control group is an essential point of comparison to detect the impact of the treatment. It is also crucial that subjects are allocated randomly into the treatment and control groups because randomisation minimises the likelihood of systematic differences between the groups. It gives each participant an equal chance of ending up in either group. If participants of the treatment group were systematically different from the control group, say more volatile in their opinions, we could not say for sure whether observed opinion changes in the treatment group were based on this difference in the subject type or on the experimental manipulation.

In a simple experiment, there is one experimental group and one control group but quite often there is more than just one treatment group. For example, we could have an experiment where subjects are randomly assigned into two different types of mini-publics and a control group that does not take part in any mini-public. Sometimes experimental designs lack a natural control group and different treatments serve as one another's controls. For example, in our study on the influence of the decision-making method on participants' opinion formation in a

mini-public, secret ballot or a consensual statement, the treatments were considered one another's controls (Setälä et al, 2010).

In the social sciences, four different types of experiments can be separated. These types can roughly be characterised according to a scale from more control to less control, and from less external validity to more external validity. Laboratory experiments allow for most control but are often criticised for a lack of external validity. In the social sciences, this critique usually stems from the use of student subjects, and conditions that lack similarity with real-world circumstances. For this reason, there has been a call for more realistic field experiments conducted in subjects' natural environments. While field experiments bring about certain advantages, the drawback is that they cannot achieve the same degree of control with laboratory experiments. Lab-in-the-field experiments fall between pure field and pure laboratory experiments and they are based on an idea of increasing external validity, e.g. by bringing experiments into natural locations without losing much of the control of variables. While randomisation is commonly seen as essential to experimentation there are experiments that lack random assignment. In quasi-experiments, subjects are not randomised into treatment and control groups. There are two types of quasi-experiments depending on whether the experimenter has control over the assignment of subjects into treatment and control conditions. In naturally occurring experiments, the experimenter has no control over the assignment of experimental subjects into treatment and control groups but these groups are formed in a natural process. An example is a study where Brazilian municipalities that use or do not use participatory budgeting are compared (Touchton and Wampler, 2014). Naturally occurring experiments have the lowest level of control over variables but they provide real-world cases that resemble experiments. When the experimenter has control over the assignment in a quasi-experimental design, subjects are allocated into treatment and control conditions based on a certain characteristic.

Experimental research is usually motivated by willingness to test theories. Based on the theory, specific hypotheses are formulated and these are then tested with a specific experimental design. Theories are then challenged or given support depending on the results of the experiment. However, in the case of democratic innovation, the relationship between theory and empirical research is a bit atypical because theories related to democratic innovations, that is, theories on deliberative and participatory democracy, are not descriptive theories describing the state of affairs but rather normative characterisations of ideal states of affairs. These theories cannot therefore be rejected simply because empirical observations are not in line with the theories. However, while normative theories cannot be tested as such, they involve empirical claims that can be tested, and this is what experiments on democratic innovations engage in (Setälä and Herne, 2014).

2 OVERVIEW OF EXPERIMENTS IN THE STUDY OF DEMOCRATIC INNOVATIONS

The introduction of this Handbook by Elstub and Escobar (see Chapter 1), identifies four conceptual families of democratic innovations: mini-publics, participatory budgeting, collaborative governance, referenda and initiatives. They define democratic innovations as 'processes or institutions that have been developed to reimagine and deepen the role of citizens in governance processes by increasing opportunities for participation, deliberation and influence'. Thus, it is obvious that most of the democratic innovations implemented so far have had clear,

policy-oriented goals – popular participation is considered to have an intrinsic value and the people behind innovations have wanted to increase the role of citizens in democratic processes (see Section 4 in this Handbook). Pure experimental approaches to democratic innovations have been rare. In practice, almost all controlled experiments have been conducted with mini-publics. Nevertheless, some of the projects where democratic innovations have been implemented can be treated as natural experiments. For example, we can try to study the impact of participatory budgeting by comparing municipalities who use it with municipalities who do not. The problem with such an approach is that the possibility of holding all other variables constant (see above) is non-existing. The researcher cannot know for sure if the variation in the dependent variable, e.g. a policy decision on budget allocations or satisfaction with the policy output, is caused by the independent variable (participatory budgeting), or some other contextual factors relating to the political culture of the municipalities involved. The best way to isolate the effect of an experimental treatment is to use a controlled setting. Touchton and Wampler (2014) provide an example of a naturally occurring experiment on participatory budgeting. With an unusual dataset, the authors compare Brazilian municipalities that have adopted participatory budgeting to those that have not, and observe a clear influence with the use of the democratic innovation, e.g. increased health care spending and decreased infant mortality. While we believe that Touchton and Wampler's results are robust, it is in principle possible that they were not able to take into account all relevant control variables. A laboratory experiment could be used to complement their study to see, for example, which types of participatory budgeting processes are likely to work best.

Since controlled experiments provide the best tests of causality, we present in detail a mini-public that has exploited the method. We also present a controlled field experiment from Indonesia, in order to shed additional light on how different kinds of democratic innovations have been experimented with.

Most deliberative mini-publics have not been controlled experiments in a sense that there would have been random allocation to treatment and control conditions. The best-known family of mini-publics is Deliberative Polling®, developed and standardised by James Fishkin in the early 1990s (Fishkin, 2009). He designed it to address the problems of measuring public opinion through 'raw' opinion polls. The aim of deliberative polling is to provide a method of measuring enlightened and reflected public opinion, by allowing people to gain information and deliberate on a political issue in small-n groups. Deliberative Polls (DP) have become a 'gold standard' in organising deliberative mini-publics (Mansbridge, 2010). The DP is a quasi-controlled experiment in the sense that the participants are invited to a venue, they are randomly allocated into small groups at tables, and they deliberate based on their own values and beliefs. In a strict sense the DP, on the other hand, cannot be considered an experiment since it fails in providing different treatments or a treatment and a control condition.¹ Thus far, DPs have been conducted in countries all over the world, including non-democracies. Fishkin is a charismatic scholar with a great amount of energy and a strong will to implement this tool for measuring informed public opinion. The aim of the DPs is to recruit a representative sample of the population, to create a 'microcosm' of the demos, inform it through reading materials and expert panels, and then randomly allocate the participants into small groups, typically at round tables in a large room. The actual deliberations take place at these tables within the group each participant was allocated to. Participants are surveyed before and after the event, in order to see how their opinions on the issue at hand develops as a result of information and deliberation. DPs show a consistent pattern. Thanks to deliberation, people's opinions change

significantly, they also learn about the issue at hand. This is tested through factual quiz-type questions (e.g. Fishkin et al, 2014: 10–11).

Even though the DP is not a strictly controlled experiment, the design clearly allows for interpretations regarding the impact of information, deliberation and reflection. From an experimental perspective, the fact that DP is a registered trademark and a standardised tool, involves both strengths and weaknesses. The main strength is the ‘most different systems logic’, i.e. testing for exactly the same mechanisms in different political systems. If the DP works in a similar manner in democracies and autocracies, countries with majoritarian and proportional electoral systems, rich and poor countries, countries with a high level of social or political trust and countries where trust levels are low, and so on, we can conclude that the mechanisms of information and deliberation have an impact, universally speaking. This increases the external validity of the findings related to DP. The weakness, on the other hand, is the fact that the DP lacks different treatments and that the standardised method does not respond well to the requirement of scientific cumulativeness. Since the experimental logic with human subjects allows for very limited variation within each experiment, the learnings from each experiment should be taken into account in future experiments. As the setting is more or less constant at each DP, it is not possible to test additional independent variables.

Most experiments with mini-publics fall in the category of lab-in-the-field experiments. This means that the participants participate with their existing preferences, but the researchers manipulate the setting in some way, i.e. give treatments, and possibly placebo. What does a controlled lab-in-the-field experiment look like in practice? Our first deliberative mini-public dealt with energy policies, especially nuclear power (Grönlund et al, 2010; Setälä et al, 2010). It utilised two treatments. Carried out as a population-based experiment, the invitation to participate was sent out to a simple random sample of 2,500 adult citizens. The invitation was included in a survey, with questions measuring opinions on energy politics in general and nuclear power in particular. The survey also included other political questions and socio-economic background variables. In designing the experiment, we were inspired by the tension between consensual ideals of deliberative democracy and real-life politics where decisions often need to be made through voting. Thus, half of the 12 small-n groups voted on the issue of a sixth nuclear power plant through secret ballot (vote treatment), whereas in the remaining six groups the decision was made by formulating a statement which everyone could agree on (common statement treatment). The participants were randomly assigned to these small-n groups. In the vote treatment, members could vote ‘yes’, ‘no’ or cast an empty ballot. In the common statement treatment, a final statement on which all group members could agree was to be written.

This first survey was answered and returned by almost one fourth of the sample ($n = 592$). Of the respondents, 244 agreed to participate in the event. This second stage of the recruitment process was based on self-selection. At the third stage of the recruitment process, the number of participants was cut down so that the target sample of 144 people, i.e. 12 small-n groups consisting of 12 persons each, could be reached. When organising experiments where the subjects are people, it is important to understand that attrition always happens. Some experimental subjects will drop out, even when they have confirmed their participation. Thus, in this experiment we invited 194 of the 244 who had volunteered to take part in the citizen deliberation event. In order to guarantee a gender and age balance at the deliberative event, the 244 volunteers were first stratified by age and gender. Thereafter, the 194 people out of the 244 volunteers were chosen through random sampling within the strata. Finally, 135 participants of the 194 invited finally showed up, i.e. we missed the target by nine individuals. The attrition

rate was 31.5 per cent at this final stage. This happened even though the invited were promised a gift voucher worth 100 euro as compensation for their participation.

The deliberation day started with a knowledge quiz, including 10 questions on energy politics and five questions addressing general political knowledge. The reason for not including the political knowledge items in the mail-in survey is that, when measured through factual questions, political knowledge should be measured under controlled circumstances. Otherwise, the norm of social desirability associated with obtaining high scores, will make some people look up the answers to the factual questions, making the measure unreliable. After completing the knowledge quiz, the participants were asked to read a briefing material on nuclear energy. Thereafter, an expert panel was heard in a plenary session. The panel consisted of two experts supporting nuclear energy, and two experts opposing it. The actual deliberation took place in 12 small groups consisting of 10–13 members. The discussions were moderated by trained facilitators and the participants were asked to follow specific written deliberative rules that emphasised respect of other participants' viewpoints, the importance of justifying views, as well as self-reflection. The deliberations lasted for three hours, after which the groups were asked to make a decision on whether a sixth nuclear power plant should be built in Finland. After this, the sessions ended with a survey. Three months later, the participants received a follow up survey in order to trace how stable the effects on opinions were.

Our results show that treatment had no systematic impact on the development of energy opinions in the experiment. Moreover, we found that there were neither indications of group pressure among the participants in general, nor significant differences between the two treatments in this respect. Overall, the participants became more critical of the use of coal and peat in energy production, and somewhat more critical of nuclear power, whereas there was an increased willingness to save energy and an increased belief in other people's willingness to save energy (Setälä et al, 2010). When it comes to gains in energy-related knowledge, participants in the common statement treatment learned somewhat more than participants in the vote treatment. Lastly, with regards to political efficacy, the common statement treatment produced a higher external efficacy after deliberation, whereas no such change could be traced in the vote treatment (Grönlund et al, 2010). The reports written by the moderators show that the participants were both motivated and serious about their task to deliberate and reach a decision. All deliberations were also audio-recorded and transcribed. The transcriptions of group discussions indicate that, in accordance with our theoretical expectations, the deliberative process was more thorough in the common statement groups than in the secret ballot groups (Himmelroos and Christensen, 2012). Our conclusion reads that the common statement procedure, which resembles a consensual ideal, seems to produce more 'deliberative' outcomes than a treatment that uses secret ballots (Grönlund et al, 2010; Setälä et al, 2010).

Figure 34.1 presents visually the experiment described above. In terms of experimental design, it reads as follows. The simple random sample was first measured through a survey. Since the impact of different decision-making methods was the main research question of the experiment, we implemented two treatments: a vote treatment and a common state treatment. In empirical terms, thus, the independent variable was dichotomous, or in experimental terms, it consisted of one factor (decision-making method) with two levels (voting and common statement). The design with two treatments and a measurement of each individual before and after they were subject to treatment makes two kinds of comparisons possible. In this experiment, both between- and within-subjects comparison can be made. Between-subjects comparisons are possible because the experimental subjects were randomly assigned to distinct treatment

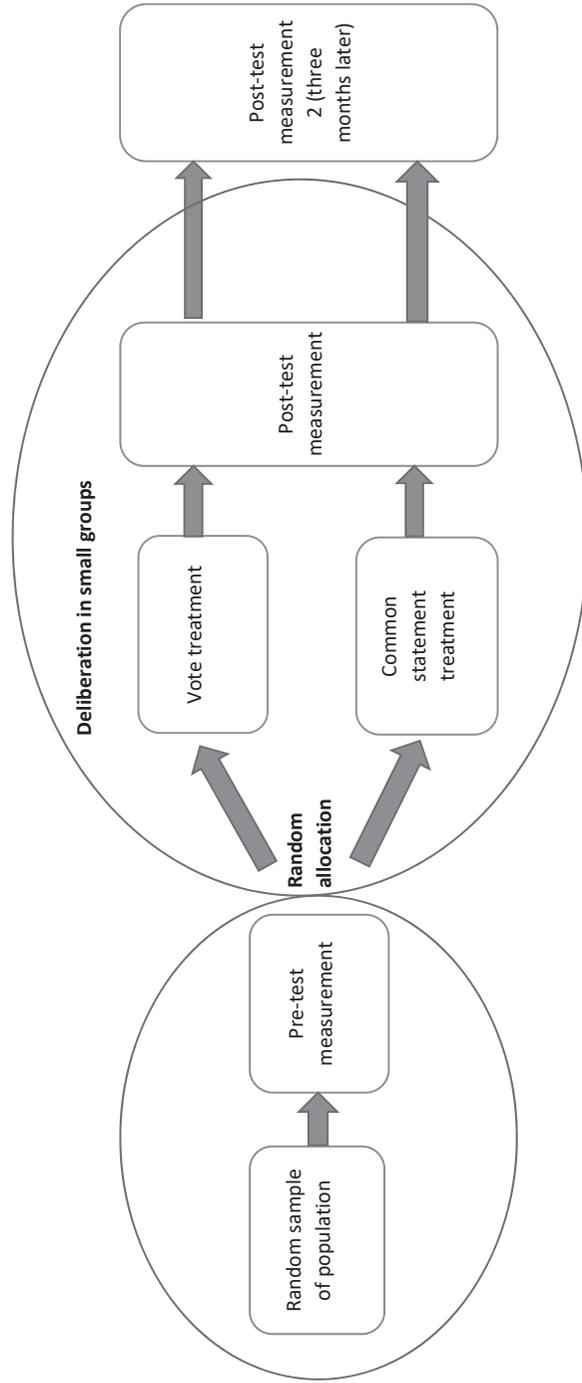


Figure 34.1 The design of the experiment comparing decision-making methods.

groups, and the within-subjects comparisons can be made thanks to the pre-test and post-test measuring of each participant (Druckman et al, 2011). Moreover, the post-test was replicated three months after the event as a check for the stability of the results.

Our research team has conducted several population-based experiments after the first one on energy policy. In fact, the experiment on energy policy was replicated online 1.5 years after the face-to-face experiment. The replication included exactly the same treatment, materials and even the same experts (we used a video from the original experiment). The only difference was modality. Participants in the online mode deliberated in online chat rooms with video and audio. Overall, the results of the face-to-face experiment were confirmed in the online mode (Grönlund et al, 2009; Strandberg and Grönlund, 2012). For a description of the two later lab-in-the-field experiments with mini-publics, see Grönlund 2016 (also Grönlund et al, 2015; Strandberg et al, 2017).

Controlled field experiments with democratic innovations are even more rare than controlled lab-in-the-field experiments. In Indonesia, however, a village-level comparison on the impact of decision-making methods was carried out in 2005–2006 (Olken, 2010). The experiment included 49 villages who chose projects to be put forward for funding in the Kecamatan Development Program. The 49 villages were chosen by the researcher so that they represented different parts of rural Indonesia (*ibid*: 244). They reached decisions in two distinct ways. In 32 villages, the decision on which projects to propose was made through representatives, whereas in 17 villages the choice was done through a plebiscite where everyone could vote. The two decision-making methods formed the treatments of the field experiment. In each village, two decisions were made. First, a general project and then a special women's project. Nothing else varied in the process between treatments. The process of setting the agenda, in which each hamlet (a sub-area) in the village nominated one general project and one women's project through a series of hamlet-level meetings, was identical in both experimental treatments. In order to make sure that he could isolate treatments effects from other effects caused by the social, political or economic context of each village (see above), Olken (2010: 247–249) checked that the randomisation resulted in a balanced set of villages in the two treatment conditions.

In the representative treatment, two village meetings were organised. At the first meeting, all representatives could discuss and decide on the general project to be put forward. At the other meeting, only women representatives met and chose a project. In the vote treatment, a similar procedure took place. All men in each village had one ballot with which they could vote for a preferred project, whereas all women had two ballots, they could vote for a general project, and for a women's project. Plebiscites resulted in a higher satisfaction among villagers and increased knowledge about the selected project. Villagers also became more willing to contribute to the selected project personally. When it comes to the nature of the projects selected, little happened. Regarding women's projects, however, the direct vote resulted in projects being located in poorer areas. The main conclusion is that the process matters. Being able to participate directly in decision making seems to increase both process satisfaction and output satisfaction (Olken, 2010).

3 CHALLENGES OF USING EXPERIMENTS

There are various challenges in organising experiments on democratic innovations. Perhaps the first and most difficult one is that organising experiments with human subjects is rather

expensive although there is a lot of variation in average and total costs, depending on the specific design of the experiment and on the number of participants needed. However, it is quite common that participants are compensated for their time spent in taking part in the experiment, and also that participants' travel and accommodation expenses are covered. Experiments on democratic innovations commonly involve surveys, often sent out to a large number of people, which also brings about costs. Setting aside the costs of experimentation, the most demanding challenges relate to the internal and external validity of experiments. As argued above, different types of experiments can be particularly vulnerable to either internal or external validity problems. In our understanding, internal validity refers to the ability of the experiment to test the causal relationship it is designed to test, whereas external validity refers to the potential to generalise the findings of the experiment. In the following, we will consider internal and external validity separately although they are not necessarily always easy to distinguish.

Shortcomings that relate to internal validity can be characterised by the so-called Duhem–Quine thesis (Duhem, 1906 [1954], Quine, 1951). The thesis concerns the acceptance or rejection of experimental hypothesis based on the observed results. Duhem and Quine argue that the decision over acceptance or rejection always involves some degree of interpretation because we cannot test a single hypothesis in isolation (Bardsley et al, 2010: 96). In other words, it is impossible to perfectly isolate the influence of a single variable and hold the other $n-1$ variables constant. Since the control of all variables is never perfect, what we actually test is a set of hypotheses consisting of the target hypothesis, the actual research hypothesis, and a set of auxiliary hypothesis. If observed results contrast with the research hypothesis, the experimenter has to decide whether the research hypothesis should be rejected or whether one of the auxiliary hypothesis does not hold. The auxiliary hypotheses consist of assumptions that guarantee that other than the target variable are actually held constant. Commonly they concern subjects, instructions or subjects' incentives. For example, the experimenter assumes that the subjects understand and follow the instructions of the experiment, that they take their tasks in the experiment seriously, that they report their opinions truthfully, that the conditions of the location, e.g. heat, coldness or noise, do not influence subject behaviour, etc.

One of the auxiliary assumptions that is highly crucial from the point of internal validity, is that subjects are assumed not to act in order to please the experimenter, or other participants, in a way that would influence their behaviour. The so-called Hawthorne effect refers to the fact that being observed can influence the behaviour of experimental subjects. A special case of the Hawthorne effect is the so-called experimenter demand effect, which occurs if subjects behave in a manner they think the experimenter expects them to behave. This can happen either consciously or sub-consciously. Schultz (1969, 221; cited in Levitt and List, 2007a) argues that the lab has a 'superior–subordinate' relationship matched only by that of 'parent and child, physician and patient, or drill sergeant and trainee'. Typical cases of experimenter demand effect are acting in a way that is perceived socially acceptable or trying to show intelligence.

Another issue of internal validity that seems highly relevant in experiments on democratic innovations is subjects' incentives. Are the incentives big enough to motivate subjects into behaviour that would correspond to their behaviour in the real world? The difference between decision making in experiments and in real-world democratic institutions is that experiments seldom concern real political decisions that would influence people's lives long into the future. It is therefore possible that subjects' behaviour in discussions and their responses to surveys are different in experiments compared to situations where deliberation is actually expected to influence collectively binding decisions.

In addition to challenges that relate to internal validity, experiments are often criticised for a lack of external validity. In behavioural economics, there is a large discussion on the use of students as subjects in experiments, and whether student subjects behave similarly with subjects drawn from the population at large, or subjects drawn from a group of experts. Indeed, there are plenty of experiments comparing student behaviour to the behaviour of other subject groups, and it seems that whether there is a difference depends largely on the specific questions studied (e.g. Dyer et al, 1989; Haigh and List, 2005; Depositario et al, 2009). Experiments on democratic innovations do not typically use student subjects but rather invite participants from a random sample or open calls in the media. However, despite random sampling there can be features related to the participants that undermine external validity. Namely, it seems that mini-public experiments are likely to be vulnerable to certain self-selection biases. Even though invitations to take part in a mini-public experiment are based on a random sample we cannot force anyone to participate and therefore cannot guarantee that no biases exist in the group of subjects that we study. In effect, there is evidence that participants in deliberation experiments tend to be more politically interested and have higher political knowledge than people in general (e.g. Fishkin et al, 2005; see also Karpowitz and Mendelberg, 2011: 268).

A further problem of external validity relevant in particular to controlled laboratory experiments is that subjects are asked to act in circumstances that are not familiar to them. Levitt and List (2007b, 348) describe the situation of a subject, who they call 'Jane', taking part in a decision-making experiment: 'Jane realises that she is entering a relationship that has no useful parallels in her everyday life.' Levitt and List are afraid that experimental subjects might use peculiar kinds of decision-making heuristics when they are in a strange situation, and can therefore end up behaving in a way that would not be typical to them in the real world.

These kinds of problem are less pronounced when we move from the laboratory to lab-in-the-field or field experiments. However, it cannot entirely be ruled out that similar issues could be of relevance to experiments in democratic innovations as well. For example, in a deliberation experiment on the topic of immigration, people with extreme anti-immigration attitudes were less likely to participate than more moderate people (Karjalainen and Rapeli, 2015).

The key to avoid as many problems as possible is naturally careful experimental design, and if possible, pilot experiments that help to reveal those parts of the design that are not functioning properly. By careful design, we can have control over the auxiliary assumptions and actually succeed in testing the target hypothesis we aim to test. Curiously, an effective way to tackle problems of experimentation is to run more experiments. Usually most of the auxiliary hypotheses can be tested. The example of testing the experimenter demand effect was based on our deliberation experiment on the topic of immigration (Grönlund et al, 2015). In the experiment, we observed no polarisation of opinions, but there is a possibility that participants gave more tolerant answers than they would without the need to please the experimenter. In other words, they gave socially acceptable answers. One way of exploring whether this actually was the case would be to run a similar experiment but with a topic that is not likely to give impetus to socially acceptable answers, and see whether similar results on polarisation would be observed.

Giving participants proper compensation is likely to increase their concentration on their tasks in the experiment as well as help to some extent alleviate the self-selection bias. External validity can also be increased by replicating the test of the hypothesis perhaps with some modifications, or in different cultural contexts. Further, comparing behaviour in the laboratory and behaviour in the field is also a way to increase external validity. For example, experiments on democratic innovations can be combined with case studies on real-world democratic innovations.

4 CONCLUSIONS

In this chapter, we provided two examples of experiments on democratic innovations, the Finnish lab-in-the-field experiment on nuclear power and a field experiment from Indonesia. The first was designed as an experiment in democratic innovations and was not a part of a political decision-making process, whereas the latter had actual political consequences. We described the processes of these experiments as well as their basic results. The outcome of both experiments suggests that the exact processes of organising democratic innovations matter. Experiments need planning and a careful implementation. After describing these cases, we discussed certain problems related to experimental research. We see the experimental demand effect and self-selection bias as especially relevant for experiments on democratic innovations. We also discussed potential means to alleviate these types of problems. In some cases, the cure to problems in experimentation is to run more experiments in different settings. This can help identifying exact problems, and to know which factors influence them. A good way to increase external validity is to run both laboratory and field experiments in order to see whether they produce similar (or different) results.

Despite the problems related to the experimental method, we consider experimentation to be a powerful method to evaluate democratic innovations. Experimental research can test whether democratic innovations, in fact, have those types of positive consequences as they are claimed to have – whether opinions change and depolarise, whether trust and political efficacy increase, whether mutual understanding is achieved, to name a few. Furthermore, by conducting experiments we can induce what are the best ways to organise democratic innovations. Our example showed a comparison of the consequences of using secret ballot or a consensual procedure in order to reach a group level verdict.

Above we have focused almost only on one form of democratic innovations, namely mini-publics. As said above, the reason for this focus is that few experiments exist of other types of democratic innovations. While there are still many things that we can study about mini-publics, e.g. how they function in different cultural contexts, it also seems that we should look at other types of democratic innovations. In many cases, these are already used in real-world political decision making, and many of them cannot be tested without a direct connection to actual decision making, e.g. participatory budgeting. Therefore, there seems to be a need for field experiments or research on naturally occurring experiments, asking what consequences these other types of democratic innovations have and how they operate in practice.

REFERENCES

- Ansolabehere, I. and S. Valentino (1994), 'Does attack advertising demobilize the electorate?', *American Political Science Review*, **88** (4), 829–838.
- Bardsley, N., R. Cubitt, G. Loomes, P. Moffat, C. Starmer and R. Sugden (2010), *Experimental Economics: Rethinking the Rules*, Princeton, NJ: Princeton University Press.
- Depositario, D.P.T., R.M. Nayga Jr, X. Wu and T.P. Laude (2009), 'Should students be used as subjects in experimental auctions?', *Economics Letters*, **102** (2), 122–124.
- Druckman, J.N., D.P. Green, J.H. Kuklinski and A. Lupia (2011), 'Experiments: An introduction to core concepts', in J.N. Druckman, D.P. Green, J.H. Kuklinski, and A. Lupia (eds), *Cambridge Handbook of Experimental Political Science*, New York NY: Cambridge University Press, pp. 15–26.
- Duhem, P. (1906), *The Aim and Structure of Physical Theory*, trans. P.P. Wiener (1954), Princeton, NJ: Princeton University Press.

- Dyer, D., J.H. Kagel and D. Levin (1989), 'A comparison of naive and experienced bidders in common value offer auctions: A laboratory analysis', *The Economic Journal*, **99** (394), 108–115.
- Fishkin, J.S. (2009), *When the People Speak: Deliberative Democracy and Public Consultation*, Oxford: Oxford University Press.
- Fishkin, J.S. and R.C. Luskin (2005), 'Experimenting with a democratic ideal: Deliberative polling and public opinion', *Acta Politica*, **40** (3), 284–298.
- Fishkin, J.S., R.C. Luskin and A. Siu (2014), 'Europolis and the public sphere: Empirical explorations of a counterfactual idea', *European Union Politics*, **15** (3), 328–351.
- Gosnell, H.F. (1926), 'An experiment in the stimulation of voting', *American Political Science Review*, **20** (4), 869–874.
- Grönlund, K. (2016), 'Designing mini-publics for constitutional deliberative democracy', in M. Reuchamps and J. Suiter (eds), *Constitutional Deliberative Democracy in Europe*, Colchester: ECPR Press, pp. 109–128.
- Grönlund, K., K. Herne and M. Setälä (2015), 'Does enclave deliberation polarize opinions?', *Political Behavior*, **37** (4), 995–1020.
- Grönlund, K., M. Setälä and K. Herne (2010), 'Deliberation and civic virtue: Lessons from a citizen deliberation experiment', *European Political Science Review* **2** (1), 95–117.
- Grönlund, K., K. Strandberg and S. Himmelroos (2009), 'The challenge of deliberative democracy online – A comparison of face-to-face and virtual experiments in citizen deliberation', *Information Polity* **14**, 187–201.
- Haigh, M.S. and J.A. List (2005), 'Do professional traders exhibit myopic loss aversion? An experimental analysis', *The Journal of Finance*, **60** (1), 523–534.
- Himmelroos, S. and H.S. Christensen (2012), 'Finding common ground – an empirical analysis of deliberative processes and opinion transformation', in 62nd Political Studies Association Annual International Conference in Belfast, Northern Ireland, April, pp. 3–5.
- Karjalainen, M. and L. Rapeli (2015), 'Who will not deliberate? Attrition in a Multi-Stage Citizen Deliberation Experiment', *Quality & Quantity*, **49** (1), 407–422.
- Karpowitz, C.F. and T. Mendelberg (2011), 'An experimental approach to citizen deliberation', in J.N. Druckman, D.P. Green, J.H. Kuklinski and A. Lupia (eds), *Cambridge Handbook of Experimental Political Science*, Cambridge: Cambridge University Press, pp. 258–272.
- Levitt, S.D. and J.A. List (2007a), 'What do laboratory experiments measuring social preferences reveal about the real world?', *The Journal of Economic Perspectives*, **21** (2), 153–174.
- Levitt, S.D. and J.A. List (2007b), 'Viewpoint: On the generalizability of lab behaviour to the field', *Canadian Journal of Economics* **40** (2), 347–370.
- Mansbridge, J. (2010), 'The deliberative system disaggregated', paper presented at the Annual Meeting of the American Political Science Association, Washington, DC, September.
- Morton, R.B. and K.C. Williams (2010), *Experimental Political Science and the Study of Causality. From Nature to the Lab*, Cambridge, NY: Cambridge University Press.
- Olken, B.A. (2010), 'Direct democracy and local public goods: Evidence from a field experiment in Indonesia', *American Political Science Review* **104** (2), 243–267.
- Plott, C.R. and M.E. Levine (1978), 'A model of agenda influence on committee decisions', *American Economic Review* **68** (1), 146–160.
- Quine, W. Van Orman (1951), 'Two dogmas of empiricism', *Philosophical Review* **60** (1), 20–43.
- Setälä, M. and K. Herne (2014), 'Normative theory and experimental research in the study of deliberative mini-publics', in K. Grönlund, A. Bächtiger and M. Setälä (eds), *Deliberative Mini-Publics. Involving Citizens in the Democratic Process*. Colchester: ECPR Press, pp. 59–76.
- Setälä, M., K. Grönlund and K. Herne (2010), 'Citizen deliberation on nuclear power: A comparison of two decision-making methods', *Political Studies* **58** (4), 688–714.
- Schultz, D.P. (1969), 'The Human Subject in Psychological Research', *Psychological Bulletin* **72** (3), 214–228.
- Strandberg, K. and K. Grönlund (2012), 'On-line Deliberation and Its Outcome – Evidence from the Virtual Polity Experiment', *Journal of Information Technology & Politics* **9** (2), 167–184.
- Strandberg, K., S. Himmelroos and K. Grönlund (2017), 'Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization', *International Political Science Review* (forthcoming).
- Touchton, M. and B. Wampler (2014), 'Improving social well-being through new democratic institutions', *Comparative Political Studies* **47** (10), 1442–1469.

NOTE

1. Some DPs have had a control group.