ACKNOWLEDGMENTS

Our sincere thanks to the authors for agreeing to contribute and for their cooperation throughout. This Handbook is their work, and without their timely and professional collaboration, it would have been an impossible task. Thanks also to Dominic Byatt, Commissioning Editor at Oxford University Press, for so wholeheartedly supporting the project. Our respective institutions, Dublin City University, the University of Bamberg, and the University of California, San Diego, hosted a number of editorial meetings, and we are grateful for the many ways in which they have supported our efforts. In addition to our home institutions, we thank the London School of Economics for hosting several editorial meetings at critical stages of the project. Martin gratefully acknowledges support for work on the Handbook from the Norwegian Research Council (Fritsam Project No. 222442) and from the School of Law and Government and the Vice-President for Research, Dublin City University. We thank Matthias Bahr, Eric Bientzle, Wolfgang Goldbach, Margret Hornstein, Erik Jäger, David Ruß, and Akinato Suzuki for help in preparing the manuscript.
CHAPTER 9
THE EXPERIMENTAL STUDY OF LEGISLATIVE BEHAVIOUR

JAMES N. DRUCKMAN, THOMAS J. LEEPER, AND KEVIN J. MULLINIX

9.1 Introduction

A. Lawrence Lowell was one of the most influential political scientists of the twentieth century, serving as president of both Harvard University and the American Political Science Association (APSA). In Lowell's presidential address to APSA, he advised against following the model of the natural sciences: "We are limited by the impossibility of experiment. Politics is an observational, not an experimental science. . ." (Lowell 1910, 7). Counter to this sentiment, experiments have become prominent, if not a central, method of inquiry in political science over the last quarter century (e.g., Druckman and Lupia 2012), just how much have they influenced the area on which much of Lowell's work focused—legislative studies? In this essay, we address this question. We begin by clarifying what we mean by an "experiment" and we discuss critical aspects of variation in experimental approaches. We then review three areas of experimental applications in legislative studies: legislative voting, parliamentary coalitions, and responsiveness. We conclude by discussing the challenges and limitations of experiments on legislatures but also by emphasizing future possibilities.

9.2 What is an Experiment and How is it Used?

In contrast to modes of research that address descriptive or interpretive questions, researchers design experiments to address causal questions. A causal question involves a comparison between two (or more) states of the world: one in which some stimulus is experienced and another in which it is not, all else constant (an untreated state). Causation is seen in the difference between these states and the fundamental problem of causal inference arises when we cannot simultaneously observe a person or entity in those distinct states (Holland 1986). Consider, for example, the causal effect of having majority versus unanimity rule. Only one of these rules can be in place at any given point in time. How can we isolate the impact of distinct rules on legislators' behaviours? To do so perfectly would require exploring a single vote, at a single point in time, with the exact same set of legislators—in one case where a majority is needed and another that requires unanimity. This is obviously impossible and thus social scientists often take one of two approaches to explore the impact of such rules (and other factors): observational research and experiments.

Observational research involves comparisons between people subjected to different treatments. In the example referenced above, it might mean analysing two similar legislatures with similar ideological compositions, identifying a similar piece of legislation, but legislatures that differ in terms of the voting rule. If the unanimity rule leads legislators to vote more strategically, for example, one might conclude the unanimity rule has a causal effect (on strategic voting) relative to majority rule. Of course the validity of this conclusion would depend on the extent to which the two legislatures and votes are truly comparable. It might be that legislatures that establish unanimity procedures in the first place already have individuals in place who are likely to vote strategically, thus any observed relationship is spurious (due to endogeneity).

As an example, a goal of much work on legislatures is to understand how institutional arrangements influence behaviours. Yet legislators themselves (whose behaviours are of interest) often design and alter institutions (see Riker 1962). Outcomes are thus endogenous to institutional arrangements. Another even more concrete example comes from Druckman and Thies (2002) finding that in bicameral regimes, when a governing coalition lacks an upper chamber majority, its stability may be reduced due to inter-chamber conflicts. Yet governments can anticipate the life shortening effects of bicameralism when forming in the first place, and they thus often form coalitions less likely to suffer from the bicameral dynamic. The type of coalition that forms impacts duration, but expected duration affects the type of coalition that forms. The inherent endogeneity makes identifying precise causal mechanisms challenging. This is especially true given that legislators' behaviour is dynamic, meaning that it occurs over time and a host of variables can influence it (Druckman 2008). These factors have stimulated a renewed interest in what the most direct way to test definitively a causal relationship: experiments.

Experimental research differs from observational research in that the entities under study are experimentally assigned to treatment. Here, treatments refer to distinct values of potentially causal (i.e., independent) variables. For example, an experimenter might randomly assign some participants to act as legislators and vote on prepositions under majority rule (one treatment) and others to do the same but under unanimity rule (a second treatment). In some designs, there also is a control group that does not receive a treatment (e.g., they vote but are not told how the outcome will be determined) and/or
multiple treatment groups (e.g. various voting rules are used such as a super-majority rule). Random assignment means that each entity (e.g. individual) being studied has a non-zero (and independent) chance of being in a particular treatment or control condition.

How does random assignment overcome the fundamental problem of causal inference? By definition, random assignment eliminates concerns about endogeneity and ensures, on average (i.e. in large enough samples), that the groups are the same in every respect (e.g. same number of "naturally" strategic and non-strategic individuals, the same average age, and so on for every possible variable). Consequently, differences in average behaviours across the two or more experimental groups (e.g. more strategic voting in the unanimity rule case) indicates confidence that there is an impact of voting rule. Although we cannot observe a given individual in both treated states (voting under majority and unanimity rule), random assignment enables the researcher to estimate the average difference between groups of individuals in the two states of the world—the average treatment effect. Prior to the intervention, the randomly assigned treatment groups (and control group) have the same expected behaviours. Apart from chance variation, random assignment provides a basis for assuming that the one treatment group behaves as the other would have behaved had it not received the given treatment. Because the experimenter controls values of the independent variable and assigns groups randomly, all confounding variables (including those not actually observed) can be ruled out in establishing the effect of the independent variable.

By offering clear causal evidence, experiments allow social scientists to address a variety of issues. Roth (1995, 22) identifies three non-exclusive roles of experiments, and a cursory review makes clear that political scientists employ them in all three ways. First, Roth describes "searching for facts," where the goal is to "isolate the cause of a non-observed regularity, by varying details of the way the experiments were conducted." These types of experiments often complement observational research. "Searching for facts" describes many experimental studies that attempt to estimate the magnitudes of causal parameters, such as the influence of minority government on cabinet stability. A second role entails "speaking to theorists," where the goal is "test the predictions [or the assumptions] of well-articulated formal theories." The third usage is "whispering in the ears of princes," which facilitates dialogue between scientists and policymakers.

It is important that we clarify four important points about experimental designs. First, one should not confuse random assignment with random sampling. Random sampling refers to a procedure by which participants are selected for inclusion in some studies, while Random assignment, by contrast, does not require that participants be drawn randomly from some large population (as with random sampling). Because the bulk of experimental studies do not include actual legislators, the question of whether the participants (e.g. college students) are comparable to actual legislators is a significant one, and one which we discuss later.

Second, many social science experiments use a between-subjects design, in which the researcher randomly assigns participants to distinct treatment groups (e.g. majority or unanimity rule). An alternative approach is a within-subjects design (in which a given participant is observed before and after receiving a treatment (e.g. there is no random assignment between subjects). For example, the same subjects vote under majority rule and then they do it again under unanimity rule (or vice versa). In principle, the within-subjects design works only (one can assume that each participant has not changed in any way relevant to the study (e.g. participation in a prior round of voting does not affect subsequent behaviour). Because social scientists have to contend with problems of memory and anticipation, a between-subjects experimental design is more typical.

That said, there are important exceptions—by definition, an experiment occurs when the experimenter intervenes (e.g. manipulates the independent variable(s) of interest) and can safely treat the units as on average comparable. Sometimes this comparability can be assumed through the employment of induced value theory. Induced value theory refers to gaining control over experimental participants' preferences via monetary rewards. There are four conditions necessary for the implementation of induced value non-satisfaction, saliency, dominance, and privacy (Guala 2005, 232–33). In nearly all cases, the payoff is some kind of financial reward. Many of the studies we discuss below employ induced value theory; this does not mean random assignment does not occur but in, theory, induced value dominates subjects preferences to such an extent that subjects are effectively homogeneous and thus the study design can meet the requirement of being an experiment sans random assignment.

One of the early experimental studies of coalitions took this approach. Specifically, Riker (1962) examined how three-person groups formed coalitions given pre-specified pay-off constraints. In this treatment-only experiment (i.e. there was no control group), subjects negotiated how to divide a pot of money among themselves within the confines of a payoff schedule (the independent variable) that offered subjects nothing for a three-person coalition and varying amounts for each other in a two-person coalition. Riker treated each run of the experiment as independent.

A third point worth noting is the distinct traditions shaping research by political economists and political psychologists. Whereas psychological experiments often include some form of deception, economists consider it unacceptable. Psychologists also rarely use subjects for specific actions they undertake during an experiment (i.e. they do not employ induced value), Economists, on the other hand, often require such payments.

Finally, there are three general contexts in which experiments take place: the laboratory, the field, and within large-scale surveys. A lab experiment involves an intervention in a setting (e.g. a research table) and controlled by the researcher; a field experiment takes place in a natural setting (e.g. society or the workplace); and a survey experiment involves an intervention in the course of an opinion survey (which might be conducted in person, over the phone, or via the web). In any of these cases the key aspect is whether the design is used to make inferences about the relationship between the treatment and outcome. If so, the design is experimental. If not, it is observational.
9.3 EXPERIMENTS ON VOTING IN LEGISLATURES

One of the most fundamental objects of inquiry in legislatures is voting. The act of voting that determines what governing coalitions survive, what laws are made, and ultimately who wins in politics. An overriding question here is given legislators' preferences, the dimensions of policy, and institutions, can one discover stable political outcomes forming in legislatures? This is the classic social choice problem and one that has been the subject of experiments for over 40 years. Experiments allow for control over these various factors and enable researchers to control induced value theory to control for preferences. In nearly all cases, this work tests theories of social choice.

The seminal social choice work, Arrow's (1951) "general possibility theorem," questioned whether it was possible for any reasonable democratic institution to aggregate collections of individual voters' preferences into a stable, majority-supported policy. One implication of Arrow's work was that any legislative coalition could be defeated by a differently configured majority coalition. McKelvey (1976) showed that in multidimensional policy space, the core is equivalent to the median voter's position. In multidimensional space, the core is the analogous "middle" of the voters' preferences depending on the configuration of preferences such a core does not always exist (e.g., a median along multiple dimensions would be a core).

Differences between configurations of preferences therefore tested the limits of majority rule across plausible conditions. The key experimental comparison in Fiorina and Plot's studies was between the two conditions—the one (with a core) where majority rule should have consistently resulted in policy outcomes near the core (and hence stability) and the second no-core condition, which tested "what would happen when anything could happen" (Miller 2011, 115). Despite McKelvey's (1976) theoretical prediction that endless cycling might occur, even without a core, Fiorina and Plot's experiments largely yielded core-like outcomes and therefore did not "explode" (Fiorina and Plot 1978, 150).

Why did the experiments not produce the infinite cycling predicted by theory? Answering this question is an example of how experiments can be used to "speak to theory". One answer is that the experimental procedures imposed some institutional particularity, which constrained behavior. Fiorina and Plot's experiments relied on the institutional rules in which those preferences operated or the outcomes of legislative voting.

Experiments on majority rule institutions began almost as soon as McKelvey's work was published (see McKelvey and Ordeshook 1990). Specifically, Fiorina and Plot (1978) examined the process of simple majority rule institutions on collective decision-making. They used an experimental design consisting of subjects tasked with choosing a policy from a non-dimensional space, with each subject having clearly defined (but private) preferences. Essential features, essentially unconstrained communication among subjects, and the group's preferences were individually observed, and the group's preferences were observed by the experimenter. The process was designed to test the extent of majority voting under a forward agenda rule. A forward agenda rule is one where possible policies are introduced and voted on immediately against the status quo, and this continues until no new proposal wins.

Fiorina and Plot's experimental approach resolved many of the shortcomings of observational methods. First, preferences and institutional rules were manipulated so that researchers were able to avoid the challenging task of observing and quantifying these critical independent variables. Second, by conducting an experiment in which decision-making was not confounded by unobservable factors of legislative processes (e.g., symbolic votes, log-rolling, party or constituency pressure, lobbying, etc.), Fiorina and Plot were able to use repeated iterations of the experiment, manipulating only preferences initially to test condition there was a majority rule equilibrium, i.e., a "core"—and test to which there was not, to clearly identify which proposals from the policy space were supported during the mock legislative processes. Formal theories had expected that majority rule institutions could only produce stable (or even remotely predictable) outcomes when a core existed and would otherwise produce chaos. In a non-dimensional policy space, the core is equivalent to the median voter's position. In multidimensional space, however, this core is the analogous "middle" of the voters' preferences, depending on the configuration of preferences such a core does not always exist (e.g., a median along multiple dimensions would be a core).

Because most policies can be discussed on multiple dimensions (e.g., health care concerns economic and social dimensions; see Riker 1986). If stable majority coalitions cannot be formed, how can governance—let alone representation—occur? Questions about majority rule instability remain but have been greatly aided by experimental investigation. Indeed, while Miller (2011) describes how early empirical efforts (e.g., Riker 1986) attempted to observe linkages between majority rule institutions, the preferences of legislators, and policy outcomes, such approaches were "seriously handicapped due to the inability to measure legislators' preferences or to a large extent". The institutional rules in which those preferences operated or the outcomes of legislative voting are the main focus of the experiment.
collective decision to any particular part of the policy space. A backward agenda, by contrast, involves voting on proposals in the opposite order from which they were introduced, eventually pitting the last successful proposal—if any proposals were successful—against the original status quo. This rule could have more predictable outcomes if the alternatives available to be voted on maximize their utility and if no proposal finds stable support of a majority coalition than the status quo wins. Wilson's (1966) experiments on forward and backward agendas (for preference configurations with a core) show that agenda rules matter: holding all else constant, a backward agenda significantly constrains instability (generally by leaving policy at the status quo) and a forward agenda offers no such guarantee of constraint. If policies are decided under Agenda rules, then, may not be the best explanation for the apparent stability in Fiorina and Plott's experiments (since agenda-setting rules are not the apparent reason for the Fiorina and Plott stability finding).

Another answer to why McKelvey's expected chaos did not occur might lie in the nature of preferences. Experiments conducted after Fiorina and Plott showed that a lack of chaos was not an artifact of their experimental design, being robust to having ideal points publicly known by all subjects, completely unconstrained discussion and no formal agenda rule (McKelvey et al. 1978). McKelvey (1986) himself offered one and no formal agenda rule (McKelvey et al. 1978). McKelvey's (1986) himself offered one explanation for the absence of chaos. He theorized that outcomes of majority voting could be "institution-free" to the extent that an "uncovered set" (an area of policy space equivalent to the core if a core exists and otherwise consisting of all alternatives that can directly or indirectly defeat all other alternatives in head-to-head voting) seems, theoretically, to encompass the set of likely outcomes across a number of different institutional arrangements. Experiments convincingly demonstrated that even without a core, outcomes consistently fall inside the uncovered set. In a "search for facts" to settle outcomes, McKelvey et al. (2008) re-analysed 322 iterations of previously theoretical inconsistencies, Bianco et al. (2008) ran majority rule voting games to show that 91.75 percent of the time, majority rule produces decisions inside the uncovered set. Thus, the original authors of these studies also concluded that their experiments demonstrated instability, but reanalysis shows that the thought their experiments demonstrated predictable outcomes from within the uncovered set. The experiments were not only indirect or indirectly dominated most other alternatives. Experiments were a vital for demonstrating the stability of social choice, even when the original experimentalists had insufficient theory to explain the empirical results.

Fiorina and Plott had wondered why trials from their no-core condition produced a fairly narrow, centrally located set of outcomes; the uncovered set appears to explain those outcomes. While theory was able to predict the effects of institutional factors (e.g., forward versus backward agenda rules) on social choice with experiments used to test those predictions, these experiments also revealed an overlooked feature of majority rule: while seemingly anything could happen, preferences themselves (even when not explicitly constrained) heavily constrain the feasible coalitions and, as a result, the set of possible collective decisions. Experiments have revealed other important insights into how institutions shape decision making. For example, subtle variations in discussion rules across studies (e.g., open discussion, no discussion, issue-by-issue discussion) have significant impacts on collective decisions. McKelvey and Ordeshook (1984), for example, compare experimental results from issue-by-issue voting (i.e., proposals can only move on one dimension at a time) under conditions of either no discussion or unrestricted discussion. Even though issue-by-issue voting constrains each proposal to a unidimensional move, discussion allowed the groups to behave more like they were operating under unrestricted majority rule than strict issue-by-issue voting. These results reiterate the importance of experimental tests of formal theories but also have implications that can "speak to princes." First, voting experiments show how coalitions form and defeat counterproposals in ways not previously predicted by theory. Compromises can be reached by legislators, not by stable majorities "bulldozing" legislation but through carefully crafted compromises that hold onto pivotal coalition members in the face of counterproposals (e.g., Jeong et al. 2011). Second, "princes" might learn that while restricted institutional rules may produce more predictable outcomes (policies in the uncovered set), those outcomes do not necessarily correspond to the outcomes produced by pure majority rule. Depending on a legislature's policy objectives, experiments suggest that relatively simple procedural rules (as opposed to significant rule changes like supermajority voting) can greatly constrain outcomes of majority rule decision-making. Experiments have confirmed how "institution-free" majority rule produces far more predictable coalitions and fairly stable outcomes than might be expected by prior theory.

9.4 Experiments on Parliamentary Coalitions

The essence of parliamentary government is coalition formation and functioning. In contrast to presidential regimes, the leader of the nation and thus the policies implemented are determined by election, but rather by coalition negotiations. Coalition formation has been a subject of intense study over the past 40 years (e.g., Goldner et al. 2001). Most of these studies concentrate on making predictions about what government will emerge after coalition negotiations, how portfolios are allocated, and how long a coalition will last. The empirical challenge is that many variables affect each of these dynamics (e.g., Martin and Stevenson, 2000). Experiments can address this challenge as they allow for control of confounds. Moreover endogeneity is always present with observational data given parties anticipate potential institutional effects and act accordingly. Thus, it is truly isolate the impact of institutions, an approach where one controls the observable presence of an institution is needed.

As such, experiments have been particularly useful for understanding how legislatures bargain over potential coalitions. Due to the necessity of control over institutional factors, almost all of this experimental work is conducted in a highly structured laboratory setting (as opposed to field or survey settings). Much of it involves testing the
disparity between theory and evidence by suggesting that a faithful implementation of the Baron-Ferejohn model in an experimental laboratory is challenging due to finite time constraints, the model's assumption of stationarity (requiring that a member take the same action in structurally equivalent subgames), and because the unique stationary equilibrium involves randomization. The sharp distinction between model expectations and McKelvey's experimental results spurred a series of experimental studies that attempted to explain the differences and overcome experimental limitations.

One body of experimental research addressed the Baron-Ferejohn model's institutional predictions. Still within the framework of a distributive model of legislative bargaining, Frerichs et al. (2003) analyze the effects of open (allow amendments) versus closed (no amendments, vote immediately) amendment rules. Consistent with the Baron-Ferejohn model, Frerichs et al. find a more egalitarian distribution of benefits and longer delays under the open amendment rule. They also find that the proposer gets a larger share of the benefits than other coalition members under both rules, and that play converges toward a minimal winning coalition (Riker 1962) under the closed amendment rule. In contrast to the Baron-Ferejohn model, Frerichs et al. demonstrate that the frequency of minimal winning coalitions is much greater under the closed rule (the model predicts minimal winning coalitions under both rules), and that the distribution of benefits is much more egalitarian than expected.

In an effort to address some of the methodological limitations of McKelvey (1991), Diermeier and Morton (2005) conduct an experiment to more directly test the Baron-Ferejohn predictions. They concentrate on a finite period using majority rule in which three players divide a fixed payoff under closed rule procedures. In contrast to McKelvey's experiment, Diermeier and Morton's design does not have to assume stationarity; the subgame perfect equilibrium does not involve randomization, and by employing a weighted majority game they are able to isolate comparative statics rather than simple point predictions. In doing so, Diermeier and Morton find that the Baron-Ferejohn model predicts "hardly better than a coin flip which coalition partner is selected by the chosen proposer" (p. 203). In contrast to model predictions, proposers do distribute much more money to other players than expected. Thus, proposers often distribute money to all players, the cheapest coalition partners are not always selected, the proposers offer too much, and many first-period proposals above the continuation value are rejected. In fact, Diermeier and Morton find that a simple equal sharing rule is the prediction that can account for half to three quarters of the accepted proposals.

The Baron-Ferejohn model is not without rivals—which is not surprising given its limited reception in experimental tests. In fact, a plethora of observational studies and field research buttressed Gamson's (1961) claim that portfolios will be distributed proportionally (based on coalition member based on their relative seat contribution to the coalition), such that it is often referred to as Gamson's law (e.g. Warwick and Druckman 2001). Motivated by the highly disparate predictions of Gamson's law and the Baron-Ferejohn model, other work focuses on demand bargaining (Morelli 1999). Morelli's demand bargaining model strips the formateur's agenda-setting power and enables proposers to make sequential demands so as to maximize their share of benefits. In contrast to the
format in the Baron-Forgey model who makes a proposal about how portolfios should be distributed, the foreman is distinguished only by the fact that he chooses the order in which sequential demands are made. Other parties are no longer restricted to accepting/rejecting the foreman’s proposal, but can make demands so as to maximize their own share of benefits. In a three-party bargaining situation, Morelli models the two-thirds of benefits to the proposer, one-third to the coalition partner, and nothing to the third party. While there are important cases that coalition partners may contribute to the total payoffs, the general conclusion is that it is possible to obtain equilibrium outcomes that approximate the proportional payoffs.

In one fruitful line of inquiry, a number of experiments directly test the Baron-Forgey, Gamsen, and Morelli models (e.g., Fréchette et al. 2005). These experiments are excellent examples of both ‘speaking to theorists’ and ‘searching for fact’. They provide explicit tests of competing theoretical propositions, and attempt to understand how the models hold up in practice. For example, the predictions of many of these experiments contrast with the expectations of Baron-Forgey in that the allocation of benefits is more egalitarian than expected and that many subjects reject offers above their continuation value.

Heretofore, the experimental context of the studies mentioned has not been discussed. Diermeier (2001) notes that much of this research follows in the methodology developed by experimental economics and game theory in which subjects interact anonymously through computer terminals with payments based on performance, and differs quite notably from the face-to-face paradigm used in most majority voting studies. Diermeier asserts that a number of promising lines of research rooted in psychology shift the experimental design away from the abstract economic setting and toward a context-rich environment.

Some of this research extends beyond Baron-Forgey to not only ask how coalitions are formed, but how they actually negotiate in ways that facilitate trust. Conceptualization of rich experiments enable scholars to examine a variety of dimensions that can facilitate trust, including nonverbal communication via face-to-face interactions in computer settings, public versus private communications, and the role of secret communication settings (Swaid et al. 2009). Face-to-face interactions are shown to increase bargaining efficiency, while private communications decrease efficiency, and secrecy appears to undermine trust and prevent efficient bargaining altogether. Enriching the experimental context, researchers are able to test theories of coalition formation and the Baron-Forgey model predictions in more realistic settings, creating opportunity for variation in communication strategies.

Disparities between model predictions and empirical results persist and many research questions remain unanswered. The lack of congruence between formal theoretical and empirical work has led some to suggest that the topic of government and coalition formation should be reconsidered from the “bottom up” (e.g., Golder et al. 2012). It may be beneficial for experiments to incorporate recent work on dynamic theories of bargaining in which the decisions of one period become the status quo in the next (Baron et al. 2011). Much work remains to be done unifying the findings from experiments with context-rich settings and those from an abstract traditional economic approach.

Of course underlying all legislative behaviour is the reality that actions are being taken by interpersonally oriented actors with an eye toward future elections. Thus, another important area of study—and one that has only recently been the subject of experimental inquiry—is how responsive legislators are to those who elect them.

9.5 Experiments on Responsiveness and Legislators as Subjects

Legislative coalitions do not operate in a vacuum. At the heart of both normative and empirical analyses of legislative behaviour is a concern for responsiveness. Observational studies have long since attempted to disentangle the relationship between constituency preferences, legislative behaviour, and policy outcomes. These studies are often plagued by problems of endogeneity. Thus recent experimentation has proven quite useful in directly testing how legislators respond to public preferences. Many of these studies also include real legislators functioning in their actual political environment. While better way to “whisper in the ears of princes” than to actually speak to legislators and hear their responses?

Butler and Broockman (2011) employ a cleverly designed field experiment to address how state legislators respond to constituent inquiries. Emails were sent to state legislators requesting help with registering to vote. The emails were randomized to contain either a racism or black alias as well as an indication of the sender’s partisan preference. Emulating the black alias received fewer replies, and controlling for partisanship did not erase the racial disparity. The study illustrates a pattern of discrimination where white legislators of both political parties reply to the black sender with far less frequency than minority legislators, suggesting asymmetric responsiveness to public concerns. This study raises important ethical considerations about experimental research on elected officials. This issue becomes particularly clear when the outcomes of interest in the experiment are political as opposed to individually beneficial. Butler and Nickerson (2016) randomly provided half of the members of the New Mexico State House with the results of otherwise unavailable, district-specific public opinion data on a particular matter. Other members were instead provided with a non-partisan fact sheet. The study demonstrates that when provided information in line with the views of their constituents and found that information did help legislators to follow voters’ preferences. This study provides unparalleled insights into the nature of legislators’ voting behaviour, and the study authors characterize their
findings as positive for democracy. Yet they make no mention of the ethical implications of withholding treatment from a random half of the legislature.

In a different direction, Neble et al. (2010) report on experiments focused on consent. In their studies, which used legislators as part of the treatment, the procedures examined whether constituencies, which use legislators as a part of the treatment. The studies examined types of constituents who were willing to engage in deliberation, with the type of deliberation varying randomly among citizens deliberating with their member of Congress or deliberation with other citizens. While most experimental work in political science involves public officials in treatments or outcomes, most of it uses lab-based treatments to aid the researcher's purpose. The research behind the Neble et al. is quite distinct in that it involves actual legislatures recruited to participate in the studies, which focus on who chooses to deliberate. For example, Neble et al. demonstrate that the opportunity to deliberate with a member of Congress significantly increases willingness to participate in deliberation (also see Esteb et al. 2011).

These experimental studies of legislative behavior, while quite different from the focus of voting studies, are innovative for their mundane realism (i.e., they use the real lives of real legislators as participants and real legislatures as contexts). Clearly, trade-offs between experimental realism, external validity, and ethical considerations are underdeveloped in the experimental literature on legislative behavior.

9.6 Challenges and the Future of Experiments

Our review of applications has hopefully made clear that experiments add considerable insight to what we know about legislatures. They enable researchers to overcome common problems (e.g., endogeneity, confounding variables) faced with observational data. Yet, there are significant challenges that may explain why experiments have not been as dominant a methodology in legislative studies and this section identifies some of these challenges.

Perhaps the most notable constraint concerns the nature of the experimental participants. The bulk of studies (other than those on responsiveness) rely on students or at least individuals who are not actual legislators. This causes many to worry that the results from such studies provide scant insight in the realities of legislative behavior. While we recognize it is a constraint, we also take a more optimistic approach, making two major points.

First, a cursory glance at related literature on international bargaining serves as evidence that results can be made by using student subjects. This can be done via role-playing which has proven successful in generating predictions/explanations that have clear relevance to the actual international negotiations. Much of this is done in simulations where participants are experimentally assigned to different roles and conditions within simulated settings/negotiations. Additionally, if one is worried about students per se, Kam et al. (2007) suggest recruiting non-student subjects, even for studies run on campus laboratories.

Second, it is important to clarify the goal of experimental inference. The goal of an experiment is to identify a causal relationship. This differs from observational research that often focuses on inferring univariate characteristics of a population. Experimental control aids the researcher by creating a reality that has precisely the features of interest to the researcher (e.g., particular voting rules). While individual subjects might vary in unknown ways, the use of induced value is thought to make these differences near irrelevant by heavily structuring their preferences within the experimental setting (and random assignment balances any differences between treatment groups). As induced value overpowers subjects' behavior, the differences between a given sample of subjects and the real population of legislators to which inference is intended become less important. Characteristics of subjects are only problematic when features of a sample dramatically differ from the target population and those differences affect the extent to which subjects respond to a treatment (see Druckman and Kam 2011). In general and particularly when induced value mimics preferences of the real legislature to which inference is intended, any subject population can be used to draw accurate causal inferences unless there is some unique feature of the subject population that makes them react dramatically different to the stimuli than would actual legislators.

Finally, many coalition processes of most recent interest are inherently dynamic—that is, they occur over time and/or focus on ongoing policy-making processes (e.g., Devan and Myatt 2012). Indeed, questions about the stability of legislative coalitions depend upon longer timeframes than are possibly emulated in a single laboratory session, but the costs and effort involved in such research can be quite high. The field, we conclude on a more uplifting note by emphasizing the potential for the expansion of experiments in legislative studies. As shown, they have proven quite successful in certain areas and we believe more can be done. One avenue is increased collaboration between those taking political psychological and political economic approaches—indeed, much of the advances in legislative studies employ economic-style approaches yet many political science experiments are more psychological in nature. Also notably absent from our review of applications are survey experiments. This is somewhat surprising given a long history of successfully launching surveys of either legislatures themselves or experts in certain areas. It would not be difficult to embed experimentalists such surveys to explore how respondent legislators would react to distinct scenarios (e.g., Tomz 2009).

In the end, experiments are one tool that will complement others but one that we believe should play a larger role along with advances in statistical methods, and the collection of large, collaborative data sets (e.g., Müller and Strom, 2006). Despite Lowell's statement with which we began, experiments have in fact long been a part of research on legislatures (e.g., Riker 1962), and we believe that, in line with much of the rest of the discipline, experiments could play an even more prominent role in the future.
Notes

1. Parts of this section are taken from Druckman et al. (2006, 2014, 2015).
2. The first two application sections draw generously on, respectively, Miller (2012) and Dienstler (2014) each of which reviews experimental applications in these areas.
3. However, outcomes within the uncovered set are not necessarily stable—the instability is simply constrained to a smaller region of the space.
4. The Baron-Ferejohn model predicts that unless there is no impatience, legislatures will prefer a closed rule to an open rule.
5. "An equilibrium is said to be stationary if the continuation values for each structurally equivalent subgroup are the same" (Baron and Ferejohn 1989, 1991).

References


**PART III**

**REPRESENTATION AND LEGISLATIVE CAREERS**