

An Experimental Approach to Citizen Deliberation

Christopher F. Karpowitz
Brigham Young University
ckarpowitz@byu.edu

Tali Mendelberg
Princeton University
talim@princeton.edu

Revised Draft

Word Count: 7,221 (excluding Works Cited List)
8,748 (including Works Cited List)

An Experimental Approach to Citizen Deliberation
Christopher F. Karpowitz and Tali Mendelberg¹
Revised Draft

Deliberation has become, in the words of one scholar, “the most active area of political theory in its entirety” (Dryzek 2007, 237). Our exploration of the relationship between experiments and deliberation thus begins with normative theory as its starting point. Experiments can yield unique insights into the conditions under which the expectations of deliberative theorists are likely to be approximated, as well as the conditions under which theorists' expectations fall short. Done well, experiments demand an increased level of conceptual precision from researchers of all kinds who are interested in deliberative outcomes. But perhaps most importantly, experiments can shed greater scholarly light on the complex and sometimes conflicting mechanisms that may drive the outcomes of various deliberative processes. In other words, experiments allow researchers to better understand the extent to which, the ways in which, and under what circumstances it is actually *deliberation* that drives the outcomes deliberative theorists expect.

Our strategy for this chapter will be to highlight the strengths of experiments that have already been completed and to point to some aspects of the research that need further improvement and development. We aim to discuss what experiments can do that other forms of empirical research cannot and what experiments need to do in light of the normative theory.

Proceeding from normative theory is not without its difficulties, as deliberative theorists themselves admit (see, for example, Chambers 2003; Thompson 2008). One difficulty is that theories of deliberation offer a wide-ranging, sometimes vague, and not always completely consistent set of starting points for experimental work – as Diana Mutz ruefully observes, “it may be fair to say that there as many definitions of deliberation as there are theorists” (2008, 525). We recognize, too, the inevitable slippage between theory and praxis that will lead almost every empirical test to be, in some sense, “incomplete” (Fishkin 1995). Finally, we agree that empirical researchers should avoid distorting the deeper logic of deliberative theory in the search for testable hypotheses (Thompson 2008). Experiments cannot “prove” or “disprove” theories of deliberation writ large. The critical question for experimental researchers, then, is not “does deliberation work?” but rather under what conditions does deliberation approach theorists' goals or expectations?

The literature on deliberation is much too large to allow us to provide full coverage here. A variety of additional research traditions from social psychology and from sociology can usefully inform our attempts to explore deliberative dynamics (see Mendelberg 2002 for an overview), but we focus on *political* discussion.² We aim to explore deliberation as practiced by ordinary citizens, which means we will set aside the literature on elites (Steiner et al. 2004). We set aside, too, a valuable research tradition focused on dyadic exchanges within social networks (Huckfeldt and Sprague 1995; Mutz 2006) to focus on discussions among groups, not dyads. Our focus on discussion of political issues and topics means that we will not cover the vast and

¹ We thank Lee Shaker for invaluable research assistance.

² See Gaertner et al. (1999) for a particularly informative study.

influential research on deliberation in juries (see, for example, Hastie, Penrod, and Pennington 1983; Schkade, Sunstein, and Kahneman 2000; Devine et al 2001), though we emphasize the value and importance of that research. Finally, we cannot do justice to experiments derived from formal theories (see, for example, Meirowitz 2007 or Hafer and Landa 2007).

1. The Substantive Issues: Independent and Dependent Variables

One of the challenges of empirical research on deliberation is the multiplicity of potential definitions (Macedo 1999). Still, many definitions of deliberation share a commitment to a reason-centered, “egalitarian, reciprocal, reasonable, and open-minded exchange of language” (Mendelberg 2002, 153; see also Gutmann and Thompson 1996; Burkhalter, Gastil, and Kelshaw 2002; Chambers 2003). While theories of deliberation do not agree about each of deliberation’s constituent aspects or about all of its expected outcomes (Macedo 1999), it is possible to distill a working set of empirical claims about deliberation’s effects (see Mendelberg 2002; Mutz 2008).

Hibbing and Theiss-Morse (2002) summarize three broad categories of effects: deliberation should lead to “better *citizens*,” “better *decisions*,” and a “better (that is, more legitimate) *system*.” Benefits for individual *citizens* may include increased tolerance or generosity and a more empathetic view of others (Warren 1992; Gutmann and Thompson 1996, 2004); a decrease in the set of pathologies of public opinion documented extensively since Converse’s (1964) seminal work, leading to more political knowledge, an enhanced ability to formulate opinions, greater stability opinions, and more coherence among related opinions (Fishkin 1995); a better understanding of one’s own interests; an increased ability to justify preferences with well-considered arguments (Warren 1992; Chambers 1996, 2003); a better awareness of opponents’ arguments and an increased tendency to recognize the moral merit of opponents’ claims (Habermas 1989, 1996; Chambers 1996; Gutmann and Thompson 1996, 2004); a sense of empowerment, including among those who have the least (Fishkin 1995; Bohman 1997); a greater sense of public-spiritedness (Warren 1992) and an increased willingness to recognize community values and to compromise in the interest of the common good (Mansbridge 1983; Chambers 1996; but see Sanders 1997 and Young 2000); and a tendency to participate more in public affairs (Barber 1984; Gastil, Deess, and Weiser 2002; Gastil, Deess, and Weiser 2008). Benefits for the quality of *decisions* flow from many of these individual benefits and include the idea that collective decisions or outcomes of deliberating groups will be grounded in increased knowledge, a more complete set of arguments, a fuller understanding of the reasons for disagreement, and a more generous aggregate attitude toward all groups in society, especially those who have the least (Chambers 1996; Gutmann and Thompson 2004).

As to the benefits for democratic *systems*, they center on the rise in support for the system that can follow from deliberation. Increased legitimacy for the system is a complex concept (Thompson 2008) but among other things, it can include a heightened level of trust in democratic processes and a greater sense of confidence that the process has been fairly carried out (see also Manin 1987). This sense becomes particularly important when the ultimate decision does not correspond well to an individual’s pre-deliberation preferences or when there is a deep or longstanding conflict at issue (Mansbridge 1983; Benhabib 1996; Chambers 1996).

These laudable outcomes for citizens, decisions, and systems are rooted in a fourth claim: that deliberation is a better decision-making *process*, one that is more public-spirited, more reasonable, more satisfying, and ultimately more just than adversarial and aggregative forms of decision making (Mansbridge 1983; Chambers 1996; Gutmann and Thompson 2004). The process is a crucial mediating variable in deliberation. In other words, normative theory leads empirical investigators to ask what aspects of discourse and linguistic exchange leads to the individual-level outcomes we have described, and what in turn causes those aspects of interaction. Examples of mediating variables include the content and style of interaction, such as whether deliberators use collective vocabulary such as “us” and “we” (Mendelberg and Karpowitz 2007), the number of arguments they make (Steiner et al. 2005), and the extent to which deliberators engage in a collaborative construction of meaning rather than speaking past each other (Rosenberg 2007). Focusing on such mediating variables allows scholars to investigate which aspects of the discourse cause those who have taken part in deliberation to feel that their voices were better heard, that the deliberating group functioned well as a collectivity, or that the process was more collaborative than other forms of interaction. In addition, researchers can investigate the effects of this sort of deliberative interaction for subsequent levels and forms of participation (see, for example, Karpowitz 2006; Gastil et al. 2008).

The variety of approaches to deliberative theory provides a rich set of procedural and substantive conditions, characteristics, and mechanisms that can be explored experimentally in a systematic way. Theorists differ, for example, on the desirability of consensus in deliberative procedures. The requirement of producing consensus is something that can be experimentally manipulated. In this way, empirical researchers can help to specify the relationship between the various potential characteristics of deliberation (the independent variables) and the positive outcomes (in other words, the dependent variables) theorists hope to see.

Of course, key independent variables relevant to deliberative outcomes may not only emerge from theory. They can also be found through careful attention to the real-world context of ongoing deliberative reform effort, where the variety of practices that might be subsumed under the broad heading of deliberation is extraordinary. Mansbridge (1983), Mansbridge et al. (2006), Polletta (2008), Cramer Walsh (2006; 2007), and Gastil (2008) are a few examples of scholars who contribute to our understanding of deliberation’s effects by insightful observation of real-world practices.

2. The Role of Experiments

Though sometimes styled as “experiments in deliberation,” much of the research to date has been purely observational – most often, these are case studies of specific deliberative events (e.g., Mansbridge 1983; Fishkin 1995; Eliasoph 1998; Fung 2003; Gastil and Levine 2005; Karpowitz 2006; Cramer Walsh 2006, 2007; Warren and Pearse 2008). The methods for evaluating and understanding deliberation are also diverse, including participant observation (Eliasoph 1998), survey research (Jacobs, Cook, and Delli Carpini 2009), and content analysis of discussion (see Gamson 1992; Conover, Searing, and Crewe 2002; Hibbing and Theiss-Morse 2002; Schildkraut 2005; Cramer Walsh 2006), just to name a few. Though our emphasis here is on experiments, other research designs are valuable in an iterative exchange with experiments and have a value of their own separate from experiments.

Experimentation is, as Campbell and Stanley put it, “the art of achieving interpretable comparisons” (quoted in Kinder and Palfrey 1993, 7). We follow the standard view that effective experimentation involves a high level of experimenter control over settings, treatments, and observations so as to rule out potential threats to valid inference. Like Kinder and Palfrey, we see random assignment to treatment and control groups as “unambiguously desirable features of experimental work in the social sciences” (1993, 7). In theory, random assignment may not be strictly necessary to achieve experimental control, but it does “provide a means of comparing the yields of different treatments in a manner that rule[s] out most alternative interpretations” (Cook and Campbell 1979, 5).

Much deliberation research is quasi-experimental (Campbell and Stanley 1963): it involves some elements of experimentation – a treatment, a subsequent outcome measure, and comparison across treated and untreated groups – but not random assignment. The lack of random assignment in quasi experiments places additional burdens on the researchers to sort out treatment effects from other potential causes of observed differences between groups. In other words, “quasi-experiments require making explicit the irrelevant causal forces hidden within the *ceteris paribus* of random assignment” (Cook and Campbell 1979, 6; see Esterling, Fung, and Lee 2010 for a sophisticated discussion of nonrandom assignment to conditions in a quasi-experimental setting).

The hallmark of experiments, then, is causal inference through control. Citizen deliberation lends itself quite well to control, since it need not take place within an official or even within a public context. Ordinarily, investigators are quite limited by the intimate link between the behavior we observe and fixed features of the political system. By contrast, citizens can deliberate with a set of strangers with whom they need have no prior connection, in contexts removed from organizational structures, official purviews or public spaces, and can reach decisions directed at no official or public target. In that sense, they present the experimenter a wide degree of control. Ultimately, deliberation often does exist in some dialogue with the political system. But unlike other forms of participation, it also exists outside of public and political spaces and it can therefore be separated and isolated for the study.

Experiments also allow us to measure the outcomes with a good deal more precision than nonexperimental, quasi-experimental, or field experimental studies. All groups assigned to deliberate do in fact deliberate; all groups assigned to deliberate by a particular rule do in fact use that rule; all individuals assigned to receive information do in fact receive information; and so on. We can measure actual behavior rather than self-reported behavior. Under the fully controlled circumstances afforded by the experimental study of deliberation, we need not worry that people self select into any aspect of the treatment. Intent-to-treat problems largely disappear.

All of this boosts our causal inference considerably. We note that the causal inference is not a consequence of the simulation that control provides. Unlike observational studies or quasi-experimental studies, experiments can and do go beyond simple simulation in two respects: first, the control can help to verify that it is *deliberation* and not other influences creating the consequences that we observe. In addition, the control afforded by experiments allows us to test specific aspects of the deliberation, such as heterogeneous or homogenous group composition;

the presence or absence of incentives that generate or dampen conflict of interests; the proportions of women, ethnic and racial groups; group size; the group's decision rule; the presence or absence of a group decision; the presence or absence of facilitators and the use of particular facilitation styles (Rosenberg 2007); the availability of information; or the presence of experts (Myers 2009).

In sum, the goal of experimental research should be to isolate specific causal forces and mediating or moderating mechanisms in order to understand the relationship between those mechanisms and deliberative outcomes as described by normative theorists. In addition, experiments allow us to reduce measurement error. This should lead to insights about how to better design institutions and deliberative settings.

3. Some Helpful Examples

To date, the experimental work on deliberation has been haphazard (Ryfe 2005, 64). Our aim here is not to provide an exhaustive account of every experiment, but to show the strengths and weaknesses of some of the work that has been completed. It is rare for a study to use all the elements of a strong experimental design. We find a continuum of research designs, with some studies falling closer to the gold standard of random assignment to control and treatments, and others falling closer to the category of quasi-experiments.

The best known and arguably the most influential investigation of deliberation is Fishkin's (1995) deliberative poll. In a deliberative poll, a probability sample of citizens is recruited and questioned about their policy views on a political issue. They are sent a balanced set of briefing materials prior to the deliberative event in order to spark some initial thinking about the issues. The representative sample is then brought to a single location for several days of intensive engagement, including small group discussion (with assignment to small groups usually done randomly), informal discussion among participants, a chance to question experts on the issue, and an opportunity to hear prominent politicians debate the issue. At the end of the event (and sometimes again several weeks or months afterward), the sample is asked again about their opinions, and researchers explore opinion change, which is presumed to be the result of the deliberative poll.

The first deliberative polls were criticized heavily on a variety of empirical grounds, with critics paying special attention to whether or not the deliberative poll should qualify as an experiment (Kohut 1996; Merkle 1996). Mitofsky (1996), for example, insists that problems with panel attrition in the response rates in the post-deliberation surveys made causal inferences especially difficult and that the lack of a control group made it impossible to know whether any change in individual opinion "is due to the experience of being recruited, flown to Austin, treated like a celebrity by being asked their opinions on national television and having participated in the deliberations, or just due to being interviewed twice" (19). Luskin, Fishkin, and Jowell admit that their approach fails to qualify as a full experiment by the standards of Campbell and Stanley "both because it lacks the full measure of control characteristic of laboratory experiments and because it lacks a true, i.e. randomly assigned, control group" (Luskin, Fishkin, and Jowell 2002, 460).

As the number of deliberative polls has proliferated, they have pursued a variety of innovations. For example, subsequent work has included pre- and post-deliberation interviews of both those who were recruited to be part of the deliberating panel but who chose not to attend and post-event interviews of a separate sample of nondeliberators. These additional interviews function as a type of control group, though random assignment to deliberating or nondeliberating conditions is not present. While still not qualifying as a full experiment, these additions function as an “untreated nonequivalent control group design with pretest and posttest” and as “a posttest-only control group design,” as classified by Campbell and Stanley (1963). When such additions are included, the research design of the deliberative poll does have “some of the characteristics of a fairly sophisticated quasi-experiment” (Merkle 1996), characteristics that help eliminate some important threats to valid inference.

Still, the quasi-experimental deliberative poll does not exclude all threats to valid inference, especially when the problem of self-selection into actual attendance or nonattendance at the deliberative event is considered (see Barabas 2004, 692). Like that of recent deliberative polls, Barabas’s analysis of the effects of a deliberative forum about social security is based on comparing control groups of nonattenders and a separate sample of nonattenders. To further reduce the potential for problematic inferences, Barabas makes use of propensity score analysis. Quasi-experimental research designs that make explicit the potential threats to inference or that use statistical approaches to estimate treatment effects more precisely are valuable advances (see also Esterling, Fung, and Lee 2010). These do not fully make up for the lack of randomization, but they do advance empirical work on deliberation.

We note one additional important challenge related to deliberative polling: the complexity of the deliberative treatment. As Luskin, Fishkin, and Jowell (2002) put it, the deliberative poll is “*one grand treatment*” that includes the anticipation of the event once the sample has been recruited, the exposure to briefing information, small-group discussion, listening to and asking questions of experts and politicians, informal conversations among participants over the course of the event, and a variety of other aspects of the experience (not the least of which is participants’ knowledge that they are being studied and will be featured on television). Perhaps it is the case that deliberation, as a concept, is a “grand treatment” that loses something when it is reduced to smaller facets, but from a methodological perspective, the complexity of this treatment makes it difficult to know what, exactly, is causing the effects we observe. Indeed, it denies us the ability to conclude that any aspect of *deliberation* is responsible for the effects (rather than the briefing materials, expert testimony, or some other nondeliberative aspects of the experience). Experimentation can and should seek to isolate the independent effects of each of these features.

At the other end of the spectrum from the “one grand treatment” approach of the deliberative pollsters are experiments that involve a much more spare conception of deliberation.

Simon and Sulkin (2002), for example, insert deliberation into a “divide-the-dollar” game in which participants were placed in groups of five and asked to divide sixty dollars between them. A total of 130 participants took part in one of eleven sessions, with multiple game rounds played at each session. In the game, each member of the group could make a proposal as to how to divide the money, after which a proposal was randomly selected and voted on by the group. A

bare majority was sufficient to pass the proposal. Participants were randomly assigned to one of three conditions – no discussion, discussion prior to proposals, and discussion after proposals. In addition, participants were randomly assigned to either a cleavage condition in which players were randomly assigned to be in either a three-person majority or a two-person minority and proposals were required to divide money into two sums – one for the majority and one for the minority – or a noncleavage condition in which no majority/minority groups were assigned. Simon and Sulkin find that the presence of discussion led to more equitable outcomes for all participants and especially for players who ended up being in the minority.

The experiment employs many of the beneficial features we have highlighted – control over many aspects of the setting and of measurement, and random assignment to conditions. In addition, the researchers ground their questions in specific elements of normative theories. However, the study artificially capped discussion at only 200 seconds of online communication, which detracts from its ability to speak to the lengthier, deeper exchanges that deliberative theory deals with or to the nature of real-world exchanges.

Other experimental approaches have also explored the effects of online deliberation (see, for example, Muhlberger and Weber 2006 and developing work by Esterling, Neblo, and Lazer 2008a, 2008b).³ The most well-known of these so far is the Healthcare Dialogue project undertaken by Price and Capella (2005, 2007). A year-long longitudinal study with a nationally representative pool of citizens and a panel of healthcare policy elites, this study explored the effectiveness of online deliberations about public policy. The research involved repeated surveys and an experiment in which respondents who completed the baseline survey were randomly assigned to a series of four online discussions or to a nondeliberating control group. In these discussions, participants were stratified as either policy elites, healthcare issue public members (regular citizens who were very knowledgeable about health care issues), or members of the general public. Half of the groups were homogenous across strata for the first two conversations, the other half included discussants of all three types. In the second pair of conversations, half of the participants remained in the same kind of group as in the first wave; the other fifty percent of the participants were switched from homogenous to heterogeneous groups or *vice versa*. Group tasks were, first, to identify key problems related to health care and, second, to identify potential policy solutions (though they did not have to agree on a single solution). To ensure compatibility across groups, trained moderators followed a script to introduce topics and prompt discussion and debate.

Price and Capella find that participation in online discussion led to higher levels of opinion-holding among deliberators and a substantial shift in policy preferences, relative to those who did not deliberate. This shift was not merely the result of being exposed to policy elites, as the movement was greatest among those who did not converse with elites. In addition, participants – and especially nonelites – rated their experience with the deliberation as quite satisfying. Because a random subset of the nondeliberating control group was assigned to read online briefing papers that deliberating groups used to prepare for the discussions, the experiments also allowed the researchers to distinguish the effects of information from the

³ See also Esterling, Neblo, and Lazer (2008a) on estimating treatment effects in the presence of noncompliance with assigned treatments and nonresponse to outcome measures.

effects of discussion. While exposure to briefing materials alone increased knowledge of relevant facts, discussion and debate added something more – an increased understanding of the rationales behind various policy positions.

Regardless of whether the findings were positive or negative from the perspective of deliberative theory, the research design employed by Price and Capella highlights many of the virtues of thoughtful, sophisticated experiments. The research includes a large number of participants (nearly 2,500), a significant number of deliberating groups (more than eighty in the first wave and approximately fifty in the second wave), and random assignment from a single sample (those who completed the baseline survey) to deliberation plus information, information only, and no deliberation, no information conditions, with respondents in all groups completing a series of surveys over the course of a calendar year. Price and Capella also leverage experimental control to answer questions that the “grand treatment” approach of deliberative polling cannot. For example, where deliberative polling is unable to separate the independent effects of information, discussion among ordinary citizens, and exposure to elites, Price and Capella are able to show that information has differing effects from discussion and that exposure to elites cannot explain all aspects of citizens’ opinion change. Given large number of groups and the elements of the research design that have to do with differing group-level conditions for deliberation, the Price and Capella design has the potential for even more insight into the ways group-level factors influence deliberators and deliberative outcomes, though these have not been the primary focus of their analyses to date. Still, their design may fail to satisfy some conceptions of deliberation, as groups simply had to identify potential solutions, not make a single, binding choice.

Experiments relevant to deliberation have also been conducted with face-to-face treatments, though we find considerable variation in the quality of the research design and the direct attention to deliberative theory. Morrell (1999), for example, contrasts familiar liberal democratic decision-making procedures, which include debate using Robert’s rules of order, with what he calls “generative” procedures for democratic talk, which include such deliberatively desirable elements as hearing the perspectives of all group members, active listening and repeating the ideas of fellow group members, and considerable small group discussion. His research design includes random assignment to either the liberal democratic condition, the generative condition, or a no discussion condition. Participants answered a short survey about their political attitudes at the beginning and the end of the experimental process. Morrell repeats the study with multiple issues and with differing lengths of discursive interaction. In this research design, participants made a collective decision about an issue, an element that is not present in Fishkin’s deliberative polls but that is critical to some theories of deliberation.

In contrast to the comparatively positive outcomes of the experiments in online deliberation we have highlighted, Morrell finds that the deliberatively superior generative procedures do not lead to greater group-level satisfaction or acceptance of group decisions. If anything, traditional parliamentary procedures are preferred in some cases. In addition, in several of the iterations of the experiment, Morrell finds strong mediating effects of the group outcome, contrary to deliberative expectations. Morrell’s findings thus call attention to the fact that the

conditions of group discussion, including the rules for group interaction, matter a great deal and that more deliberative processes may not lead to the predicted normative outcomes.

Though we see important strengths in Morrell's approach, we note that the reported results do not speak directly to the value of the presence or absence of deliberation. The dependent variables Morrell reports are nearly all focused on satisfaction with group procedures and outcomes, measures for which the nondeliberating control condition are not relevant. In other words, Morrell's test as reported contrasts only different types of discursive interaction. Moreover, as with our earlier discussion of the "grand treatment", the treatments in both cases are complex, and it is not entirely clear which aspects of "generative" discussion led to lower levels of satisfaction. Finally, we note that Morrell's experiments were based on a very small number of participants and an even smaller number of deliberating groups. All this makes comparison to other, conflicting studies difficult.

Druckman's (2004) study of the role of deliberation in combating framing effects is a good example of the way experiments can speak to aspects of deliberative theory. The primary purpose of Druckman's research is to explore the conditions under which individuals might be less vulnerable to well-recognized framing effects. The relevance to deliberation lies in investigating how deliberation can mitigate the irrationality of ordinary citizens and improve their civic capacities. The study advances the literature on deliberation by assessing the impact of different deliberative contexts.

Druckman presented participants with one of eight randomly assigned conditions. These conditions varied the nature of the frame (positive or negative) and the context in which the participant received the frame. Contexts included: a control condition, in which participants received only a single, randomly chosen frame; a counter-framing condition, in which participants received both a positive and a negative frame; and two group conditions, in which participants had an opportunity to discuss the framing problems with three other participants. In the homogenous group condition, all members of the group received the same frame, and in the heterogeneous condition, half of the group received a positive frame and half received the negative frame. Participants in the group condition were instructed to discuss the framing problem for five minutes. Druckman recruited a moderate number of participants (580), with approximately 172 taking part in the group discussion conditions. This means that just over forty deliberating groups could be studied.

As with the Simon and Sulkin experiment, exposure to group discussion in this research design is limited and may, therefore, understate the effect group discussion might have. But what is most helpful from the perspective of deliberative theory is a systematic manipulation of both the presence of discussion and the context under which discussion occurred. Druckman finds that the presence of discussion matters – participants in both the homogenous and heterogeneous conditions proved less vulnerable to framing effects than in the control condition. This would seem to be positive evidence for the relationship between deliberation and rationality, but the story is somewhat more complicated than that. Neither discussion condition reduced framing effects as much as simply giving the counterframe to each individual without requiring group discussion. Moreover, homogenous groups appeared to be comparatively more vulnerable to

framing effects compared to heterogeneous groups. Results were also strongly mediated by expertise.

Druckman's research design reflects several attributes worthy of emulation. First, the number of groups is sufficiently sizeable for meaningful statistical inference. Second, Druckman has used the key levers of experimental control and random assignment appropriately. This allows him to make meaningful claims about the difference between discussion across different contexts and the difference between discussion and the simple provision of additional information. As we discussed in the previous section, one of the key problems of causal inference in the "grand treatment" design has been whether deliberation is responsible for the observed effects or whether one particular aspect of it – the provision of information – is responsible. Given that information is not unique to deliberation, finding that the effects of deliberation are due primarily to information would considerably lessen the appeal and value of deliberation as a distinct mode of participation. Druckman's results do raise further questions, however, especially with respect to what is actually happening during the discussion period. Druckman does not look inside the "black box" of discussion to understand how the dynamics and the content of discussion vary across the homogenous and heterogeneous conditions.

Finally, we add a few words about our own experimental work on deliberation (Mendelberg and Karpowitz 2007; Karpowitz and Mendelberg 2007; Karpowitz, Mendelberg, and Argyle 2008). We do this in order to highlight a few of the methodological issues that have emerged as we have conducted the research. Our interest in experiments began when we reanalyzed data collected earlier by Frohlich and Oppenheimer (1992). Participants in the experiment were told that they would be doing tasks to earn money; that the money they earned would be based on a group decision about redistribution; but that prior to group deliberation, they would not be told the nature of the work they would be doing. This was meant to simulate the Rawlsian veil of ignorance, as individuals would not know the specifics of how their decision would affect them personally because they would not know how well or poorly they might perform.

During deliberation, groups were instructed to choose one of several principles of justice to be applied to their earnings, including the option not to redistribute at all. The principle chosen would simultaneously govern the income they earned during the experiment (which was translated into a yearly income equivalent) and apply (hypothetically) to the society at large. Groups were randomly assigned to one of three conditions: imposed, unanimous, and majority rule. In the imposed condition, groups were assigned a principle of justice by the experimenters. In the unanimous and majority conditions, groups had to choose a principle of justice either unanimously or by majority vote, respectively.

The key finding of Frohlich and Oppenheimer's original study was that, when given an opportunity to deliberate behind the veil of ignorance, most groups choose to guarantee a minimum income below which the worst-off member of the group would not be allowed to fall. In our reanalysis of their data (Mendelberg and Karpowitz 2007), we noted that Frohlich and Oppenheimer paid very little attention to the ways in which the group context shaped participants' attitudes and group-level outcomes. Our reanalysis showed that important features of the group context, such as the group's gender composition and its decision rule, interacted to

significantly affect group- and individual-level experimental outcomes. But the earlier data were also limited to a significant extent. First, participants in that experiment were not randomly assigned to conditions. In addition, the data did not include a sufficient number of groups of varying gender composition to be entirely confident in our statistical results.

For those reasons, we chose to conduct our own updated version of the experiment, this time with random assignment, a sufficient number of groups (nearly 150), and systematic manipulation of the various gender/decision rule conditions. We also carefully recorded each group discussion in order to explore more fully the dynamics of the group interactions themselves, tying the verbal behavior of each participant during deliberation to their pre- and post-discussion attitudes about the functioning of the group, the need for redistribution, and a host of other variables. Our analysis is still in its initial stages, but we do find evidence that the group-level factors, especially the interaction of group gender and decision rule, affect various aspects of the group's functioning and deliberative dynamics. We also find significant differences between groups that deliberate and control groups that did not.

In sum, in our work we have attempted to advance the study of deliberation methodologically in several ways. We use a larger N, particularly increasing the group N; we employ random assignment; and our design both controls on deliberation itself and isolates the effects of specific aspects of deliberation, some of which derive from empirical studies of citizen discussion (decision rule, the group's heterogeneity or homogeneity), some of which focus on normatively relevant processes of communication (such as equal participation in discussion, use of linguistic terms reflecting a concern for the common good), and some of which supplement these theories by focusing on sociologically important variables such as the group's demographic composition. In conducting these experiments, we have also begun to confront directly some of the practical challenges inherent in attempting to implement random assignment of individuals to group conditions.

Having outlined both positive features and further questions that emerge from several highlighted experiments, we turn next to some of the challenges of effective experimentation about deliberation.

4. Challenges

We have detailed an argument in which we urge more investigations using experimental methods, more care in designing treatments that manipulate various aspects of deliberation, and particularly more frequent use of random assignment to conditions. However, the more control the investigator seeks, the greater the tradeoffs. Control brings artifice and narrow, isolated operationalizations of rich and complex concepts. The behavior of interest is often embedded in the contexts of institutions and social relationships and must therefore ultimately be moved back out of the lab, where every effort was made to isolate it, and studied all over again with attention to these contexts. There are other difficulties involved in the use of experiments – they may be more expensive and effortful than other methods. Here we consider these tradeoffs.

One of the challenges of experimentation is the operationalization of idealized normative theory. Experimental approaches may be particularly vulnerable to the disagreements between

theorists and empiricists to the extent that their heightened levels of control bring more stylized and more artificial operationalizations of complex and multifaceted theoretical concepts. We illustrated the issue with a contrast between Fishkin's "grand treatment" versus Simon and Sulkin's decision to trade off the complexity of deliberation against the ability to control it. The tradeoff is understandable, but not necessary. It is possible to design a controlled experiment with random assignment with multiple conditions, one of which resembles the "grand treatment" notion, others of which isolate each of the major elements of the deliberation, and with a control condition identical in every way but lacking all of these elements.

A related second challenge for experimentation is external validity. It is difficult not only to adequately operationalize the key concepts of normative theory – to achieve construct validity (Campbell and Stanley (1963) – but to simulate the causal relationships as they occur in the real world. We need to know ultimately how deliberative efforts interact with real-world actors and institutions. For example, Karpowitz's (2006) study of a local civic deliberation suggests that the deliberators' knowledge that they could pursue their preferences after the deliberation was over by lobbying the city council, writing letters to newspapers, and filing law suits in the courts, significantly affected various aspects of deliberation, including the ability of the deliberation to change minds, enlarge interests, resolve conflicts, and achieve other ends envisioned by normative theorists. This presents a challenge to the external validity of experiments in that their deliberative situation is abstracted from interaction with real institutions. On the other hand, the cumulation of findings such as Karpowitz's from observational studies can, in turn, lead to further hypothesis testing using experimental designs, where the impact of particular institutional contexts can be isolated and studied rigorously.

Mansbridge's (1983) study of a New England town meeting is the classic example of how a careful observational study can lead to theoretically rich insights into group discussion and decision making as practiced in the real world. Mansbridge shows, for example, how residents of the town struggle to navigate their common and conflicting interests in group settings and how the presence or absence of conflicting interests shapes the dynamics of the discussion and patterns of attendance at the town meeting. The textured details of real-world observation found in Mansbridge's work are often lacking in experimental studies, but her work can also be seen as articulating a set of hypotheses that can be explored much more deeply with the control that experiments provide.

Another aspect of external validity is that, in real-world settings, citizens often have to choose to deliberate. Karpowitz's (2006) analysis of patterns of meeting attendance in a national sample suggests that people who attend meetings are not a random sample of the adult population – they are more opinionated (though not more ideologically extreme) than nonattenders, but also more interested in politics, more knowledgeable about it, and more likely to discuss political issues frequently. In a controlled experiment, people also exercise some level of choice as to whether to participate, but to a much lesser extent. An experiment described up front as focused on deliberation may better approximate a real-world setting in which people choose to participate in deliberation, and people may choose to participate in the experiment for the same reasons they choose to participate in real-world deliberations. But some deliberation experiments may not be described that way. The question then becomes to what extent are the

processes and effects of deliberation generalizable from the sample in the experiment to the samples in the real world.

Attending to the relationship between deliberating groups and the wider political context, and to the differences between those who choose to deliberate and those who do not, also raises the question of how deliberation might affect those who do not participate directly, but who view the deliberation of others or who merely read about the work of deliberating groups. Given the problem of scale, deliberation is unlikely to be all-inclusive, and those who sponsor opportunities for deliberation must also communicate their processes and results to the wider public. How those who were not part of the discussion understand deliberating groups is a topic worth considerable additional study, including with experimental approaches.

A third and final difficulty of randomized experiments is of a practical nature. It is extremely difficult to implement random assignment in the study of deliberation. One variant of this problem comes in the “grand treatment” design. There, the holistic treatment and attempt to approximate the ideal conditions specified by normative theory require a substantial commitment of time and effort by deliberators. A significant percentage of those assigned to a demanding deliberation condition may well refuse treatment, and the decision to drop out of the treatment condition may well be nonrandom, introducing bias into the estimates of causal effects. Lab-based deliberation experiments may face less severe problems because random assignment takes place after subjects come to the lab, so that participants are less likely to opt out of the treatment due to its demanding nature.

Another variant of this problem presents itself when variables of interest are at the group level. This requires a large number of groups, which in turn requires a much larger individual n than lab experiments typically use. The practical challenges of random assignment to group conditions can be significant, especially when potential participants face a variety of different time constraints. For example, in our work on group composition and deliberation, we found that simultaneously accounting for differences in participants’ availability and instituting a random assignment procedure that ensures each recruited participant a roughly equal chance of being assigned to all relevant groups is a complicated exercise.

5. Conclusion -- What’s Next?

We began with the notion that empirical research can usefully evaluate the claims of deliberative theorists, and we have developed an argument about the special utility of controlled experiments. The control afforded by experiments allows not only strong causal inference but also the ability to measure, and therefore to study, mediating and outcome variables with a heightened level of precision and accuracy. We have argued that despite a proliferation of self-titled deliberative “experiments,” methodologically rigorous research design with sufficient control and random assignment is still a relative rarity. We are anxious to see experiments with an increased number of participants and especially an increased number of groups. Experimental approaches can also use their high level of control to measure the exchange of language – that is, we can train the analytical microscope more directly on the process of deliberation itself, though this practice is also still rare.

While experimental control allows for unique causal inference, experiments miss some of the richness of real-world deliberative settings. In-depth observational case studies can fill the gap and uncover the meaning of key concepts (e.g., Mansbridge 1983; Eliasoph 1998). Indeed if we were forced to choose between Mansbridge's classic work and many experiments, we might prefer Mansbridge's. The ideal research design is an iterative process in which experimentation in the lab is supplemented and informed by observation of real-world settings. Our ecumenism is not, however, a call for a continuation of the hodgepodge of studies that currently characterizes the field. Instead, we need a tighter link between the variables observed in real-world discussions and those manipulated in controlled settings. In addition, the external validity concerns typical of experiments generally apply in the case of deliberation, and may be addressed by supplementing controlled experiments with field experiments that make use of the explosion of deliberative reform efforts in cities and towns across the United States. Field experiments will be especially helpful if they allow the investigator access to accurate measures of mediating and outcome variables. It is unclear whether they do in fact allow such a degree of access or not, but the effort is worth making.

References

- Barabas, Jason. 2004. "How deliberation affects policy opinions." *American Political Science Review* 98: 687–701.
- Barber, Benjamin. 1984. *Strong Democracy: Participation politics for a new age*. Berkeley: University of California Press.
- Benhabib, Seyla. 1996. "Toward a deliberative model of democratic legitimacy." In *Democracy and Difference: Contesting the Boundaries of the Political*, ed. Seyla Benhabib. Princeton, NJ: Princeton University Press.
- Bohman, James. 1997. "Deliberative Democracy and Effective Social Freedom: Resources, Opportunities and Capabilities." In *Deliberative Democracy*, eds. James Bohman and William Rehg. Boston, MA: MIT Press.
- Burkhalter, Stephanie, John Gastil, and Todd Kelshaw. 2002. "A Conceptual Definition and Theoretical Model of Public Deliberation in Small Face-to-Face Groups." *Communication Theory* 12: 398-422.
- Campbell, Donald T., and Julian Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. New York: Wadsworth Publishing.
- Chambers, Simone. 1996. *Reasonable Democracy*. Ithaca: Cornell University Press.
- Chambers, Simone. 2003. "Deliberative Democratic Theory." *Annual Review of Political Science* 6: 307-26.
- Conover, Pamela J., David D. Searing, and Ivor Crewe. 2002. "The Deliberative Potential of Political Discussion." *British Journal of Political Science* 32: 21-62.
- Converse, Philip E. 1964. "The nature of belief systems in mass publics." In *Ideology and discontent*, ed. David E. Apter. New York: Free Press.
- Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. New York: Wadsworth Publishing.
- Cramer Walsh, Katherine. 2006. "Communities, Race, and Talk: An Analysis of the Occurrence of Civic Intergroup Dialogue Programs." *Journal of Politics* 68:22-33.
- Cramer Walsh, Katherine. 2007. *Talking About Race: Community Dialogues and the Politics of Difference*. Chicago: University of Chicago Press0073
- Devine, Dennis J., Laura Clayton, Benjamin B. Dunford, Rasmy Seying, and Jennifer Price. 2001. "Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups." *Psychology, Public Policy, and Law* 73(3): 622-727.
- Druckman, James N. 2004. "Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects." *American Political Science Review* 98: 671-686.
- Dryzek, John S. 2007. Theory, Evidence, and the Tasks of Deliberation. In *Deliberation, Participation and Democracy: Can the People Govern?* ed. Shawn W. Rosenberg. Basingstoke, UK: Palgrave Macmillan.
- Eliasoph, Nina. 1998. *Avoiding Politics How Americans Produce Apathy in Everyday Life*. Cambridge, UK: Cambridge University Press.
- Esterling, Kevin, Archon Fung, and Taeku Lee. 2010. "How Much Disagreement Is Good for Democratic Deliberation? The California Speaks Health care Reform Experiment." Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.

- Esterling, Kevin, Michael Neblo, and David Lazer. 2008a. "Estimating Treatment Effects in the Presence of Selection on Unobservables: The Generalized Endogenous Treatment Model." Paper presented at the annual meeting of the American Political Science Association, Chicago, IL.
- Esterling, Kevin, Michael Neblo, and David Lazer. 2008b. "Means, Motive, and Opportunity in Becoming Informed about Politics: A Deliberative Field Experiment." Paper presented at DG.O2007, the annual meeting of the Digital Government Society, Philadelphia, PA.
- Fishkin, James S. 1995. *The Voice of the People*. New Haven, CT: Yale University Press.
- Frohlich, Norman, and Joe Oppenheimer. 1992. *Choosing Justice*. Cambridge: Cambridge University Press.
- Fung, Archon. 2003. "Deliberative Democracy and International Labor Standards." *Governance* 16: 51-71.
- Gamson, William A. 1992. *Talking Politics*. Cambridge: Cambridge University Press.
- Gaertner, Samuel L., John F. Dovidio, Mary C. Rust, Jason A. Nier, Brenda S. Banker, Christine M. Ward, Gary R. Mottola, and Missy Houlette. 1999. "Reducing intergroup bias: elements of intergroup cooperation." *Journal of Personality and Social Psychology* 76:388-402.
- Gastil, John. 2008. *Political communication and deliberation*. Thousand Oaks, CA: Sage.
- Gastil, John, E. Pierre Deess, and Phil Weiser. 2002. "Civic awakening in the jury room: A test of the connection between jury deliberation and political participation." *Journal of Politics* 64: 585-595.
- Gastil, John, E. Pierre Deess, Phil Weiser, and Jordan Meade. 2008. "Jury service and electoral participation: A test of the participation hypothesis." *Journal of Politics* 70:1-16.
- Gastil, John, and Peter Levine. 2005. [*The deliberative democracy handbook: Strategies for effective civic engagement in the twenty-first century*](#). San Francisco, CA: Jossey Bass.
- Gutmann, Amy, and Dennis Thompson. 1996. *Democracy and Disagreement*. Cambridge, MA: Harvard University Press.
- Gutmann, Amy, and Dennis Thompson. 2004. *Why Deliberative Democracy?* Princeton: Princeton University.
- Habermas, Jurgen. 1989. *The Structural Transformation of The Public Sphere: An Inquiry Into a Category of Bourgeois Society*, trans. by T. Burger with the assistance of F. Lawrence. Cambridge, MA: MIT Press.
- Habermas, Jurgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.
- Hafer, Catherine, and Dimitri Landa. 2007. "Deliberation as Self-discovery and Institutions for Political Speech." *Journal of Theoretical Politics* 19: 325-55.
- Hastie, Reid, Steven D. Penrod, and Nancy Pennington. 1983. *Inside the Jury*. Cambridge, MA: Harvard University Press.
- Hibbing, John, and Elizabeth Theiss-Morse. 2002. "The Perils of Voice: Political Involvement's Potential to Delegitimate." Presented at the annual meeting of the American Political Science Association, Boston, MA.
- Huckfeldt, Robert, and John Sprague. 1995. *Citizens, politics and social communication: Information and influence in an election campaign*. Cambridge and New York: Cambridge University Press.

- Jacobs, Lawrence, Fay Lomax Cook, and Michael X. Delli Carpini. 2009. *Talking Together: Public Deliberation and Politics in America*. Chicago: University of Chicago Press.
- Karpowitz, Christopher F. 2006. "Having a Say: Public Hearings, Deliberation, and American Democracy." Dissertation. Princeton University.
- Karpowitz, Christopher F., and Tali Mendelberg. 2007. "Groups and Deliberation." *Swiss Political Science Review* 13: 645-62.
- Karpowitz, Christopher, Tali Mendelberg, and Lisa Argyle. 2008. "Group Effects and Deliberation: The Deliberative Justice Experiment. Paper presented at the annual meeting of the ISPP 31st Annual Scientific Meeting, Sciences Po, Paris, France.
- Kinder, Donald R., and Thomas R. Palfrey. 1993. *Experimental Foundations of Political Science*. Ann Arbor, MI: University of Michigan Press.
- Kohut, Andrew. 1996. "The Big Poll That Didn't." *Poll Watch* 4: 2-3.
- Luskin, Robert C., James S. Fishkin, and Roger Jowell. 2002. "Considered Opinions: Deliberative Polling in Britain." *British Journal of Political Science* 32: 455-487.
- Macedo, Stephen. 1999. Introduction. In *Deliberative Politics: Essays on Democracy and Disagreement*, ed. Stephen Macedo. New York: Oxford University Press.
- Manin, Bernard. 1987. "On Legitimacy and Political Deliberation." *Political Theory* 15: 338-368.
- Mansbridge, Jane. 1983. *Beyond Adversary Democracy*. Chicago: University of Chicago Press.
- Mansbridge, Jane, Janette Hartz-Karp, Matthew Amengual, and John Gastil. 2006. "Norms of deliberation: An inductive study." *Journal of Public Deliberation* 2: Article 7.
- Meirowitz, Adam. 2007. "In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values." *Journal of Theoretical Politics* 19: 301-327.
- Mendelberg, Tali. 2002. "The Deliberative Citizen: Theory and Evidence." In *Political Decision Making, Deliberation and Participation: Research in Micropolitics, Volume 6*, eds. Michael X. Delli Carpini, Leonie Huddy, and Robert Y. Shapiro. Greenwich, CT: JAI Press.
- Mendelberg, Tali, and Christopher Karpowitz. 2007. "How People Deliberate about Justice." In *Deliberation, Participation and Democracy: Can the People Govern?* ed. Shawn W. Rosenberg. Basingstoke, UK: Palgrave Macmillan.
- Merkle, Daniel M. 1996. "The National Issues Convention Deliberative Poll." *Public Opinion Quarterly* 60: 588-619.
- Mitofsky, Warren J. 1996. "It's Not Deliberative and It's Not a Poll." *Public Perspective* 7: 4-6.
- Morrell, Michael. 1999. "Citizens' Evaluations of Participatory Democratic Procedures: Normative Theory Meets Empirical Science." *Political Research Quarterly* 52: 293-322.
- Muhlberger, Peter, and L. M. Weber. 2006. "Lessons from the Virtual Agora Project: The Effects of Agency, Identity, Information, and Deliberation on Political Knowledge." *Journal of Public Deliberation* 2: Article 6.

- Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative vs Participatory Democracy*. Cambridge and New York: Cambridge University Press.
- Mutz, Diana C. 2008. "Is Deliberative Democracy a Falsifiable Theory?" *Annual Review of Political Science* 11: 521-38.
- Myers, C. Daniel. 2009. "Information Sharing in Democratic Deliberation." Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL.
- Polletta, Francesca. 2008. "Just Talk: Public Deliberation after 9/11." *Journal of Public Deliberation* 4: 1-24.
- Price, Vince, and Joe N. Cappella. 2005. "Constructing electronic interactions among citizens, issue publics, and elites: The Healthcare Dialogue project." Proceedings of the National Conference on Digital Government Research, Atlanta, GA, pp. 139-140. Digital Government Research Center.
- Price, Vince, and Joe N. Cappella. 2007. "Healthcare Dialogue: Project highlights". Proceedings of the National Conference on Digital Government Research, Philadelphia, PA, p. 178. Digital Government Research Center.
- Rosenberg, Shawn W... 2007. *Deliberation, Participation and Democracy: Can the People Govern?* Basingstoke, UK: Palgrave Macmillan.
- Ryfe, David M. 2005. "Does deliberative democracy work?" *Annual Review of Political Science* 8: 49-71.
- Sanders, Lynn M. 1997. "Against deliberation." *Political Theory* 25: 347-376.
- Schildkraut, Deborah. 2005. *Press One for English: Language Policy, Public Opinion, and American Identity*. Princeton, NJ: Princeton University Press.
- Schkade, David, Cass R. Sunstein, and Daniel Kahneman. 2000. "Deliberating about Dollars: The Severity Shift." *Columbia Law Review* 100: 1139-1176.
- Simon, Adam, and Tracy Sulkin. 2002. "Discussion's Impact on Political Allocations: An Experimental Approach." *Political Analysis* 10: 402-411.
- Steiner, Jurg, Andre Bachtiger, Markus Spornli, and Marco Steenbergen. 2005. *Deliberative Politics in Action: Analyzing Parliamentary Discourse (Theories of Institutional Design)*. Cambridge, UK: Cambridge University Press.
- Thompson, Dennis 2008. "Deliberative democratic theory and empirical political science." *Annual Review of Political Science* 11: 497-520.
- Warren, Mark E. 1992. "Democratic Theory and Self-Transformation." *American Political Science Review* 86: 8-23.
- Warren, Mark E., and Hilary Pearse. 2008. Introduction. In *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*, eds. Mark E. Warren and Hilary Pearse. Cambridge, UK: Cambridge University Press.
- Young, Iris Marion. 2000. *Inclusion and Democracy*. New York: Oxford University Press.