# Experimental Political Science and the Study of Causality

## From Nature to the Lab

Rebecca B. Morton

Kenneth C. Williams

3. Controlling observable confounding factors, and

4. Random assignment of subjects to manipulations to control for unobservable confounding factors and observable confounding factors that are difficult to control for otherwise.

In Chapter 9 we comprehensively investigate the choice of a target population. But we have argued that the key component of experimentation is the second feature, intervention and manipulation in the DGP. Some research with human subjects, where such intervention and manipulation is not the intended goal, we do not include as experiments. Similarly, other research activities that have the label experiment, such as natural experiments and computer simulations, we also do not include because they do not involve intervention or manipulation in the DGP by a researcher.

Furthermore, we have argued that although the second two aspects – control and random assignment – are valuable features of experimentation, experiments can vary significantly in how these two components are manifested. Sometimes control is manifested in the use of baseline manipulations and a high degree of control over the environment of the experiment; other times control is only minimally used by an experimentalist. Correspondingly, sometimes random assignment is implemented nearly perfectly, in which all subjects comply with assignments and their choices are fully observable; other times experimentalists must deal with noncompliance with manipulations and an inability to observe subjects' behavior (nonresponse).

Because these two aspects of experimentation are not binary but are closer to continuous variables, we do not define experiments by whether they have a given level of control or random assignment. Instead, we argue that the degree that control and random assignment are used depends on the goal of the experiment, something that we explore more expansively in the chapters to come, with an extensive discussion of control in Chapter 4 and random assignment in Chapter 5. Before we turn to our detailed examination of these aspects of experimentation, we present the Rubin Causal Model, which is one of the main approaches that underlies causal inference in experimental political science. We discuss the second approach to causality in experimental political science, the Formal Theory Approach, in Chapter 6.

## 2.6  Appendix: Examples

**Example 2.1 (Newspaper Field Experiment):** *Gerber et al. (2007) report on an experiment conducted during the 2005 Virginia gubernatorial*

*election, designed to see if biased information sources affected voter behavior.*

**Target Population and Sample:** *Gerber, Kaplan, and Bergan selected a set of residents in Prince William County, Virginia, which is 25 miles from Washington, D.C. The subjects were selected about one month before the gubernatorial election in two waves from two different lists – a list of registered voters and a consumer database list. The registered voter list provided names of 54% of the first wave and 46% of the second wave; the consumer database provided the other names. Once these names were selected, the residents were surveyed in September 2005 and were asked if anyone in the household received either the* Washington Post *or the* Washington Times *newspapers. Respondents who answered "yes" or refused to answer any one of the questions in the survey were excluded from the study. The other questions were about the respondent's newspaper readership in general, the respondent's gender, whether he or she had voted in previous elections, and who she or he supported in the coming gubernatorial election.[11] This yielded a total of 3,347 subjects.*

**Environment:** *Gerber, Kaplan, and Bergan make use of the fact that Washington, D.C., has two national newspapers, the* Washington Post *and the* Washington Times; *the first is generally viewed as a liberal newspaper and the second is widely viewed as a conservative paper. Furthermore, researchers have found empirical evidence in support of the popular perceptions. Groseclose and Milyo (2005) compare the similarity of the experts used by media outlets with those cited by conservative and liberal members of Congress. From this comparison, they construct a measure of the ideological bias of newspapers. They find that the* Times *is the most conservative of the six papers they evaluate and the* Post *is much more liberal. Furthermore, the* Post *had endorsed the Democratic candidate for the Virginia gubernatorial election and the* Times *had endorsed the Republican. Thus, they have the opportunity to compare the effects of exposure on voting behavior in the Virginia gubernatorial election of two apparently very different news sources where the framing of stories, priming on issues, and persuasive efforts are arguably biased toward different candidates.*

**Procedures:** *The subjects were randomly assigned to one of three groups: a group that received a free one-month subscription to the* Post, *a group that received a free one-month subscription to the* Times, *and a group that received neither offer. Prior to the randomization, the sample was stratified into groups based on who they planned to vote for, whether they subscribe to another*

---

[11] Initially half of the subjects were to be asked if they would like a free one-month subscription to a national newspaper as a thank you for completing the survey. But this question was dropped early in the study.

(non-Post, non-Times) newspaper, whether they subscribe to news magazines, and whether they were asked if they wished they read the paper more (50% of the subjects were asked this question). The stratification was designed so that Gerber, Kaplan, and Bergan had a balance on these covariates across the groups and the proportion of subjects in the groups was constant across strata. This randomization took place in two waves to maximize the time that subjects received the newspapers. We discuss such stratification techniques in Chapter 6. Households were given the option of canceling the subscriptions; approximately 6%, roughly equal between the Post and Times, canceled.

The newspapers were unable to deliver to some of the addresses (76 of those assigned to the Times and 1 of those assigned to the Post). Gerber, Kaplan, and Bergan employed a research assistant to monitor the delivery of newspapers to a random sample of households in the newspaper groups. While the Post had been delivered, the Times was not observed at all of the assigned addresses. Gerber, Kaplan, and Bergan spoke to the Times circulation department and called a small random sample of households assigned to receive the Times to verify their delivery.

Moreover, when the lists of households to receive the newspapers were sent to the newspapers, 75 of the households assigned the Post were reported to already subscribe to the Post (although it is unclear if they were subscribers only to the Sunday Edition or to the regular newspaper as well) and 5 of those assigned the Times were reported to already subscribe to the Times.

After the gubernatorial election, Gerber, Kaplan, and Bergan reinterviewed 1,081 of the subjects, a response rate of approximately 32%. Gerber, Kaplan, and Bergan reported that (p. 11)

[t]he remainder was not reached because the individual refused to participate in the follow-up survey (29.7%), the individual asked for was not available at the time of the call (10.3%), the operator reached an answering machine (9.8%), or the individual only partially completed the survey (6%). The operators were unable to reach the remainder for a number of different reasons, including reaching a busy signal, being disconnected, or getting no answer on the phone.... The follow-up survey asked questions about the 2005 Virginia Gubernatorial election (e.g. did the subject vote, which candidate was voted for or preferred), national politics (e.g. favorability ratings for Bush, the Republicans, the Democrats, support for Supreme Court nominee Samuel Alito), and knowledge of news events (e.g. does subject know number of Iraq war dead, has subject heard of I. Lewis Libby).

**Results:** Gerber, Kaplan, and Bergan found that those assigned to the Post group were eight percentage points more likely to vote for the Democratic candidate for governor than those not assigned a free newspaper. They also found similar evidence of differences in public opinion on specific issues and attitudes, but the evidence was weaker.

**Comments:** The results provide evidence that the biases in the news can affect voting behavior and political attitudes. The experiment is also a good illustration of randomization within strata and the measurement of causality using Intention-to-Treat, both of which we discuss in Chapter 5.

**Example 2.2 (Clientelism Field Experiment):** Wantchekon (2003) reported on a field experiment in Benin during a naturally occurring election in which candidates manipulated their campaign messages to test voter responses to messages of clientelist versus public policy messages.

**Target Population and Sample:** With the help of consultants, Wantchekon approached the leaders of four of the six political parties in Benin, which included the candidates of the two major parties. In Benin, voters are divided into eighty-four electoral districts. Wantchekon chose eight districts that are noncompetitive: four dominated by the incumbent government and four dominated by the opposition government. He also chose two competitive districts. The selection of districts was done in consultation with the campaign managers of the candidates. Within these districts, Wantchekon drew random samples for his postelection survey using standard survey sampling methods.

**Environment:** Benin is considered one of the most successful cases of democratization in Africa with a tradition of political experimentation. The election was a first-round election in which all expected a subsequent run-off election between the two major parties' candidates. Finally, candidates typically use a mixture of clientelism and public policy appeals in their election campaigns.

**Procedures:** Within each experimental district, two villages were chosen. In noncompetitive districts, one village was exposed to a clientelist platform, and the other a public policy platform. In the competitive districts, the manipulation differed; in one village one candidate espoused a clientelist platform and the other candidate a public policy platform, and in the other village the roles were reversed. The remaining villages in the selected districts were not exposed to the manipulation. The noncompetitive districts were ethnically homogenous and were less likely to be exposed to the nonexperimental manipulated campaign. The villages within each district were similar in demographic characteristics. Wantchekon took care to select villages that were physically distant from each other and separated by other villages so that, given normal communications, the manipulation was contained.

The experimental platforms were carefully designed in collaboration with the campaign managers. The public policy message emphasized "national unity and peace, eradicating corruption, alleviating poverty, developing agriculture and industry, protecting the rights of women and children, developing rural credit, providing access to the judicial system, protecting the environment, and fostering educational reforms." The clientelism message consisted of a specific

*promise to the village for things like government patronage jobs or local public goods, "such as establishing a new local university or providing financial support for local fishermen or cotton producers."*

After devising the platforms, ten teams of campaign workers were created and trained. Each team had two members, one a party activist and the other a nonpartisan research assistant. The team trained, monitored, and supervised campaign workers. There were also statisticians who served as consultants. Wantchekon (p. 410) describes how the messages were conveyed to voters:

*During each week for three months before the election, the campaign workers (one party activist and one social scientist) contacted voters in their assigned villages. With the help of the local party leader, they first settled in the village, contacted the local administration, religious or traditional authorities, and other local political actors. They then contacted individuals known to be influential public figures at home to present their campaign messages. They met groups of ten to fifty voters at sporting and cultural events. They also organized public meetings of fifty to one hundred people. On average, visits to household lasted half an hour and large public meetings about two hours.*

In the post-election surveys, voters were asked demographic characteristics, degree of exposure of messages, and voting behavior.

**Results:** Wantchekon found that clientelism worked as a campaign message for all types of candidates but was particularly effective for regional and incumbent candidates. He also found that women had a stronger preference for public goods messages than men.

**Comments:** Wantchekon's experiment is an unusual example of a case where political candidates were willing to manipulate their messages substantially in a naturally occurring election. The experiment raises some ethical issues of the influence of experimentalists in the DGP, although most of the manipulation took place in noncompetitive districts in an election that was widely seen as not significantly consequential given the likelihood that a run-off election would be held. We return to these ethical issues in Chapters 11 and 12.

**Example 2.3 (Negative Advertising Internet Survey Experiment):** *Clinton and Lapinski (2004) report on an Internet survey experiment on the effects of negative advertising on voter turnout.*

**Target Population and Sample:** *Clinton and Lapinski used a national panel in the United States created by Knowledge Networks (KN). Information on KN can be found at http://www.knowledgenetworks.com/index3.html. Another Internet-based survey organization which has been used by political scientists is Harris Interactive (see http://www.harrisinteractive.com/). The*

*panelists were randomly selected using list-assisted random-digit-dialing sampling techniques on a quarterly updated sample frame from the entire U.S. telephone population that fell within the Microsoft Web TV network, which at the time of the study was 87% of the U.S. population. The acceptance rate of KN's invitation to join the panel during the time of the study averaged 56%. Clinton and Lapinski randomly selected eligible voters from the KN panel for their study.*

**Subject Compensation:** *The panelists were given an interactive television device (Microsoft Web TV) and a free Internet connection in exchange for participating in the surveys. Participants are expected to complete one survey a week to maintain the service.*

**Environment:** *Clinton and Lapinski conducted their experiment during the 2000 presidential general election campaign and they used actual advertisements aired by the two major candidates, Bush and Gore. Subjects took part in the experiment in their own homes, although the subjects had to use the Web TV device to participate in the experiment. This somewhat reduced the variance in subjects' survey experience.*

**Procedures:** *An email was sent to the Web TV account of the selected subjects, informing them that their next survey was ready to be taken. Through a hyperlink, the subjects reached the survey. The response rate was 68% and on average subjects completed the survey within 2.7 days of being sent the email. The subjects were asked a variety of questions, both political and nonpolitical, for other clients of KN, but Clinton and Lapinski's questions were always asked first. During the survey, those subjects who had been randomly chosen to see one or more political advertisements were shown a full-screen advertisement and then asked a few follow-up questions.*

*The subjects were approached in two waves. The two waves in the experiment tested between different manipulations. In wave I, Clinton and Lapinski investigated the effect of being shown a single or pair of advertisements on Gore on the likelihood of voting, and in wave II, Clinton and Lapinski investigated the effect of seeing a positive or negative Bush advertisement conditioned on seeing a Gore negative advertisement. In wave I, subjects were divided into four groups depending on the types of advertisements shown: manipulation A (Gore negative and positive), manipulation B (Gore positive), manipulation C (Gore negative), and a group that was not shown an advertisement. Wave I took place between October 10, 2000, and November 7, 2000, with the median respondent completing his or her survey on October 12, 2000. In wave II, subjects were divided into three groups: manipulation D (Gore negative, Bush positive), manipulation E (Gore negative, Bush negative), and a group that was not shown an advertisement. Wave II took place between October 30,*

*2000, and November 5, 2000, and the median respondent completed his or her survey on November 1, 2000. Overall 2,850 subjects were assigned to groups A, B, and C; 2,500 were assigned to groups D and E. In wave I, 4,614 subjects did not see an advertisement; in wave II, 1,500 did not see an advertisement.*

*After being shown the ad or ads in both waves, subjects were asked the likelihood that they would vote in the presidential election. The question wording was slightly different in the two waves, with five options in wave I, and ten options in wave II. Finally, after the election, subjects were surveyed again, asking whether they had voted; 71% of the subjects responded to the post-election survey request.*

**Results:** *Clinton and Lapinski found no evidence that the negative advertisements demobilize voters either using the initial probability of voting question or the post-election self-reported turnout question. They also found, when they controlled for respondent characteristics, that there is no mobilization effect of the campaign advertisements either. They argued that their results suggest that the effects of the manipulations are dependent on voter characteristics and the issues discussed in the advertisements and not the overall tone of the ads.*

**Comments:** *The group that did not see an advertisement in wave I was not a random sample devised by Clinton and Lapinski, but due to a technical difficulty that was known prior to the administration of the survey. However, Clinton and Lapinski state that the group was "essentially" random and that they used demographic controls in the analysis. Clinton and Lapinski analyzed the data using both the manipulations as independent variables and other demographic variables that can matter for turnout and that varied by manipulation group. We discuss the reasoning behind these estimation strategies in Section 4.2.8.*

**Example 2.4 (Candidate Quality Lab Experiment):** *Kulisheck and Mondak (1996), hereafter Mondak1; Canache et al. (2000), hereafter Mondak2; and Mondak and Huckfeldt (2006), hereafter Mondak3, reported on a series of experiments investigating how voters respond to information about the quality of candidates, independent of issue positions. Mondak et al. refers to all three experiments.*

**Target Population and Sample:** *Mondak et al. used as subjects undergraduate students enrolled in political science classes at universities in Mexico, Venezuela, and the United States. Mondak1 used 452 students at the University of Pittsburgh; Mondak2 used 130 students at two universities in Caracas (Univeridad Católica Andrés Bello and Universidad Simón Bolívar) and 155 students at three universities in Mexico (Universidad de las Américas, Universidad Autónoma de Méjico-Xochimilco, and Centro de Investigación y Docencia Económica de Méjico); and Mondak3 used 223 students at Indiana University.*

**Subject Compensation:** *Mondak et al. did not report whether the subjects were compensated for their participation. Presumably, however, they were compensated by credit in the political science classes they were taking.*

**Environment:** *The experiments reported on in Mondak1 and Mondak2 were conducted in classrooms using pen and paper. The experiments reported on in Mondak3 were conducted using computers. In all of the experiments, the candidates that subjects were presented with were hypothetical. In Mondak1 and Mondak2, subjects were given detailed information about the hypothetical candidates and asked to read material similar to what would appear in a local newspaper in a "meet the candidates" format. In Mondak3, subjects were presented information about the candidates via computer, but the information was more limited. An important factor in the experiments was that while in the United States there is usually much discussion about personal skill and the integrity of candidates for Congress, for legislative positions in Mexico and Venezuela the electoral system in place at the time of the experiments did not encourage voter discussion of these issues and voters rarely had this sort of information about the candidates.*

*MediaLab and DirectRT are computer software programs designed for psychology experiments and used by political psychologists like Mondak et al. Information on this software can be found at http://www.empirisoft.com/medialab.aspx and http://www.empirisoft.com/DirectRT.aspx.*

**Procedures:** *First subjects took a survey about their political attitudes and attentiveness. Then subjects were presented with the information about the hypothetical candidates, either in paper form or via computer. All of the experiments varied the content of the information presented and subjects were randomly assigned to these manipulations. In all of the experiments, the researchers focused on manipulations of evaluations of the skill and integrity of the candidates. In Mondak1, the subjects were asked to give feeling thermometer-like ratings to the two candidates in each set and to identify which one would receive their vote. In Mondak2, the subjects were only asked which candidate they would vote for. And in Mondak3, subjects were asked whether they favored or opposed each candidate on an individual basis; that is, the candidates were presented not as pairs but singly on the computer screens. Also in Mondak3 the researchers measured the response time of subjects (how long it took for a subject to express his or her choice after seeing the information about a candidate).*

**Results:** *Mondak1 and Mondak2 found significant evidence that the qualities of the candidates affected the subjects' choices. They found this result was robust even when controlling for the importance of political issues for subjects and the distance between subjects' views on ideology and the candidates. Mondak3 found a similar effect but also found that subjects' attitudes toward candidates' competence and integrity were highly cognitively accessible*

*(as measured by the response time). But they found no evidence that candidate character serves as a default basis for evaluation when things like partisanship and ideology are unavailable.*

**Comments:** *In Mondak3, the researchers also reported on survey evidence that supports their conclusions.*

**Example 2.5 ("In-Your-Face" Discourse Lab Experiment):** *Mutz (2007) reported on laboratory experiments designed to evaluate the effects of televised political discourse on awareness of opposing perspectives and views of their legitimacy.*

**Target Population and Sample:** *At least 171 subjects were recruited from temporary employment agencies and community groups. Mutz did not report the community from which the subjects were drawn, although probably they were from the area around her university.[12]*

**Subject Compensation:** *Subjects from the temporary employment agencies received an hourly rate that depended on whether the subjects came to campus to participate in this particular experiment or a set of studies over several hours. The subjects from civic groups participated as a fund-raising activity for their organizations.*

**Environment:** *The experiments took place in a university facility where subjects were shown a 20-minute mock television program while sitting on a couch. The program was produced professionally with paid actors and a professional studio talkshow set was used to tape the program. The program was also professionally edited. The program was an informal political discussion between two "candidates" for an open congressional seat in a distant state, with a moderator who occasionally asked the candidates questions. Subjects were led to believe that the program and candidates were "real."*

*The candidates in the video had opposing views on eight different issues. The views drew on arguments from interest groups and the issues were topical at the time of the experiment. Four versions of the video were produced; two were civil versions and two were uncivil ones. In all four the same issue positions and arguments were expressed in the same words. As Mutz relates (p. 625):*

*The only departures from the script that were allowed for purposes of creating the variance in civility were nonverbal cues (such as rolling of the eyes) and phrases devoid of explicit political content (such as "You have completely missed the point here!"). The candidates in the uncivil condition also raised their voices and interrupted one another. In the civil version, the politicians spoke calmly throughout and were patient and respectful while the other person spoke.*

[12] Mutz did not report the number of subjects who participated in experiment 3, so this sum only includes the participants in experiments 1 and 2.

*Mutz did a manipulation check with pretest subjects who rated the candidates on measures of civility.*

*Mutz also manipulated the camera perspective. That is, in one of the civil versions and one of the uncivil versions there was an initial long camera shot that showed the set and location of the candidates and moderator, and then the subsequent shots were almost exclusively tight close-ups. In contrast, in the medium version the candidates' upper bodies were shown.*

**General Procedures:** *After giving consent to participate, subjects were seated on the couch and given a pretest questionnaire. They were then asked to watch the video program and informed that they would be asked some questions after the program was concluded. Afterward, a paper-and-pencil questionnaire was administered. Subjects only saw four issues discussed, which varied by experimental session. In the questionnaire, Mutz asked open-ended questions designed to measure the extent that subjects recalled the arguments and the legitimacy of the arguments. Mutz also asked subjects to rate the candidates on a feeling thermometer. Using these basic procedures, Mutz conducted three different experiments.*

**Experiment 1 Procedures:** *In the first experiment, which used 16 subjects, the subjects saw a discussion using all four different combinations of camera perspective and civility (the issues also varied). These subjects' arousal during the video was measured using skin conductance levels by attaching two electrodes to the palm of each subject's nondominant hand. According to Mutz (p. 626), "Data collection began at the start of each presentation, with a 10-second period of baseline data recorded while the screen was blank prior to the start of each debate."*

**Experiment 2 Procedures:** *The second experiment used 155 subjects and the subjects saw only one of the four possible experimental manipulations. Subjects were randomly assigned to manipulations. Also included was a group of subjects who were randomly assigned to watch a nonpolitical program for the same amount of time and received the same questionnaire.*

**Experiment 3 Procedures:** *Mutz did not report the number of subjects used in this experiment. Experiment 3 is a partial replication of experiment 2 with one exception: in this experiment, all subjects saw the close-up versions of the videos and were randomly assigned to either civil or uncivil discourse.*

**Results:** *In experiment 1, Mutz found that uncivil discourse was significantly more arousing than civil discourse and that the close-up camera perspective was also significantly more arousing than the medium perspective. In experiment 2, she found that the awareness of rationales for arguments was also affected by the manipulations in the same direction, and uncivil close-up conditions led to the most recall. Furthermore, she found that the difference in thermometer ratings between the subjects' preferred and nonpreferred candidates*

*was not affected by civility in the medium camera perspective. However, in the close-up camera condition, this difference was significantly greater in the uncivil condition. The effect worked in both directions; that is, in the civil close-up condition the difference in ratings fell and in the uncivil close-up condition the difference rose, in comparison to the medium camera condition. Mutz found a similar relationship in the perceived legitimacy of opposing arguments in both experiment 2 and experiment 3.*

**Comments:** *Mutz's experiments are a good example of how control can be exercised over unobservable variables in experiment 1, which we discuss further in the next two chapters. Experiment 3 is an interesting instance of a researcher replicating a previously found result, something we discuss in Chapter 7.*

**Example 2.6 (Swing Voter's Curse or SVC Lab Experiment):** *Battaglini et al. (2008, 2010) report on a series of experiments conducted to evaluate the predictions from the Swing Voter's Curse.*

**Target Population and Sample:** *Battaglini, Morton, and Palfrey recruited student volunteers at Princeton University (84 subjects) and New York University (NYU; 80 subjects) from existing subject pools which had been recruited across each campus. No subject participated in more than one session. The subject pool for the experiments had been recruited via email to sign up for experiments conducted at either the Princeton Laboratory for Experimental Social Sciences at Princeton or the Center for Experimental Social Sciences at NYU. One free online recruitment system is ORSEE (Online Recruitment System for Economic Experiments, devised by Ben Greiner at the University of New South Wales; see ttp://www.orsee.org/).*

*As is typical in political economy laboratory experiments, more subjects than the required number were recruited because the experiments were designed for specific numbers of participants. Subjects were chosen to participate on a first-come/first-serve basis, and subjects who arrived after the required number of participants had been met were given the show-up fee as payment.*

**Subject Compensation:** *Subjects were paid in cash based on their choices during the experiment as described in the procedures. Average earnings were approximately $20. In addition, subjects were also given a show-up fee of $10. Subjects were assigned experiment-specific identification (ID) numbers and payments were made to subjects by ID numbers such that records were not kept that could be used to match subject identity with payments received or choices in the experiment.*

**Environment:** *The experiments used a standard setup for computerized laboratory experiments by political economists. That is, the experiments were conducted in computer laboratories via computer terminals and all*

*communication between the experimenter and subjects was conducted via the computer interface. Each subject's computer screen was shielded from the view of other subjects in the room through privacy screens and dividers. Subjects were first presented with instructions about the experiment and then took a short quiz regarding the information in the instructions before they were allowed to continue to the experiment. Subjects were told all the parameters of the experiment as described in the procedures. The experimental parameters were chosen specifically to evaluate game-theoretic predictions and a formal model was used to derive these predictions explicitly. The software used for the experimental program was multistage, which is an open-source software program for laboratory experiments developed at the California Institute of Technology (see http://multistage.ssel.caltech.edu/). Of particular usefulness for experiments is the free software z-Tree for the Zurich Toolbox for Readymade Economic Experiments (see http://www.iew.uzh.ch/ztree/index.php), which was devloped by Urs Fischacher.*

**Procedures:** *Before the experiment began, one subject was randomly chosen to be a monitor. The monitor was paid a flat fee of $20 in addition to his or her show-up fee. Each session was divided into periods. In five of the sessions conducted at Princeton University, 14 subjects were randomly assigned to two groups of 7 voters. The group assignments were anonymous; that is, subjects did not know which of the other subjects were in their voting group. In two sessions at Princeton, only 7 subjects were in each session, so in each period the group of voters was the same. At NYU all subjects in each session were in the same group; three sessions used 21 subjects and one session used 17 subjects.*

*In each period and group, the monitor would throw a die to select one of two jars, red or yellow. Although subjects could see the monitor making the selection, they were not able to see the selection made. The red jar contained two red balls and six white balls, the yellow jar contained two yellow balls and six white balls. These were not physical jars, but were represented on the computer monitors. Subjects then were shown the jar with eight clear balls, that is, a jar without the colors. They then could click on one of the balls and the color of the ball selected would be revealed. If a red or yellow ball was revealed, they learned which jar had been chosen. If a white ball was revealed, they did not learn which jar had been chosen.*

*After choosing a ball and finding out its color, the group members simultaneously chose where to abstain, vote for red, or vote for yellow. The computer casts a set number of votes for the red jar. The jar that received the majority of the votes was declared the winner, including the computer votes (ties were broken randomly by the computer). If the jar chosen by the majority was the correct jar, all the subjects earned a payoff of 80 cents in the period, and if the*

jar chosen by the majority was the incorrect jar, all the subjects earned a payoff of 5 cents in the period.

Subjects were told the outcome of the period in their group. They were then randomly reassigned to new groups for the next period if applicable and the procedure was repeated. The colors of the balls within each jar were randomly shuffled each period so that if a subject repeatedly chose to click on the same ball, whether they were revealed a white ball was randomly determined by the percentage of white balls in the jar (i.e., the probability of observing a white ball was always 75%). A check of the data shows that the procedure worked as desired: approximately 75% of subjects saw a white ball and 25% saw either a red or yellow ball. There were a total of 30 periods in each session.

Each session was divided into three subsessions of ten periods each. In one of the subsessions, the computer had zero votes. In sessions with groups of 7 voters, in one subsession the computer had two votes and in one subsession the computer had four votes. In sessions with groups of 17 and 21 voters, in one subsession the computer had six votes and in one subsession the computer had twelve votes. The sequence of the subsessions varied by session. The following sequences were used depending on the number of voters in the groups: (0,2,4), (0,4,2), (2,4,0), (4,0,2), (4,2,0), (0,6,12), and (12,6,0).

Battaglini, Morton, and Palfrey also varied the probability by which the monitor would pick the red jar. In some sessions, the probability of a red jar being selected was equal to the probability of a yellow jar, which was equal to 1/2. This was done by having the monitor use a six-sided die; if 1, 2, or 3 were shown, the red jar was selected, and if 4, 5, or 6 were shown, the yellow jar was selected. In other sessions the probability of a red jar being selected was equal to 5/9 while the probability of a yellow jar was equal to 4/9. This was done by having the monitor use a ten-sided die; if 1, 2, 3, 4, or 5 were shown, the red jar was selected; if 6, 7, 8, or 9 were shown, the yellow jar was selected; and if 10 was shown, the die was tossed again.

**Results:** Battaglini, Morton, and Palfrey found that subjects who were revealed either a red or yellow ball voted for the red or yellow jar, respectively. They also found that when the number of computer votes was equal to zero, most of the subjects who were revealed a white ball abstained. As the number of computer voters increased, uninformed voters increased their probability of voting for the yellow jar. These results occurred even when the probability of the red jar was 5/9.

**Comments:** The results strongly support the SVC theoretical predictions. Uninformed voters are more likely to abstain than informed voters, but when there are partisans (as operationalized by the computer voters), the uninformed voters appear to vote to offset the partisans' votes, even when the probability

is higher that the true jar is the red jar (the partisans' favorite). Battaglini, Morton, and Palfrey also considered some alternative theoretical models to explain some of the subjects' errors, an analysis that we explore in Chapter 6.

**Example 2.7 (Polls and Information Lab Experiment):** Dasgupta and Williams (2002) reported on a laboratory experiment that investigates the hypothesis that uninformed voters can effectively use cues from public opinion polls as an information source to vote for the candidate they would choose if informed.

**Target Population and Sample:** Dasgupta and Williams recruited 119 undergraduate student volunteers at Michigan State University. The authors recruited subjects who were unaccustomed to psychological experiments and were unfamiliar with spatial voting models and formal decision theory.

**Subject Compensation:** Subjects were paid based on their choices as described in the procedures. As in many political economic experiments, the subjects' payoffs were denominated in an experimental currency that was then converted at the end of the experiment into cash. Dasgupta and Williams called this experimental currency "francs." We discuss reasons for using an experimental currency in Section 10.1.4; one reason for doing so in the Dasgupta and Williams experiment is that it allowed them to have different exchange rates for different types of subjects, as described later. The exchange rates were fixed and known by subjects. Subjects earned on average $22 for the two-and-a-half hours plus a show-up fee.

**Environment:** As in Example 2.6, the experiment was conducted via a computer network. Subjects were seated so that they were unable to see the computer monitors and choices of other subjects. The experimental parameters were chosen to fit the formal model presented by Dasgupta and Williams.

**Procedures:** Dasgupta and Williams conducted six experimental sessions with 17 subjects in each session. At the beginning of a session, two subjects were chosen to be incumbent candidates; the remaining subjects were assigned as voters. The sessions were divided into two subsessions lasting 10 periods each. The incumbent candidates were randomly assigned to separate subsessions and only participated in the experiment in the subsession to which they were assigned. At the beginning of each subsession, the incumbent candidate was assigned an issue position of either 0 or 1000, which was held fixed throughout the subsession. This issue position was publicly announced to all subjects. In each period in a subsession, the subjects were randomly divided into three equal-sized groups of 5 voters each with issue positions of 250, 500, and 750, respectively. So in each period, a subject's issue position was a new random draw. All subjects knew the distribution of issue positions of voters.

Table 2.1. *Voters' payoffs*

| Incumbent Quality | Incumbent's issue position = 0 | | | | | |
| | Voter issue positions | | | | | |
| | Incumbent wins | | | Challenger wins | | |
| | 250 | 500 | 750 | 250 | 500 | 750 |
| 10 | 40 | 10 | 10 | 25 | 10 | 20 |
| 20 | 45 | 45 | 25 | 30 | 45 | 40 |
| 30 | 50 | 50 | 50 | 35 | 50 | 60 |

*Before the subsession began, the incumbent candidate was provided with an initial endowment of 900 francs. The candidate chose an "effort" level equal to either 10 or 20. Effort levels were costly to the candidate; if he or she chose an effort of 10, the cost was 30 francs. The cost of an effort of 20 was 170 in one subsession and 90 in the other. The cost of the effort chosen by the candidate was deducted from his or her endowment. The computer program then assigned the candidate a "quality" of either 10, 20, or 30, with equal probability. Note that the incumbent candidate was not told his or her quality before choosing an effort level. The effort level and quality were combined to produce an "output." Three voters in each of the voting groups were randomly chosen to be told the output. The remaining two voters in each voting group received no information about the output and were uninformed.*

*Voters then participated in an election. They could either vote for the incumbent or the challenger (who was an artificial actor). If the incumbent won the election, he or she would receive 300 francs. The voters' payoffs depended on their own issue position, the issue position of the incumbent candidate, and the quality of the incumbent. Table 2.1 presents the payoffs to voters by issue position when the incumbent candidate's issue position equals 0 (the case where the incumbent candidate's issue position equals 1000).*

*Before the election was held, three separate opinion polls were taken. In each poll, voters were asked which of the two candidates – incumbent or challenger – he or she currently preferred. All subjects were revealed the aggregate outcomes of the polls. After the polls were completed, the election was held, which was decided by majority rule (ties were broken randomly by the computer). The incumbent candidate and the voters received their respective payoffs and a new period began with voters reassigned to new voter types.*

*Dasgupta and Williams also conducted an additional seventh session which was the same as the other sessions except that the candidate's issue position was held constant at 0 in both subsessions and all voters were uninformed.*

***Results:*** *Dasgupta and Williams found significant support for the theoretical argument that uninformed voters will use poll information to make more accurate voting decisions and that the electorate's behavior is consistent with complete information. Compared to the session in which no voters were informed, voters were much more likely to choose optimally. However, voters make errors when incumbent candidates choose suboptimally. Specifically, in the low-cost subsessions, according to the theory, the candidates should always choose a low level of effort and in the high-cost trials the candidates should choose a high level of effort. The majority of the time candidates chose effort levels as predicted, with errors occurring early in the subsessions. Voters' mistakes were most likely to occur when candidates did not choose effort levels as predicted.*

***Comments:*** *The experiments are an extension of earlier experiments and theoretical work reported by McKelvey and Ordeshook (1984, 1985a,b, 1986a,b), and Collier et al. (1987) on a rational expectations model of the relationship between voting and information. Dasgupta and Williams also compared the predictions of several alternative models in which uninformed voters are uninfluenced by polls and found support for the rational expectations model.*

**Example 2.8 (Decentralization Experience and Turnout Natural Experiment):** *Lassen (2005) presents a study of the effects of information on voter turnout in Copenhagen. He exploits a naturally occurring situation that affected voter information in a naturally occurring referendum and studies the effect of the information on turnout in the referendum.*

***Target Population and Sample:*** *Lassen used a telephone survey of voters in the city of Copenhagen. The voters in the city were partitioned into five strata reflecting five different areas of the city. In Denmark every citizen has an identification number. Everyone eligible to vote automatically receives a ballot card at their address in the census registry. The voter then presents the ballot card at their local polling station and if the individual votes, it is registered. The survey was commissioned by the city, so the sample was drawn from the city's list of eligible voters in the election. The response rate to the survey was 55%, resulting in a sample of 3,021 observations. Of this sample, one-third refused to answer questions about their yearly income or whether they had voted in the referendum, which left 2,026 observations. Lassen used the larger sample for some analysis as well as the smaller sample.*

***Environment:*** *In 1996 the city of Copenhagen decided to conduct a policy experiment in decentralization of city government by having some areas of the city experience decentralized services while other areas continued to experience centralized services. The policy experiment lasted for four years and in 2000 a*

*citywide consultatory referendum was held on whether the program should be extended to or abolished in the entire city.*[13]

***Procedures:*** *The city was divided into fifteen districts (divisions which did not exist prior to the policy experiment): eleven districts where government services continued to be centrally administered and four districts where approximately 80% of government services were administered by locally elected district councils, which are labeled pilot city districts or PCDs. The designers of the policy experiment attempted to choose four districts that were representative of the city. The strata for the survey were the four PCDs and the rest of the city.*

*The consultatory referendum was held on the same day as a nationwide referendum on whether Denmark should join the common European currency. It was possible for individuals to vote in only one of the referenda if they wished.*

*The survey asked respondents if they voted in the last municipal election, whether they voted in the nationwide referendum, whether they voted in the municipal referendum, their opinion of the decentralization experiment, their opinion of the responsiveness of municipal council members, how interested they are in political issues, and a set of demographic questions.*

***Results:*** *Lassen found that turnout was significantly higher among informed voters than in the districts where services remained centralized. The effect is robust to a number of different specifications and is strongest among those voters who had zero cost of voting; that is, those who voted in the nationwide referendum and so had paid the cost of going to the polling place.*

***Comments:*** *Lassen dealt with a wide variety of methodological issues in identifying and estimating the causal relationship between information and voting. We discuss his study more expansively as we explore these issues. As in the other experiments we have so far discussed, in Lassen's study the manipulated variable is not the same as the treatment variable. The manipulated variable is whether the respondent lived in the district that experienced the decentralization policy experiment, whereas the treatment variable is the information level of the respondents. As a proxy for the treatment variable, Lassen uses whether a respondent reported an opinion of the decentralization experiment, classifying those with no opinions as uninformed. He cites empirical evidence that finds a strong correlation between other measures of voter information and the willingness to express an opinion. Lassen also finds, however, that this treatment variable is endogenous and affected by other observable variables from the survey. We discuss how Lassen deals with this endogeneity in identifying and estimating the causal effects of information on turnout in Chapter 5.*

---

[13] The referendum was nonbinding because the Danish constitution does not allow for binding referenda at the municipal level.