

Donald P. Green

*on evidence-based
political science*

The advent of the evidence-based movement in fields as varied as medicine, criminology, and education represents not simply a new thirst for evidence, but for evidence of a particular kind. Far from issuing banal appeals for data, advocates of evidence-based practice emphasize the need for experimental

Donald P. Green, a Fellow of the American Academy since 2003, is A. Whitney Griswold Professor of Political Science and director of the Institution for Social and Policy Studies at Yale University. His scholarship covers a broad array of topics, such as campaign finance, party identification, rational action, voter turnout, and prejudice. He is the coauthor of three books: "Pathologies of Rational Choice Theory: A Critique of Applications in Political Science" (1994), "Partisan Hearts and Minds: Political Parties and the Social Identities of Voters" (2002), and "Get Out the Vote! How to Increase Voter Turnout" (2004). Outside of political science, he is known for his psychological research on the measurement of mood and his sociological research on the causes of racially motivated crime.

© 2005 by the American Academy of Arts & Sciences

research conducted in real-world settings.

The term 'experimental' in this context refers to studies in which units of observation are assigned at random to treatment and control conditions. This type of investigation stands in sharp contrast to observational research in which some natural process determines which individuals or groups receive a treatment. In a medical experiment on hormone replacement therapy, for example, women are randomly assigned to receive either the treatment or a placebo; the corresponding observational study simply compares women who, for whatever reason, receive hormone replacement therapy to those who do not, often with statistical controls designed to make the two groups equivalent.

The strength of experimental research derives from the use of randomization procedures, which ensure an unbiased comparison between treatment and control groups. 'Unbiased' is a term of art that statisticians use to refer to estimation procedures that have no systematic tendency to over- or underestimate the true effect. Any given randomized study might over- or underestimate the effects of, say, hormone replacement therapy, but on average these errors will cancel out. Observational research, by contrast, forces the investigator to impose strong and often untestable assumptions about the comparability of groups that do and do not receive a treatment. If healthier women happen to take hormone replacements, there will be a systematic tendency to overestimate the treatment's effects.

Although political science is predominantly an observational discipline, it has gradually drifted in the direction of experimental research. The 1950s saw some initial forays into randomized experiments on voter turnout. The 1960s

and 1970s drew political scientists into the study of group bargaining and public goods dilemmas, usually with college students standing in for legislators. The 1980s witnessed a dramatic increase in the use of randomized survey design, whereby respondents answered questions that varied in wording and content. The late 1990s ushered in the current renaissance of field experimentation when a small but rapidly growing group of scholars rekindled the experimental study of voter mobilization, conducting dozens of field experiments designed to examine the effects of door-to-door canvassing, direct mail, phone calls, leaflets, electronic mail, and televised public service announcements.

These studies, which Alan Gerber and I summarize in *Get Out the Vote! How to Increase Voter Turnout*, refocus the method and substance of research in the field of electoral behavior. With regard to method, these studies randomly assigned hundreds, thousands, and, in some cases, tens of thousands of registered voters to treatment and control groups. Those assigned to the treatment group received some kind of intervention designed to encourage them to vote. After each election, public records were examined to gauge voter turnout rates in the treatment and control groups.

Note the contrast with conventional survey analysis, which in this instance assesses the effects of phone calls by asking respondents whether they voted and whether they received some sort of contact from a campaign. Obviously, surveys of this kind confront serious misreporting problems. But even if reporting were flawless, the problem of drawing causal inferences from nonexperimental data would remain: do those contacted by campaigns vote at higher rates because they were contacted, or are campaigns simply prone to target likely vot-

ers? If the latter were true, we might see a correlation between voting and receiving phone calls even if contact from a campaign had no effect on voter turnout. Of course, the survey analyst could assume away this problem by stipulating that those who are contacted by campaigns are just like those who are not contacted. Indeed, these kinds of assumptions are so routinely invoked that researchers are often oblivious to them. The point of experimental research is to call attention to these assumptions. The challenge is to devise experimental designs that free researchers from invoking them.

Three healthy developments flow from the exercise of confronting causal questions with experimental data. Ordinarily, bold pronouncements about research methodology and statistical analysis are insulated from criticism by the fact that the causal parameters of interest are unknown. Indeed, one cannot understand the intellectual currents in political science without appreciating the role of chronic causal uncertainty.

First, in the absence of an accurate causal inference or the prospect of getting one, methodological disputes are often a matter of style, with the advantage typically going to the most technically sophisticated statistical procedure. But experiments change the terms of these methodological debates. In recent decades, researchers in economics and medicine have begun to evaluate the performance of observational research methods by asking how closely their results concur with experimental benchmarks. The results have been quite sobering. Observational methods often perform poorly, even when the data are analyzed using cutting-edge statistical techniques.

Second, the fact that observational methods are sometimes shown to pro-

Evidence-based political science

duce wildly inaccurate conclusions engenders a healthy skepticism about canonical research findings that have not been confronted with experimental data. Consider, for example, the relationship between education and voter turnout in the United States. Thousands of surveys spanning several decades have documented the powerful correlation between educational attainment and electoral participation: no two variables in social science have a more robust statistical relationship. On the other hand, the implications of this individual-level result seem to be at odds with historical trends. The fact that many Western industrialized democracies have experienced dramatic long-term gains in education but no gains in voter turnout raises questions about whether education exerts a causal effect on voter turnout or is instead a marker for other factors that are the true causes (for example, exposure to political discussion during childhood). The solution to this puzzle, which has vexed political scientists for decades, is to study the long-term consequences of exogenous interventions that have increased educational attainment—for instance, do randomly assigned college scholarships or reductions in class size produce higher rates of voter turnout? This approach represents a radical departure from current practice in political science.

Third, as the foregoing example illustrates, the experimental perspective encourages political scientists to attend to experimental developments in other disciplines. When education researchers devise an experiment that produces a significant increase in high school graduation rates, the stage is set for what Alan Gerber and I have termed a ‘downstream experiment.’ Once education is manipulated in a random way, the task for political scientists is to trace the

downstream effects on voter turnout, support for civil liberties, and other outcomes thought to be driven by education. Analogous arguments could be advanced for economic experiments: when individuals experience a rise in income, do they change their political orientations? Or for sociological experiments: when people are randomly recruited to participate in environmental or cultural organizations, do their levels of social trust and civic engagement increase?

Political scientists whose subject matter falls outside the field of political behavior tend to regard experimentation as impractical or unethical, or both. Few researchers are comfortable with the idea of randomly altering constitutional arrangements, alliances, or political cultures. This is usually where the discussion of experimental methods ends. But experimental thinking can be useful even if researchers can do no more than approximate the features of an experiment. Unfortunately, researchers seldom avail themselves of these opportunities. Those who conduct comparative research, whether qualitative or quantitative, rarely design their investigations around near-random events, such as technological or climatic developments that in the short run change trade flows, military capabilities, or mass access to new information.

Admittedly, many questions in political science do not lend themselves to experimentation. Practical and ethical constraints provide a justification for observational methods. However, those who are forced by circumstances to rely on nonexperimental evidence should not lose sight of its inherent limitations. To bring these limitations into sharper focus, Alan Gerber, Edward Kaplan, and I have spelled them out in something we dubbed the Illusion of Learning Theorem. Put simply, the argument runs as

follows: Suppose you are confronted with two kinds of evidence – experimental and observational. Thanks to random assignment, you can extract information about causality from the experimental data (at least for the narrow setting in which the experiment takes place); the larger the experimental study, the less uncertainty that surrounds this causal inference. The observational data present two sorts of uncertainty. Sampling error, one type of uncertainty, diminishes with the size of the study. A second source of uncertainty concerns bias – the tendency of research method to over- or underestimate a causal relationship. Even when the sample size is infinite, uncertainty about the bias associated with the observational research design remains. The weight you assign to observational evidence hinges on the second source of uncertainty. When you know nothing about the bias of the observational research procedure, you simply ignore the observational results entirely.

The logic of this argument presents an important challenge to those who claim that observational research has advanced our knowledge of cause and effect in political science. Those who make this claim are implicitly insisting they know the inherent biases in the nonexperimental methods that are routinely used. (They are encouraged in this view by reporting conventions, which present statistical results as though these biases were known with perfect certainty.) For most applications, one might reasonably ask how researchers came to know the biases of their nonexperimental approach. I am aware of no research program in political science that endeavors to assess whether scholars can successfully predict the biases of different types of research designs. The history of medicine is replete with examples of

therapies that were lauded on the basis of observational research only to be repudiated by randomized trials – for example, hormone replacement therapy. The subtle biases of observational research often become evident only in hindsight.

A more persuasive defense of observational research notes that the questions it addresses are often bigger than those that lend themselves to experimental inquiry. If we imagine that the expected value of a research program is the product of the importance of the research question times the increase in knowledge that results from the investigation, we may conclude that some observational studies are probably good investments. The question, then, is whether the research portfolio of political science is appropriately diversified. Although experimental methods have made inroads in recent years, the overwhelming majority of research in the discipline remains nonexperimental. Political science is arguably too enamored of long shots.

One final concern about allocating more research effort to experimental investigation is that its narrow empirical focus comes at the expense of broader theoretical inquiry. Knowing whether this or that voter mobilization technique raises turnout, so the argument goes, does not tell us much about the broader conditions under which people engage in collective action. My colleague Alan Gerber has argued forcefully against this proposition, pointing out that the value of rigorous science is that it provides a firm foundation on which theories can be erected. The gradual accumulation of secure causal propositions aids theory building. Theories of collective action alone may or may not predict that door-to-door canvassing stimulates voter turnout while a torrent of direct mail

*Evidence-based
political
science*

*Note by
Donald P.
Green*

and robotic phone calls does not. But once these facts are established, theorists know not to bother advancing arguments that imply that door-to-door canvassing is a waste of time because it fails to resolve the public goods dilemma, or that the content of mail and phone messages allows voters to overcome the costs of acquiring political information. Experiments provide the stubborn facts that inspire theoretical innovation, which in turn suggests new lines of empirical inquiry. To Kurt Lewin's aphorism that "there is nothing as practical as a good theory," evidence-based political science would add that there is nothing as theoretically informative as a reliable causal inference.