Good Theories Deserve Good Data

Frank R. Baumgartner, *Texas A&M University*
Beth L. Leech, *Texas A&M University*

At base, the disagreement that sets us apart from Austen-Smith and Wright has to do with the relative importance that we ascribe to observation and empirical demonstration as opposed to model building. In our view, the construction of rigorous empirical tests of a model should be seen as an integral part of the process of developing new theories of politics, not as secondary to the more important task of building new models. The strategy followed by Austen-Smith and Wright is to construct a formal model and later to search for a data set that might allow for its testing. By focusing on the formal aspects of the research project rather than by integrating the theoretical and empirical aspects, researchers often are led to use data sets collected for purposes that do not correspond closely to the theory they hope to test. They dilute the power of their theories by settling for less than convincing evidence for them.

In our comment on Austen-Smith and Wright's work, we raised several problems with their theory, but we focused mostly on the strength and appropriateness of the evidence. We focused on one article because we saw it as an archetype of an increasingly common approach that denies equal weight to the dual goals of developing new theories and constructing rigorous empirical tests. Indeed it is the commonality of the problems we pointed to, not their uniqueness, that motivated our comment and justifies this discussion.

In adopting an existing data set to test a new theory, one should be aware that if the theory is truly innovative, pre-existing data sets are unlikely to include all the relevant variables and to measure them appropriately. Further, these data sets may or may not be at the appropriate unit of analysis or use the best research design. The use of existing data sets too often leads to acceptance of measurements that are not exactly what the theory requires, omission of theoretically endogenous variables from the empirical test, and inclusion of theoretically exogenous variables as ill-justified measurement controls. Many of the problems that we reviewed are common components of a research strategy that gives first priority to the construction of an elegant and mathematically coherent formal model, but relegates the testing of that model to a low priority.

Editor's note: To provide closure on this debate, Professors Austen-Smith and Wright were not asked to respond to this manuscript.

American Journal of Political Science, Vol. 40, No. 2, May 1996, Pp. 565–569

The approach to empirical testing adopted by Austen-Smith and Wright is made clear in their response when they assert that a substantive interpretation of their results is impossible because of the nature of their dependent variables. They write that "the substantive implication of an increase . . . in the values of these variables is unclear." We believe that the substantive interpretation of these values is clear. More importantly, however, we believe that the substantive interpretation of empirical results is a central component of any research project. To design a study and reach conclusions on the basis of data that cannot be interpreted indicates a lack of concern with the empirical component of the research project and an acceptance of only a loose fit between theory and data. Substantive conclusions require substantive evidence.

The apparent substantive ambiguity of the empirical results does little to stop the authors from interpreting their results as if they were very important indeed; in fact it seems to give them greater freedom. "Given the evidence of counteractive lobbying, it follows that legislators do sometimes change their positions on the basis of interest group lobbying. Although we did not conduct a direct test of the influence of lobbying . . . , our findings nevertheless are consistent with the notion that groups are influential." On the basis of this logic, they assert that the literature on interest-group lobbying activities is not contradictory, but rather that many studies conclude that lobbying often matters. Taken in these terms, of course, it is certainly apparent that a consensus exists that groups sometimes matter. But which groups, when, under what circumstances, by using what strategies, on what kinds of issues, and with what results?

If one considers it important that theory, data, and conclusions be tightly integrated, then the choice of the dependent variable is fundamental. The use of data that one believes to be immune from substantive interpretation is so limiting as to preclude generalization. It breaks the link between theory and conclusions. Austen-Smith and Wright clearly disagree with us about the importance of this link, since they draw broad generalizations despite their critical view of their own evidence.

Another area of disagreement concerns the importance of the choice of a research design. The comparative statics approach to testing the predictions of many formal models in political science is so prevalent that it merits considerable reflection, and we hope that this exchange adds to that discussion. Like any research design, tests using cross-sectional data have strengths and weaknesses. Some problems that we pointed out are avoidable, such as the failure to include all theoretically relevant variables and to exclude all others. Other drawbacks, however, are not mistakes or errors on the part of the authors, but are unavoidable consequences inherent in the nature of this research approach. While cross-sectional tests are clearly appropriate in some circumstances, the approach also has flaws, especially when applied to a single case.

The most important drawbacks to the design adopted by the authors have to do with generalizability and replication. Despite the use of impressive statistical techniques, the evidence remains tied to a single case of policymaking. Comparative static tests need not be limited to the analysis of a single case, but in practice this is the predominant research strategy. To the extent that variables held constant by the choice of a single case are of theoretical importance, generalization becomes hazardous. In spite of the discipline's hopes that the accumulation of case studies over the years might lead to a truly cumulative science, experience shows the difficulties of comparing the results of one such study with those of the next. To argue whether the authors can generalize best from the particular case that they chose or whether a more consensual case might have been more typical is to miss the point that alternate designs might have avoided this dilemma in the first place.

The one-case-study-at-a-time approach to theory building makes it inherently difficult to replicate a study and to test for rival hypotheses if these data were not collected in the original research effort. In reaction to charges that evidence is unconvincing because of the omission of potentially important variables or the mis-measurement of others, authors can argue that critics have no proof. In response to evidence drawn from another case, one could argue that the contexts differed. The authors adopt and defend a research strategy that, by its very design, inhibits the direct comparison of alternative theoretical models.

The comparative static approach is indirect in that it relies on evidence about outcomes to draw inferences about processes. The approach is valuable in many instances; however it may also mislead. Often, several distinct process models are consistent with a single set of outcomes. Such a situation can be resolved in two ways: either generate and test rival hypotheses about outcomes, or gather direct evidence about the process. Our original paper pointed out that Austen-Smith and Wright do neither.

We also pointed to the question of robustness. Because the approach involves no direct observation of the process, and yet seeks to reach conclusions about the process, the robustness of the underlying model is especially important. If the model is not robust, then slight changes in the assumptions or operationalizations may dramatically alter the predicted outcomes; similarly, small changes in observed outcomes may be taken as evidence for dramatically different underlying processes. This is what we found when we compared the particular definition of counteractive lobbying in the Austen-Smith and Wright 1994 article with that used by Wright in his article in 1990. The reversal in conclusions about the process stemming from subtle changes in operationalization raises questions about the robustness of either model. The contradictions in the literature that we pointed out stem largely from the instability of predictions from incomplete models. Were the models more robust, contradictions would be fewer and confidence in the findings would be greater.

A further complication from the use of cross-sectional data to draw inferences about processes concerns ambiguities of language. Austen-Smith and Wright have in mind a particular definition of the term "counteractive" as it relates to their model, as we noted in our comment. While the common usage of the term implies a time-ordering, their particular usage does not. They are careful to point out that information about the sequence of behavior is absent from their model, as both we and they have emphasized. Our disagreement in this area does not stem from difficulties in interpreting equilibrium results, but from their willingness to move from a finding of support for their particular definition of "counteractive" lobbying to an assertion in their discussion that they have found support for the common definition of the term.

In view of these problems in the authors' use of the comparative static design, we concluded that direct evidence about the process of lobbying would be more convincing. Our concern is not with the comparative static design per se, but with its specific application here and with its adoption as the single dominant research approach in testing game-theoretic propositions, even about dynamic situations. In our view, the power of new theories of politics would be greatly enhanced if scholars adopted a broader range of research strategies and recognized the importance not only of the theoretical work, but also of the difficult empirical work. This would certainly lead to more direct (and more convincing) tests of such ideas as counteractive behavior.

Careful attention to the construction of a complex and highly formalized model of a small part of the political process is worthwhile. But to the extent that the resulting theory is really innovative, and to the extent that one hopes to convince others of the merits of the new theory, certainly one should be willing to make the effort to collect the best evidence. In our view, the research process is only at its beginning after the development of a theory and the generation of logical empirical hypotheses. The difficult part in showing the validity of a theory comes in choosing the most appropriate research designs, in devising measurements, and in gathering the data. These empirical endeavors are not mere add-ons to the interesting theoretical work: a research project is much more likely to fail or to mislead because of problems in measurement and in data collection than because of problems in the generation of hypotheses. The deemphasis of these empirical questions is where we differ most strongly from those we criticize.

Researchers proud of their theoretical propositions should realize that their influence will be much greater if they combine theoretical innovations with well crafted and accurate data collection efforts—time-consuming, difficult, and frustrating as these may be. A massive effort in generating new theories of politics deserves at least as much effort in generating convincing tests of them.

Final manuscript received 24 August 1995.