Developing Public Policy Theory: Perspectives from Empirical Research*

GEORGE D. GREENBERG
Office of the Assistant Secretary for Planning and Evaluation; U.S. Department of Health, Education, and Welfare

JEFFREY A. MILLER
University of Wisconsin, Madison

LAWRENCE B. MOHR
The University of Michigan

BRUCE C. VLADECK
Columbia University

There has been considerable interest in the development of theories of public policy formation, but theoretical efforts to date have not demonstrated adequate recognition of the distinctive qualities of the dependent variable as a focus of research. Facets of public policy are far more difficult to study systematically than most other phenomena investigated empirically by political scientists. Our attempt to test hypotheses with some rigor demonstrated that public policy becomes troublesome as a research focus because of inherent complexity–specifically because of the temporal nature of the process, the multiplicity of participants and of policy provisions, and the contingent nature of theoretical effects. We use examples of policy making taken from the case study literature to show concretely how such complexity makes it essentially impossible to test apparently significant hypotheses as they are presented by Lowi, Dahl, Banfield, and others. Our effort here is to enhance theoretical development by carefully specifying and clarifying the major shortcomings and pointing out the apparent directions of remedy.

The emerging discipline of public policy studies is characterized by a growing disjunction between theory and research. While there are many provocative and potentially important theories, systematic empirical research to test them has largely been lacking.[1] Most importantly, there seems to have been little progress in the critical intermediate stage of refining and then operationalizing the important variables.

In the course of our own research–an attempt to employ aggregated case-study data in order to test quantitatively a number of important public policy hypotheses[2]–we experienced enormous difficulty at this intermediate stage. We attempted to translate leading hypotheses into operational variables suitable for employment in quantitative analysis. The difficulty we experienced did not lie in the question of whether the hypotheses were supported or not, but in whether they were testable or not. Although the theories seemed perfectly applicable to the few cases used by their authors to illustrate them originally, the propositions did not fit so neatly when applied to a number of examples not expressly chosen for explanation and illustration. As we attempted to deal with problems of operationalization in connection with a variety of theories and dozens of cases, we noted that many of the problems recurred. While the theories were plausible, they seemed to fall into certain traps that appear to be endemic to the systematic study of public policy. This led us to adopt the purpose of the present paper–to comment on

*The authors gratefully acknowledge support from the Rackham School of Graduate Studies, The University of Michigan, and the assistance of Marc Holzer and an anonymous referee.

[1]By "systematic empirical research" we mean the use of quantitative data gathered from a number of instances of policy making in order to test specific hypotheses. Examples of recent work in this direction are provided by Robert K. Yin and Douglas Yates, *Street-Level Governments*, R–1527–NSF (Santa Monica: Rand, 1974); William A. Lucas, *The Case Survey Method*, R–1515–RC (Santa Monica: Rand, 1974); and Robert K. Yin and Karen A. Heald, "Using the Case Survey Method to Analyze Policy Studies," *Administrative Science Quarterly*, 20 (September 1975), 371–81.

[2]For a description of that project, see our "Case Study Aggregation and Policy Theory," *Proceedings*, 1973 Annual Meeting of the American Political Science Association, New Orleans, La., September 4–8, 1973.

these peculiarities of public policy theory and describe these pitfalls, as an avenue toward both building better theory and illuminating some of the significant characteristics of public policy itself.

The plausible and provocative theories of public policy formation offered by Lowi, Banfield, Froman, Gamson, Dahl, and others, appear in general to be no more or less adequate than those in any area of political science. *But the kinds of phenomena that the theories seek to explain* are radically different from most of the other phenomena we study. It is this difference in the objects of analysis that is the source of the difficulty.

In brief, public policy as a focus of systematic comparative analysis is *more complex* than such phenomena as electoral votes, legislative roll calls, incidents of political violence, and elite ideologies. It is more complex on at least four counts, which we will elaborate and illustrate in the discussion to follow. These are:

1. The policy process takes place over time, sometimes over a long period of time. This leads to difficulty in explaining "the process" as a simple unit. Even if one attempts to explain specific outcomes, the explanatory forces invoked almost invariably involve characteristics of this long and shifting process. Two sorts of difficulty arise:

a.) As the process proceeds over time, it can involve a large number of decision points, e.g., the decision of a subcommittee chairman, a Senate roll call, a presidential compromise, and the decision of an appellate court. The contents of each of these outputs might be called "public policy" and might be predictable by public policy theory. But we do not want theories to be oriented toward or tested upon inconclusive or tentative decisions. Nor do we want them constructed so as to predict the characteristics of the rubber-stamping process. We want somehow to focus only on "significant" outputs.

b.) The idea of a predictive theory of public policy demands that the values of the predictors be determined at some beginning point. Such values, however, are likely to change with the unfolding of the process itself, their final status being achieved only at its termination. Many presumably predictive theories are thereby weakened substantially, and become, in final analysis, post hoc explanations.

2. Any given policy proposal, or "output," or "outcome"[3] is in itself complex; it may have several important aspects. This multiplicity can make the whole policy extremely difficult to place in any single category, as is demanded, for example, by the categorization schemes that currently abound in public policy theory.

3. As a focus of analysis, policy making is complicated by the presence of a large number of participants. When a characteristic of the participants becomes a variable of interest, as it often does, variation among participants with regard to that characteristic causes difficulty. The difficulty takes two forms:

a.) Subjective. The state of the world as perceived by participants yields many important policy-analytic variables. But perceptions vary considerably, of course, depending upon the participant consulted or described.

b.) Objective. Still more variables are generated in existing theory by "objectively" determined participant characteristics—as determined, that is, by the researcher, interviewer, casewriter, or other outside observer. Ambiguity is introduced when the heterogeneous group of all participants, or heterogeneous subcollections of participants, must be assigned a single score on such a characteristic (e.g., level of involvement, or point of access to decision makers).

4. Lastly, public policy as a research focus is complex because the process cannot be described by simple additive models. On the contrary, the forces interact; the impact of one depends in large measure upon the value of another.

In sum, "public policy" is almost never a single, discrete, unitary phenomenon. Indeed, the appeal of public policy studies as a focus of intellectual endeavor lies precisely in its richness; the complexity of the unit of analysis simply and appropriately reflects the fact that an action of government is rarely meaningful if conceived of as a discrete, disembodied event, and that the impacts of a single government action on society are not understood properly if taken in isolation from one another.

In this essay, then, we seek to specify in some detail the kinds of difficulties these characteristics create, illustrating them by exploring the problems arising from attempts to apply specific theoretical propositions to concrete events taken from the case-study literature. We hope that our conclusions will not only provide some assistance to developers of

[3]Our use of the terms "output" and "outcome" follows Ranney's definition of those terms. See his "The Study of Policy Content," in *Political Science and Public Policy*, ed. Austin Ranney (Chicago: Markheim, 1968), pp. 8–9.

public policy theory who are concerned to avoid problems of conceptualization and operational definition, but will also highlight certain important characterisitcs of public policy itself that have heretofore been inadequately understood or explained.

An Initial Illustration

Many of the problems arising from the complexity of using "policy" as a unit of analysis can be illustrated by the work of Theodore Lowi. No single theoretical construct has been more important to the development of public policy studies than Lowi's categorization scheme,[4] yet the way in which he defines his fundamental terms is seriously weakened by the problems inherent in the unit of analysis, which make his hypotheses almost impossible to operationalize meaningfully in order to test them empirically. Since these problems of operationalization bear upon several of the challenges inherent in developing public policy theory, and since Lowi's idea is so well known and so well received, we refer to his work by way of general introduction, to be recalled briefly at several points in the more schematic discussion to follow.

The heart of Lowi's argument is that "policies determine politics." He developed, in a series of articles, a typology containing "regulatory," "distributive," "redistributive," and (subsequently) "constituency" policies and argued that policy processes will differ significantly depending on the policy type involved. As a result of those differences in process, he suggests, relationships among important concepts or variables that may be quite strong when policies of one kind are involved may be much weaker or even totally absent when other types of policies are concerned. Instead of attempting to find relationships that hold across the entire range of public policy, he argues that one should focus one's investigation within one of the four policy classifications.

Lowi generates a number of important hypotheses from his typology. For example, he predicts that congressional committees will be able to retain control of the process of coalition building on distributive amendments and that, therefore, few amendments will be offered to committee bills on the floor.[5] He also predicts that peak associations can be expected to be more cohesive when confronted with redistributive issues that unite their memberships than when they are faced with regulatory issues that divide them.[6]

The explicitness of the predictions Lowi derives from his policy classification is not matched, however, by a similar explicitness in explaining how one can determine the correct policy type for a given policy. Since the correct determination of policy type is obviously central to the predictive ability of Lowi's theory, this limitation has crucial importance. Lowi contends that distributive decisions are made without reference to their implications for other decisions; that regulatory decisions imply a direct choice "as to who will be indulged and who will be deprived"; and that redistributive decisions involve the greatest interconnection, since they imply choices among broad classes of individuals.[7] But the "decisions" of which Lowi is speaking are not simple and discrete; they are *policy* decisions. Many can be expected to contain at least some characteristics of each of Lowi's types, and there is little guidance for determining how a policy is to be classified in any but the simplest cases.

The problem of classifying a policy correctly thus becomes a central stumbling block to the empirical testing of Lowi's hypotheses. For instance, policies often have distributive programmatic characteristics but redistributive or regulatory financing mechanisms, as the case of the Chicago Transit Authority aptly illustrates. The CTA approached the state of Illinois for subsidization—a handout, precisely in the distributive tradition of the pork barrel and the pre-1960s tariff. The matter might easily have been processed and settled as a distributive issue, but the method for financing the subsidy became controversial. The proposal was even-

[4] We refer, of course, to the ideas presented in the series of three articles: "American Business, Public Policy, Case-Studies and Political Theory," *World Politics,* 16 (July 1964), 677–715; "Decision Making vs. Policy Making: Toward an Antidote for Technocracy," *Public Administration Review,* 30 (May-June 1970), 314–25; "Four Systems of Policy, Politics, and Choice," *Public Administration Review,* 32 (July-August 1972), 298–310.

[5] This is why Lowi found the '60s tariff to be so different. In earlier times, tariff policy was distributive, but "The true nature of tariff in the 1960s emerges as regulatory policy. . . . Issues that could not be thrashed out through the 'group process' also could not be thrashed out in committee but had to pass on to Congress and the floor," Lowi, "American Business," p. 701. Lowi presented data on amendments to bills in Congress from the floor in "Decision Making vs. Policy Making," pp. 321–22.

[6] "If there is ever any cohesion within the peak associations, it occurs on redistributive issues, and their rhetoric suggests that they occupy themselves most of the time with these" (Lowi, "American Business," p. 707).

[7] Ibid., pp. 690–91.

tually defeated, largely because of the dispute on regulatory or redistributive concerns—the question of whose particular pockets the money was to come from.[8] More generally, for *any* policy lacking explicit income transfers, there is some financing method either built in or implied that may or may not fall into the same category (in Lowi's typology) as the policy's other characteristics.

James Q. Wilson discusses this difficulty at some length. He maintains that a single policy may have aspects of each policy type, and he offers several examples. "A bill barring discrimination in public accommodations," for instance:

> could be seen as a measure regulating the use of hotels and restaurants or as one redistributing a benefit (access to hotels and restaurants) from one social stratum to another. . . . Urban renewal programs regulate the use of land, redistribute the housing supply, and distribute benefits to certain contractors and labor unions. Monetary and fiscal policy has both regulatory and redistributionist implications depending on whether one thinks of it as simply controlling the interest rate or as benefiting creditors at the expense of debtors (or vice versa).[9]

Lowi never specifically addresses the problem of how to classify a policy correctly when it has attributes of more than one of his policy types, but one can infer from the general tone of his articles that he does not expect classification to present a major problem. His basic argument in this respect appears to be that expectations based on past experience with similar issues objectively structure policy choices for the entire policy process. Thus he might well respond to Wilson by contending that one particular aspect of any of the policies Wilson cites can be expected to dominate the expectations of the actors in the process, so that there will be no problem in making the classification. Hotel and restaurant accommodations policy, for example, would undoubtedly evoke regulatory expectations from virtually all those connected with the policy process, based on their past experience and the past debate on similar issues.

That answer, however, is adequate only when expectations are virtually unanimous. Yet the policy process is as likely to be characterized by multiple perceptions as policies are likely to be characterized by multiple attributes. Whenever there are any significant disagreements among perceptions of a policy, the problem in classification according to Lowi's typology has simply shifted from one of determining "which aspect" to one of determining "whose perceptions." In the CTA example, it would be quite difficult to decide whose views should predominate—those concerned with mode of financing or those concerned with the costs and benefits of the transit service. Moreover, as Wilson notes, the problem becomes even more difficult in the cases of new or innovative policy, where there is little relevant past experience to structure the perceptions of any of the participants.[10]

Variations in perception of a policy are especially likely to occur when participants in the process actively seek to redefine the issue. Lowi himself suggests that "one of the important strategies in any controversial issue is to attempt to define it in redistributive terms in order to broaden the base of opposition or support."[11] In classifying a policy in which this strategy has been employed, Lowi's method will work only if the strategy has been almost a complete success (almost all perceptions have been redefined) or almost a complete failure (almost no perceptions have been redefined). If the strategy succeeds in altering some perceptions but not all, one has no guidance for deciding which perceptions will provide the basis for classification.

Take the controversy over Medicare.[12] Initially, since the proposal called for an additional tax on workers to provide benefits for the elderly, one would probably classify the policy as redistributive. The AMA, however, attempted to redefine the issue as one of government regulation of physicians and the practice of medicine. The AMA's argument undoubtedly influenced the perceptions of at least some individuals on the issue, but it certainly did not succeed in restructuring the perceptions of all. There is simply no objective way to determine which set of perceptions should be dominant in classifying a policy issue when there is substantial disagreement among the participants themselves about what is at stake.

[8]Edward C. Banfield, "The Chicago Transit Authority," in *Political Influence* (New York: Free Press, 1961), pp. 91–125.

[9]James Q. Wilson, *Political Organizations* (New York: Basic Books, 1973), p. 329.

[10]Ibid., p. 339.

[11]Lowi, "American Business," p. 707, footnote 28.

[12]Theodore R. Marmor, "The Congress: Medicare Politics and Policy," in *American Political Institutions and Public Policy*, ed. Allan P. Sindler (Boston: Little, Brown, 1969), pp. 3–68.

The problems in operationalizing Lowi's hypotheses illustrate the special difficulties of theory construction and testing in the field of public policy studies. His provocative ideas cannot be meaningfully operationalized without considerable effort by the researcher to add greater specificity and precision, a process in which the researcher must often, without adequate guidance, make important assumptions about what the theory is really trying to say. As it stands, Lowi's theory is not testable because the basic concepts are not operationalizable; in order to operationalize them, researchers must make a number of guesses and assumptions that create a situation where they can no longer be sure just whose theory they are testing.

The discussion of Lowi's paradigm illustrates the obstacles one encounters in attempting to apply policy theory across many cases, each of which is a full-scale instance of the operation of a policy-making system. We will now consider somewhat more explicitly the forms of complexity enumerated earlier, after which we will conclude with a brief listing of implications.

Problems of Temporality

Multiple Outputs. Several important hypotheses become mired in ambiguity for want of a time-stopping criterion. A whole host of variables is difficult to operationalize without being arbitrary until some decision is reached as to when, over a protracted period of development and struggle, "policy is made." For example, many of the most important hypotheses attempt to predict, as a dependent variable, the success or effectiveness of employing various resources or strategies in influencing outputs. Both Gamson and Dahl have discussed this problem at length, and both suggest measuring success by comparing the output of the process with the intentions of the relevant actors.[13] The output, however, keeps changing! There are simple cases, to be sure, such as those studied in Gamson's research on fluoridation, but there are complex cases as well. In fluoridation, points of beginning and ending were generally clear; someone at some time initiated the controversy with a proposal to fluoridate the water and policy was "made" when a community voted. More frequently, though, policy making consists of an ongoing process in which beginning and ending states are unclear, and in which both outputs and intentions are continually modified.

Policy struggles are usually preceded by a background period during which intentions and opinions are being formed. Later, outputs are issued, modified, and remodified as administrators respond to continuing pressures or as losers initiate appeals to higher authorities. At times, reviews by higher authorities create real opportunities to change results; at other times the decision is merely rubber-stamped. At what points are researchers to slice into the ongoing policy process to measure the relevant intentions and outputs? How many and what sorts of appeals are to be considered before a decision process is regarded as complete? These questions are important because the measurement of such variables as the number of participants, the duration of conflict, the resources employed, etc., will be significantly altered depending on the time period over which they are measured.

An illustration can help clarify some of these difficulties. In the case study of "The Glavis-Ballinger Dispute,"[14] the controversy progressed through several stages. At stake was the validity of the Cunningham claim to large tracts of public land in Alaska. At first there was an administrative determination that the claim was valid. Then Glavis, an official in the Bureau of Land Management, began a routine investigation into the validity of the claim, ultimately recommending that the land not be turned over to Cunningham. Secretary of the Interior Ballinger overruled Glavis. Glavis then appealed to President Taft, who supported Ballinger. But Forest Service Director Gifford Pinchot then helped Glavis publicize the issue and take his case to the Congress, and ultimately a law was passed which prohibited the sale of public lands in the future, but which still did not resolve the Cunningham dispute. Several years later a federal court found against Cunningham.

It may seem obvious that the court decision decided the case, but is this the most important point to examine? And what if the court had found *for* Cunningham? The determinative decision might then be considered to have been either that of Ballinger, or Taft, or Congress, or the court, with all of the subsequent decisions essentially ratifying the first. Similarly, one might argue that the case began at any of several different points: congressional authori-

[13] William A. Gamson, *Power and Discontent* (Homewood, Ill.: Dorsey, 1968), p. 71; Robert A. Dahl, *Modern Political Analysis* (Englewood Cliffs, N.J.: Prentice-Hall, 1963), pp. 39–54.

[14] [Winifred McCulloch,] "The Glavis-Ballinger Dispute," in *Public Administration and Policy Development,* ed. Harold Stein (New York: Harcourt, 1952), pp. 77–87.

zation of the sale of land; Cunningham's violation of the law in amassing his claim; Glavis's investigation; or the rejection of Glavis's recommendations. Depending upon the points chosen, a number of important variables, such as the degree of conflict in the dispute and the number of participants, would vary.

Treatment of this particular quandary not only affects decisions about the analysis of an individual case, such as the Glavis-Ballinger dispute, but might also artificially predetermine certain results of a quantitative analysis. Defining the duration of policy conflicts in such a way as to be as inclusive as possible, for instance, might well yield the finding that most important administrative decisions are determined in the courts, while in fact one suspects that there are many appeals in which there is no real hope of overruling an administrative determination—appeals that are taken largely pro forma in order to symbolically satisfy aggrieved interest groups. Another artificial result of inclusiveness would, no doubt, be the finding that congressional committees hardly ever have the final say about anything—although in fact they make an enormous amount of policy—simply because their decisions are routinely passed upon by the full chamber, conference committees, and the president. On the other hand, defining the time parameters of the process too narrowly—by attempting, say, to single out each discrete decision—would force the systematic analyst to consider thousands of decisions, jamming any data set with useless, unimportant, and misleading information.

It is thus essential, both when policy hypotheses are proposed and when they are tested, to have in mind some reasonable criteria for demarcating the process temporally. The reader and writer must agree somehow on the outputs that are theoretically crucial, otherwise doubt and confusion must inevitably arise regarding the consistency of theory and data.

Postdictive Theory. Another complication of temporality arises from the annoying tendency of variables not to stand still as the process unfolds. The duration, scope, or complexity of a given policy struggle are rarely strictly determined at the outset; not only are outcomes often unpredictable, but so too is the process. Thus predictive hypotheses that fail to account for this contingency often become extremely difficult to apply to concrete cases and are often best tested—and thus rendered most meaningful—by being converted into postdictive hypotheses. But that change may seriously dilute their significance.

An example is provided by the concept of "requisite actions" suggested by Banfield in *Political Influence*. Banfield introduces the concept as a way of explaining why some policy proposals are adopted and others are not. He employs the following definitions:

> Performance of a specified set of actions by specified actors, or by a specified number of, or proportion of, the actors who constitute a specified group, constitutes *adoption*, i.e., adoption is defined as the performance by these actors of these actions, which will be called *requisite* actions. . . . An actor who can perform a requisite action has *authority* over the action. He may perform it or not as he likes, or, in the language to be used here, he may *give* or *withhold* it from the system of activity being concerted toward adoption of the proposal.[15]

In any situation, certain actors will be controlled while others are autonomous—that is, free to withhold the performance of requisite actions. In highly centralized systems, there are few autonomous actors; in highly decentralized systems there are many. Adoption of proposals in a highly decentralized system is uncertain, according to Banfield, because it cannot be predicted whether or not autonomous actors will perform the requisite actions.

Banfield goes on to derive a number of interesting and nontrivial hypotheses from his argument. For example, he argues that, "as the number of autonomous actors in a situation increases, the probability of adoptions decreases,"[16] and that "corruption will tend to increase as the distribution of authority widens."[17] Banfield believes that these derived propositions can be tested: "If when tested—and some of them cannot be tested by any data in this volume—these derived hypotheses prove false, doubt will be cast upon the factual premises from which they were deduced."[18] But in order to test Banfield's model, it is necessary to identify and count the number of requisite actions that must be concerted in order for a policy to be adopted. And while Banfield asserts that it is uncertain that any given requisite action will be performed, it must be possible, in order for his theory to have predictive value, to at least specify what the requisite actions are before the policy struggle begins.

The number of requisite actions may, however, change dramatically during the course of a

[15] Banfield, p. 309.
[16] Ibid., p. 318.
[17] Ibid., p. 322.
[18] Ibid., p. 308.

policy struggle, as controversy heats up; and it may become clear that some actions were in fact requisite actions only after the policy struggle is over. Take the case, *Defending "The Hill" Against the Metal Houses*, as an example.[19] A developer wished to erect a questionable type of housing in a lower-middle-class neighborhood. Given the legal requirements of the jurisdiction, certain zoning and building clearances would undoubtedly have to be obtained. Are there other requisite actions? How long and harsh will the decision be in the making? Can we predict the probability of adoption from our initial count of the requisite actions?

If we are to count the neighborhood residents as "an autonomous actor," the struggle could be protracted, but the residents of such neighborhoods typically do not organize to make their wishes known. Thus, predictively, the number of requisite actions would seem to be, say, two or three. As the particular case turned out, however, the residents of the neighborhood did organize, and as a result—though not even then a clearly predictable result—a substantial number of aldermen, the mayor, the city's corporation counsel, the state board of health, and the opposition candidate for mayor all eventually became autonomous actors as well. The developer was defeated, but only after a long and emotional struggle.

The specific dynamics of the policy process, in this case at least, thus determined not only whether or not significant actors would give or withhold requisite actions, but how many requisite actions there would have to be. This is not to say that Banfield's theory is not correct as a post hoc generalization about the extent to which two interesting attributes of the policy process covary; but the implied causality in the original hypothesis becomes impossible to test empirically.

It is now clear that Lowi's central hypothesis, that characteristics of the process can be predicted from characteristics of an input (policy type) is fettered by this same difficulty. The policy type and perceptions of it by relevant actors often are determined well after the process has begun. Of course, one might predict new characteristics of the process continuously as a policy changes or is perceived to change from distributive to regulatory, regulatory back to distributive, and so forth, but this complicates the job of comparative analysis to the point of impracticability.

The Problem of Multiplicity of Policy Aspects

A public policy is only rarely the result of a simple binary decision, or even a chain of such decisions. Even the simplest government policy is likely to spring from a complex chain of causes and relationships and to have a set of subtly interrelated consequences for the general social network. Unless there is adequate awareness of this complexity, and adequate precautions are taken to cope with it, even those theories of public policy which seem most sensible are likely to evaporate on close inspection into a cloud of ambiguity. We have already discussed at some length the problems of complexity arising from the temporal nature of policy making. The point of this and succeeding sections is that policy research would be complex even if the process were instantaneous.

Lewis Froman, for instance, hypothesizes that homogeneous communities will adopt "areal" policies, while heterogeneous communities will adopt "segmental" policies. He defines areal policies as those which affect the total population of a city simultaneously by a single action, and segmental policies as continuing policies which affect different people at different times in separate sections of the city.[20] In his segmental category, Froman thus seems to take cognizance of the fact that the same policy can have different impacts in different places over time. Yet that awareness is of no help to a researcher trying to fit a policy such as, for example, "promoting industrial development," into Froman's categories. Industry may locate in a carefully zoned, narrowly circumscribed area, but (1) its pollution may affect surrounding neighborhoods; (2) it may provide jobs to residents of a much larger area; and (3) its property taxes may pay for services uniformly consumed throughout the whole jurisdiction. There is *no* a priori way for the researcher to determine which of its impacts are to be considered in deciding whether the policy is "areal" or "segmental."

The same problem is illustrated by Alan Altshuler's description of an intercity freeway dispute: should the freeway be built along a southerly route through a black neighborhood

[19] William K. Muir, Jr., *Defending "The Hill" Against the Metal Houses*, ICP Case #26 (University, Ala.: University of Alabama Press, 1955).

[20] Lewis A. Froman, Jr., "An Analysis of Public Policy in Cities," *Journal of Politics*, 29 (February 1967), 94–108.

or along a railroad line to the north?[21] The route through the black neighborhood was "scientifically chosen" for the shortest and best traffic patterns. The benefits from that route certainly appeared to be areal. The highway would be used by virtually everyone, and even those who didn't use it would benefit from the generally improved transportation and commerce in the area. More importantly, state and federal assistance was available for that route, lowering the tax burden for everyone in choosing it. On the other hand, the costs of the decision were quite clearly segmental. Many families were displaced from the area immediately surrounding the construction, with little or no counseling or provision for replacement housing. The displaced residents moved primarily into nearby buildings, seriously increasing already severe congestion and suffering the noise and pollution consequences of their continued proximity to the freeway. From the point of view of the winners, then, the freeway location decision was areal; from the point of view of the losers, it was segmental. From the point of view of a researcher trying to assign the policy to one of Froman's types in order to test his hypotheses, the decision was not easily classifiable.

The difficulty we encountered above in treating Lowi's stimulating hypothesis in its illustrative application to the subsidy for the Chicago Transit Authority also fits into this category. The policy proposal took on at least two separate and salient aspects, service and financing, which made difficult its classification into one and only one category of Lowi's policy typology.

The Problem of Multiple Participants

Even if the policy-making process were both instantaneous and unitary, it would still be complex as a focus of analysis because of the multiplicity of participants involved and the towering importance of participant characteristics, both subjective and objective, as elements of policy theory.

Subjective. The perceptions of relevant actors are important determinative variables in many theories about public policy, but policy theories rarely specify whose perceptions are to be taken into account. An illustration of the necessity of specifying from whose point of view a concept is to be defined is provided by hypotheses in which the status quo is an important concept. Gamson, for example, argues that it takes fewer resources to defend the status quo successfully than to bring about change.[22] While the status quo seems to be an objective characteristic of a state of the policy process, it may often be defined only in terms of participants' perceptions, which may in fact differ substantially from one another. Take the following example: in the case study, *Shooting Down the Nuclear Plane,*[23] all agreed that the specific purpose of an existing $75 million appropriation was to carry out a small development program. From the Air Force's perspective, however, that appropriation was only a first step and constituted a commitment to the actual construction of a nuclear plane. To the Air Force, it was that commitment which defined the status quo. Failure to expand the program and build the plane would be a negation of the commitment and thus a serious departure from the status quo. From the perspective of the members of the congressional appropriations committees, on the other hand, the small development program represented only a limited venture; from their point of view, confining the future program to reactor development, for example, would represent no change whatever in the status quo. In a situation of this sort, how is one to evaluate Gamson's hypothesis? Each protagonist, the Air Force and the Congress, thought that it alone was defending the status quo and that the other was opposing it; there is no objective criterion on which the outside observer can base a decision as to whose perception was correct. It might well be considered sound to decree in this case that the most relevant perception of the status quo is that of the Congress. Its perception anchors the concept of "change" in Gamson's hypothesis; *Congress* must be changed. The generalization to be recognized, however, is that such a determination must be made on a priori theoretical grounds for each variable and each hypothesis subject to this kind of ambiguity in research.

Again, the example helps to pinpoint a problem noted in our earlier discussion of Lowi's typology. Lowi argues explicitly that the perceptions of actors determine the category into which a given policy must be classified, but he does not treat theoretically the question of whose perceptions must be dominant in the event that there are differences.

[21] Alan Altshuler, *Locating the Intercity Freeway,* ICP Case #88 (New York: Bobbs-Merrill, 1965).

[22] Gamson, p. 63.

[23] W. Henry Lambright, *Shooting Down the Nuclear Plane,* ICP Case #104 (New York: Bobbs-Merrill, 1967).

Objective. In many hypotheses, the independent variable seeks to describe a subjective characteristic of participants—that is, a state of the world as seen through a particular lens. When theorists fail to specify adequately whose subjectivity is to be measured, the problem described in the previous section arises. Other hypotheses, however, seek to link *objective* characteristics of participants—their resources, explicit attitudes, or demographic characteristics—to aspects of process or output. But those hypotheses are of little help in determining *which* participants the characteristics describe. It is useful to keep subjective and objective characteristics of participants distinct. Hypotheses regarding the former often are deceptively simple, because they refer explicitly to impersonal conditions (e.g., the status quo, the divisibility of the benefits), which in reality are subjective perceptions of conditions; whereas the latter tend to refer quite obviously to people (e.g., resources used, strategy used). Both, of course, are troublesome for the same basic reasons—the multiplicity and heterogeneity of participants.

In *Democracy in the United States,* Dahl suggests that, "how severe a conflict is depends on how much is at stake."[24] Empirically, we might estimate how much is at stake ourselves as outside observers, or we might depend on the relevant actors' own views of what is to be lost or gained. Even if we elect the former, "objective" course, however, so that perceptions of actors are not relevant, we are still left with the unanswered question: What is at stake *for whom?* Variation is possible here, just as it is for factors such as social class or resources employed. The following concrete instance illustrates how answering that question may be critical to an empirical test of Dahl's hypothesis.

The issue in the case of *The Florida Milk Commission Changes Minimum Prices* was a proposal for the complete deregulation of milk prices, replacing a system of controls at all levels from the dairymen to the consumer.[25] Three different groups were primarily affected. Consumers could be expected to benefit in the short run from increased competition within the milk industry. The "big three" dairies would be hurt in the short run, since they would have to lower prices and curtail profits. Most seriously affected of all, however, would be the independent distributors, who would not be able to compete with the "big three" and might eventually be forced out of business altogether, if past history in the market area were any guide. In the long run, the "big three" would then benefit from the reinstitution of an oligopolistic market in which they each had a larger market share. Conversely, consumers could expect the long-run outcome to be neutral at best, and possibly negative.

Dahl's hypothesis could thus be applied to this case in at least three different ways. If one looked primarily at the consumers, one would predict very little conflict; for them, the stakes were quite small. For the "big three," the stakes were large but not overwhelming. For the independents, the issue was virtually one of life or death, since they could not compete in either buying power or production efficiency if prices were deregulated. The choice of which group to focus on is thus crucial. If one takes the party with the most at stake, one would expect the conflict in this case to be quite severe; if the party with the least at stake were used, very little conflict would be expected. To take an average would similarly be to predict a relatively low or at most a moderate level of conflict. More to the point—unless one attaches a way of surmounting the problem to the original formulation of the hypothesis—the theory cannot be tested. It remains, in an important sense, incomplete.

The Problem of Interaction

Interaction among independent variables in determining an outcome is of course common in social theory and in research findings. By "interaction" we mean that the existence or strength of an effect is contingent upon some other condition or the value of some additional variable; for example, the effect of financial resources upon success in a policy struggle may depend upon the arena of decision—legislative, executive, or judicial. The phenomenon is quite obviously not peculiar to public policy. We suggest, however, as the complexity emphasized repeatedly in the foregoing discussion would indicate, that interaction is endemic to public policy—it is perhaps its most salient characteristic—yet, it is rarely recognized in theoretical offerings.

Many important theories about public policy are probably incorrect as generalizations encompassing everything that falls within *anyone's* image of policy making, while essentially accurate for that subset of examples which

[24] Robert A. Dahl, *Democracy in the United States,* 2nd ed. (Chicago: Rand-McNally, 1972), p. 303.

[25] Harmon Zeigler, *The Florida Milk Commission Changes Minimum Prices,* ICP Case #77 (University, Ala.: University of Alabama Press, 1963).

theorists implicitly take as their *own* definition of the universe of applicable instances. The failure to demarcate that universe is potentially a failure to recognize important statistical interaction. We will offer a specific illustration in a moment, but we note in passing that the general definition of "policy" itself may be a contingent condition upon which the applicability of a theory is meant to depend. There exists in the literature a rather astounding number and variety of suggested boundaries (or lack thereof) about the concept, "policy": all government action;[26] a program of goals, values, and practices;[27] the impacts of government activity;[28] general rules to subsume future behavioral instances;[29] the consequences of action and inaction;[30] important government decisions;[31] and "a particular object or set of objects which are intended to be affected ... [together with] a desired course of events ... a selected line of action ... a declaration of intent ... and an implementation of intent. ..."[32] It is perhaps too much to ask at this stage that we all agree on our usage of the term "policy," but in the absence of such agreement, it is well to recognize that hypotheses might receive more or less support if tested on all government decisions and actions or on some one of the many alternative subsets referred to as "policy" in current theoretical writings.

To illustrate the problem of unspecified interaction more specifically, we will consider hypotheses about group size as a political resource. Pluralists in general postulate that group size is an important potential resource. Murray Edelman, on the other hand, argues that large groups are more likely to be bought off with symbolic reassurances than small, well-organized groups.[33] E. E. Schattschneider argued still a third position: that group size may be of relatively minor significance because the relevant group may be able to involve wider publics or disinterested government officials in a dispute.[34] It is easy to think of examples to illustrate the persuasiveness of each of these conflicting hypotheses, and, indeed, each author provides several. None, however, attempts to place his propositions in the context of more general theory by specifying the conditions under which they are valid. One is thus left with three opposing theories, all of which may very well be valid for a broad class of decisions, but with no clue as to when they are valid and when they are not, and why.

By way of contrast, it might be useful to offer the work of one theorist who appears to have adequately stated the conditions under which his hypotheses can be expected to be supported. In his *The Logic of Collective Action*, Mancur Olson takes exception to the pluralist position concerning group size as a resource, contending that large groups may be ineffective in pursuing their interests in comparison with smaller groups.[35] Although his hypothesis is derived from more general economic theory, Olson also relies heavily on the kind of illustrative material used by Edelman and Schattschneider. Yet Olson carefully delimits the scope of his theory by specifying the circumstances under which the behavior he predicts is most likely to occur. Large groups, for example, will be able to organize effectively when they are seeking collective goods if their members benefit disproportionately, or if they

[26]Lowi criticizes Dror for defining policy as simply any output of any decision maker in his book *Public Policy Making Reexamined* (San Francisco: Chandler, 1968). Dror never formally defines policy but his discussion indicates Lowi is correct. See Lowi, "Decision Making vs. Policy Making," p. 317. Thomas R. Dye defines public policy as "Whatever governments choose to do or not to do," *Understanding Public Policy* (Englewood Cliffs, N.J.: Prentice-Hall, 1972), p. 1.

[27]Harold D. Lasswell and Abraham Kaplan define policy as "a projected program of goal values and practices," *Power and Society* (New Haven: Yale University Press, 1970), p. 71. See also Carl J. Friedrich, *Man and His Government* (New York: McGraw-Hill, 1963), p. 70.

[28]For example, Easton writes: "Arriving at a decision is the formal phase of establishing a policy; it is not the whole policy in relation to a particular problem. A legislature can devise to punish monopolists; that is the intention. But an administrator can destroy or reformulate the decision by failing either to discover offenders or prosecute them vigorously. The failure is as much a part of the policy with regard to monopoly as the formal law. When we act to implement a decision therefore we enter the second or effective phase of a policy." *The Political System*, 2nd ed. (New York: Knopf, 1971), p. 130.

[29]This is our own preferred definition of policy. See our "Case Study Aggregation and Policy Theory," p. 11.

[30]Dye includes the consequences of inaction as well as action, whether intended or not, in his definition of policy. Dye, p. 2.

[31]Lowi suggests the need to look at only important substantive government decisions, "Decision Making vs. Policy Making," p. 317.

[32]Ranney, p. 7.

[33]Murray Edelman, *The Symbolic Uses of Politics* (Urbana: University of Illinois Press, 1974).

[34]E. E. Schattschneider, *The Semi-Sovereign People* (New York: Holt, 1960), pp. 3–77, passim.

[35]Mancur Olson, *The Logic of Collective Action* (Cambridge, Mass.: Harvard University Press, 1965).

are able to coerce their memberships or provide members with side payments of selective goods. Because Olson makes such contingent conditions explicit, his hypothesis is testable without requiring further refinement—without the researcher having to guess at the theorist's intentions.[36]

Implications

We have argued from our own research experience that the systematic study of public policy is seriously complicated by the nature of the beast. "Policy" is complex—because of temporality, because of multiplicity of aspects and participants, and because of interaction. Yet complexity need not prohibit systematic empirical research. So long as it is accounted for and confronted directly, so long as hypotheses are sufficiently precise, theoretical models may adequately represent reality. We suggest that most of the lessons to be learned have to do, not so much with the conduct of data analysis, as with the formulation of hypotheses and the elaboration of theory.[37] Specifically, we suggest the following:

1. **Temporality: Multiple Outputs.** Optimally, we should all agree on how to identify key developmental points in the policy process, such as the beginning point and the point at which it might be said that policy was indeed "made." Such agreement is undoubtedly premature. In its absence, however, systematic comparative analysis still requires that key stages be identified. In this one instance, it is the data analyst who perhaps has more to contribute than the theorist. We can learn by experience which identification criteria are reliably operationalizable and which among them seem to yield fair tests of hypotheses. In our own work, we have found it both practical and productive to identify a "point of last significant controversy" and, somewhat less importantly, a "point of first significant controversy." The terms are almost self-explanatory, although an elaboration of their meaning and use is available elsewhere.[38] We commend these criteria to others for consideration and trial and urge the formulation and trial of alternatives as well. Most importantly, we emphasize that objective comparisons cannot be made unless a criterion is consistently applied to each case in order to determine which of its many decision points establishes "policy."

2. **Temporality: Postdictive Theory.** One must simply be sensitive to the problem and avoid creating predictive hypotheses that fall into the postdictive trap. Specifically, one must not attempt to predict characteristics of the policy *process* either (a) by variables (such as number of requisite actions or resources committed to the struggle) whose value or score for a given case is not known until too late in the process itself, or (b) by variables (such as "policy type") that are offered predictively by theorists as inputs but that in actuality are characteristics of outputs. For a policy hypothesis to be predictive, the causal variable must obviously be observable and measurable at a point in time before the effect and should not be subject to significant change beyond that point of evaluation.

3. **Multiplicity of Policy Aspects.** The general implication of this problem is that we must either avoid variables that may differ significantly in value depending upon the aspect of a given policy to which they are applied, or that we must provide a means of selecting the aspect that is most relevant for the operationalization of a given variable. Two guidelines on this issue stand out in our experience to date. One is that any typology of whole policies runs a substantial risk of ambiguity at the stage of operationalization; typologies can be extremely valuable, but they should optimally be offered along with well-considered ground rules for classification. The other guideline is that a major (but not exclusive) source of ambiguity lies in the possible divergence of the benefits from the costs or the substantive from the financial aspects of a public policy. In creating hypotheses, it may be profitable for the theorist to consider whether the variables are subject to

[36]This does not mean that Olson's theory is necessarily correct. There may be additional contingent conditions which he failed to state that would further refine or modify his theory. John Chamberlin, for example, has recently suggested some additional contingent conditions under which large groups are likely to provide large amounts of a collective good in contradiction to Olson's predictions. All we want to imply is that the concepts in Olson's theory are reasonably well specified and the theoretical relationships among concepts are stated with reasonable precision, allowing others successfully to re-examine and test them. See John Chamberlin, "Provision of Collective Goods as a Function of Group Size," *American Political Science Review*, 68 (June 1974), 707–16.

[37]For illustrations of how this process can both permit the elaboration of theory and provide operationalized hypotheses for empirical testing, see Jeffrey A. Miller, "Welfare Criteria and Policy Outcomes: An Empirical Assessment" (Ph.D. dissertation, University of Michigan, 1975).

[38]See our "Case Study Aggregation and Policy Theory," p. 14.

differential rating depending upon whether benefits or costs form the basis of evaluation.

4. **Multiple Participants: Subjective.** The question of whether it is most important to consider a given possible state of the world as it in some sense actually or objectively exists or as it is perceived by participants in the policy process is generally given inadequate attention. If there are good theoretical or methodological grounds for choosing the latter, it is essential to consider whether different participants might perceive the status of affairs differently. If so, it is necessary to translate those grounds into guidelines on whose perceptions are to govern the scoring of that particular dimension for research purposes.

5. **Multiple Participants: Objective.** This problem is potentially the most troublesome of all, since characteristics and behaviors of participants emerge so commonly in policy theory as important variables. The theoretical considerations that necessitate testing a hypothesis comparatively—by assigning a value to each policy struggle to represent some objective characteristic or behavior of "participants"—should make possible some decision as to how those characteristics should be operationalized. Theory might provide criteria for selecting *one* participant or homogeneous group as the basis for assigning a value to the variable, or, perhaps, criteria for aggregating the value across one or more groups of heterogeneous participants. We have had substantial success in dividing the participants in each struggle into two opposing camps and, for most participant-oriented variables, giving an aggregate score on the variable to each camp. However, this attempt at a universal method is much less satisfactory for some variables than for others, and for some cases of policy making than for others. Sometimes it does not work at all. Other schemes might be devised, but such tinkering by the data analyst is in general less desirable than explicit criteria generated within the theoretical proposition to be tested; for example: "In predicting the severity of conflict by 'how much is at stake,' the true predictor is the average of the perceived stakes across all major participants."

6. **Interaction.** It is clear that not all hypotheses can be valid for all types of policy making in all kinds of circumstances. To the extent that a theory fails to specify major conditions defining its applicability, it is inadequate theory. If the conditions of validity are specified, not only is the analyst's job made easier and a fair test of the proposition likely to ensue, but also direct benefits result for the development of the content of policy theory itself. Theory should be parsimonious, to be sure, but not oversimplified.

Conclusion

Because of the complexities of public policy as an object of study, we may never be able to obtain hard knowledge of the policy process of the type available in the advanced physical sciences. Yet improving our understanding of policy phenomena is clearly possible, if only through advancing the conceptual sophistication of theoretical formulations.[39] Our own experience has convinced us that such advances will be more rapid and certain as theory encounters systematic empirical data. As the brief comments in this paper illustrate, the collision between theory and data, while perhaps frustrating at first, can have important benefits for both researchers and theorists. Those lacunae in theory, painfully identified by difficulties in operationalization, become foci for efforts at additional theorizing that, however much they may do violence to the intentions of the original theorist, can add considerably to the richness and utility of hypotheses. The necessity of separating aspects from cost aspects in theories of policy type, or the notion of "last significant controversy" in a policy process, are small, but not trivial examples. If present theoretical levels in public policy are to progress, obstacles of the kind identified here should not be viewed as roadblocks. They cannot be wished away, nor can they be evaded. But by grappling with them directly, it is possible to add to the precision and sophistication of theory while proceeding with the essential conduct of supporting empirical research.

[39] For an illuminating discussion of the role of "hard" research in developing social theory of various types, see the following two papers by Anatol Rapoport: "Various Meanings of 'Theory'," *American Political Science Review*, 52 (December 1958), 972–88, and "Explanatory Power and Explanatory Appeal" (Paper prepared for the Conference on Explanatory Theory in Political Science, Department of Government, University of Texas at Austin, Feb. 19–23, 1968). Rapoport argues that because of the limitations of social science, improved conceptualization is often a more important criterion in judging good theory than predictive power. In this regard, the value of the application of hard methods in social science would not necessarily lie so much in improved prediction as in better conceptualization and the reformulation of thinking about problems.