

# Presidential Address— The Extrapolation Problem: How Can We Learn from the Experience of Others?

*Eugene Bardach*

I want to pose a very common but largely unrecognized problem of professional practice. It is also a problem of great intellectual complexity. I call it The Extrapolation Problem. It goes like this: Suppose you are a young policy analyst. Your boss calls you in and says, “We’ve got this problem here in Jackson County of big fire hazards building up in the rural-urban interface. Go find out what they’re doing about it over in Washington, Jefferson, and Lincoln counties. See if anything they’re doing about it might make sense for us.” What methodology should you use in fulfilling this request? How do you conceptualize “what they’re doing,” and see which if any elements might “make sense”?

As an additional challenge, you must assume that strict and faithful replication is not in the cards. That just isn’t the way of the world. For lots of very good reasons, if people wish somehow to borrow from the wisdom and practice of others, they want to adapt it, customize it, localize it. If you want to perform well for your boss—not to mention the public interest—you must think not only about what you do see, but also, to some extent, about what you don’t see. That is, you have to speculate about those modest variations that would show up, both by accident and by design, when some practice would move from those source sites in Washington, Jefferson, and Lincoln counties to your target site, Jackson County.

## PREVALENCE

How common is the extrapolation problem? Very common. In the workshop classes I have taught at the Goldman School, where students work on real projects for real clients, over the last 5 years, 45 of 103 candidate projects for the first-year workshop entailed checking out what might be learned from the experience of various source sites. For instance:

- The Oakland, California, Police Department wanted to know whether to offer child care to its employees. Students looked to the U.S. military and to Portland, Oregon, for guidance.
- The University of California Medical School in San Francisco was thinking of setting up an organization to manage clinical trials for its own and perhaps other researchers. Harvard and Duke were two sources of useful guidance.

- The Health Department in San Mateo County, California, wanted to increase Medicaid enrollment in the county. The student policy analyst looked to the experience of other Bay Area counties.

### EXTRAPOLATING FROM "BEST PRACTICE"

These are examples of the extrapolation problem as seen by those who seek to learn from the experience of others. But it also looms for those who would teach from their own or others' experience. Compendia of "best practice" in dozens of policy areas are now exploding on the Internet.<sup>1</sup> These compendia purport to offer guidance to all and sundry, usually drawn from experiences in one or a few sites. A recent symposium in *JPAM*, in the Professional Practice section, of which I am editor, had the originators or keepers of four such compendia explain how they went about their work (Bardach, 2003). These authors generally made reasonable cases as to what they chose to include in their compendia and how they chose to present the best practice entries. But these and many other compendia I have seen assume that users are considering replication rather than adaptation or inspiration. Perhaps one reason is that the compendium-makers simply don't know how to teach about adaptation.

### ERRORS OF INTERPRETATION

Let us return to the Jackson County fire prevention problem and consider some of the possible mistakes that occur when the experience of a source site is misinterpreted. I shall present three vignettes. In the first two, the policy analyst charged with scouting out others' experiences underestimates the potential for what is observed. In the third, he underestimates the risk of failure and even worse:

- In Washington County they have a regulation that property owners must maintain a 15-foot clearance between flammable shrubbery and residential structures. It is enforced by fines and, in certain cases, county employees clear the shrubbery themselves. Due to property-owner resistance, Washington County hardly enforces the regulation. Jackson County officials mistakenly conclude that "regulation" won't work. But it might in fact have worked if Washington County had either added a public education campaign or gotten community leaders to lead by example.
- In Jefferson County officials have encouraged the public agencies managing large tracts of land near the rural-urban interface to jointly hire a goatherd contractor to control vegetation on their domain. Their joint action creates limited monopsony power that keeps the contractor's rates low and the work affordable. Jackson County replicates Jefferson County's success, but fails to see the potential for additional collaboration by these agencies with regard to more comprehensive inter-jurisdictional planning for fire prevention.
- In Lincoln County, they create a special assessment district for property owners in the vulnerable areas to pay for more intense fuel management and public education programs. The program is a great success. Jackson County officials set out to do the same thing. But pretty quickly, some angry property

<sup>1</sup> See, for practices related to reducing poverty, Kruger, 2002. A recent Google search on the Internet returned 286 entries for "compendium of best practice" and 15,700 entries for the less restrictive "compendium," "best practice."

owners move to block the new district, on the grounds that the county owes them fire-prevention services, for free, under its general obligation to protect against fire. The Jackson County commissioners, partly in consequence, make it a point of principle to withhold extra service for a group some call “those rich, selfish newcomers.” The result is that the needed additional investment is not forthcoming.

## THE ROLE OF METHODOLOGY

Now one might say that the failure of our young policy analyst in these vignettes is not one of methodology, as I claim, but of imagination or intelligence. If only the analyst were not such a dim bulb! But assigning blame is not a matter of “either/or.” Good methodology can to some degree substitute for imagination and intelligence. It can also stimulate imagination and intelligence. That is, it can work as both substitute and complement. Thus, it is so much the more to be valued.

Perhaps “methodology” is not the right word. It is perhaps too prescriptive. Perhaps a “conceptual framework” is more to the point, a conceptual framework that would prescribe what relationships to look for and what uncertainties to calibrate. To better illustrate what I have in mind, let us contrast the interpretive skills of an architect and her client who together go on a home-and-garden tour hoping to gain inspiration for their own project from one or more of the sites they visit.<sup>2</sup> On seeing one particular site, the client exclaims, “I just love all those view windows! I’d like those in my house.” The architect, more aware of the constraints and trade-offs in their own project, says, “I’m afraid we’ll have to adapt what they did on that magnificently large scale to look and work a little differently. Please tell me more precisely what you like about those view windows. The light? The view? The openness? The sense of added space? The absence of boundaries?” The architect is not ready to commit to any particular features without understanding what functions and purposes are to be served. Most importantly, the architect has at her disposal a whole menu of options, conceptualized within a framework of interrelated forms, materials, and functions with which to create alternative templates for her client. In a word, if she is reasonably sophisticated and experienced, she knows deep in her bones what the interesting counterfactuals are; that is, those qualitative variations on what has been observed which, though perhaps small in some sense, might have large effects. Whether it is a “methodology” or a “conceptual framework” that the architect has, it is lacking in the toolkit of the policy analyst trying to make sense of other people’s experiences and practices.

So much for my definition of the problem. I hope I have convinced you that the extrapolation problem is real and that the solution to it, if there is one, is not simple. I hope I have also convinced you, through my invocation of architects and architecture, that a systematic approach to at least some kinds of adaptation problems is feasible. Now I wish to sketch some ideas for a solution.

## EXAMPLE: THE INTENSIVE POLICY EXERCISE

Let us work through a particular example, a practice pioneered at the Kennedy School of Government at Harvard, and subsequently adopted at the Ford School at the University of Michigan, which I will call the “Intensive Policy Exercise,” or IPE,

<sup>2</sup> I happen to know that the architect is a “her,” since I have in mind my architect wife, Nancy.

though Michigan calls it the “Integrated Policy Exercise” and Harvard the “Spring Exercise.” As a potential adapter at a target site, the Goldman School of Public Policy at Berkeley, I have been fairly close to the data, which is an advantage in speaking about it. A disadvantage is that it is very close to home and takes place inside the “ivory tower.” It might therefore seem idiosyncratic. However, I do not think that is so. Certainly a curriculum design problem is a stock problem in education, at whatever level. It is also the sort of stock problem that shows up for line-level service delivery or enforcement agents in a great many organizations, be they public, nonprofit, or for-profit. *Mutatis mutandis*, it could be about motivating compliance with fire codes, signing up Medicaid enrollees, or developing watershed protection plans.

A few years ago, in our seemingly perpetual effort to improve curriculum, a faculty-student committee at the Goldman School was considering how better to teach problem-solving skills. The IPE came to our attention. Looking at materials on the Kennedy School and Ford School web sites, and talking to a few colleagues at these institutions, we described its main features this way:

- While the exercise is in progress, all other courses are shut down or not yet begun.
- Students work on a single large project, of global, national, or statewide import, e.g., welfare reform, global warming, Social Security reform, mandatory living wage.
- Students work in teams organized around sub-projects.
- The problem is multi-faceted and complex.
- Varied disciplinary skills are relevant.
- Students prepare a written report (or memos) and an oral briefing.
- Their audience is high-profile officials whom faculty invite to participate.
- Faculty and experts provide much or all data to the students.

From what we learned, we concluded that the IPE worked reasonably well, that it did in fact improve students’ problem-solving skills. We believed our colleagues at the Ford School and the Kennedy School would not keep a practice like this in the curriculum if they did not think it worked; and we assumed that their thinking was probably good. Further, we had anecdotal evidence that students enjoyed the experience and believed they learned from it. Harder measures of success, or failure, would of course have been desirable. But in the absence of these, we were prepared to rely on our colleagues’ hunches about the matter and student reports. In the language of the program evaluation field, we assumed that the practice was efficacious and that the problem of internal validity in reaching this conclusion was adequately solved.<sup>3</sup>

<sup>3</sup> It is not always possible to solve the internal validity problem, to be sure, and one ought always to be humble. But the logic of quasi-experimental research design and efforts to rule out competing explanations can often be applied (Shadish, Cook, and Campbell, 2001). In any case, in the event that we would have gotten more serious about replicating or adapting the IPE, we would have revisited the question of whether it worked, and tried to answer it with greater rigor. For instance, we would have asked our colleagues at these institutions pointed questions about the evidence they used to convince themselves of its efficacy, and we would have urged them to apply the logic of quasi-experimental research design which would have been familiar to them. Obviously, this approach to internal validity only works when your informants are as sophisticated as you are about the pitfalls of causal inference.

## THE ARCHITECTURE OF "HOW IT WORKS"

This leaves us with the problem of external validity. Now, if strict replication of a practice in some target site under conditions identical to those in the source sites is at issue, there are known methodologies for assessing the suitability of such a move. The program evaluation field is rich with advice. But if the target-site environment introduces variations from what was analyzed in the source sites, the problem becomes much more difficult. You are extrapolating, not replicating. The most accessible, if not always the most reliable, method of solving the extrapolation problem is what Shadish, Cook, and Campbell (2001, pp. 369–371) call “causal explanation.”<sup>4</sup> For instance (p. 369),

...knowledge of how electricity is generated allows us to provide such power to satellites in space where electricity may never have been available or studied before; and knowledge of the pharmacologically active ingredients of aspirin might allow inventive persons to create the drug’s equivalent out of local plants that have never before been used to cure headaches.

Shadish, Cook, and Campbell are at pains to stress the uncertain state of causal theory in the social sciences—and, I would add, in folk thinking—and hence to underline the current limitations of this approach to solving the extrapolation problem.<sup>5</sup> However, for my purposes, it is the main alternative to mindlessly reproducing all the surface features of what the source site is doing minus the ones that happen to be inconvenient or unnoticed and plus the ones that politics, chance, and whimsy happen to introduce. And as the principal vehicle for utilizing causal theory, I rely on the idea of mechanism.

### Basic Mechanisms and Their Purposes

In any practice, the driving elements come packaged as a “basic mechanism,” a mechanism that has some sort of causal power.<sup>6</sup> The IPE has two basic mechanisms of interest. One involves the emotional and intellectual arousal that occurs under moderate pressure. Psychologists tell us that a moderate degree of arousal enhances learning and memory. The arousal here comes from deadline pressure, the high expectations of performance from faculty and fellow students, and the absence of any (school-related) distractions. Secondly, there is a mechanism made up of peers-

<sup>4</sup> The extrapolation problem is a special case of the generalization, or the external validity, problem.

<sup>5</sup> Interestingly, another major textbook in the evaluation field, Rossi, Freeman, and Lipsey (1999, pp. 272–273), says relatively little about the problem of generalizing beyond the evaluation site.

<sup>6</sup> In recent years there has been a flurry of interest in the social sciences in the idea of mechanism as an explanatory device. It is, in one sense, a schema for filling in the causal linkages within what would otherwise be black boxes connecting variables on the left- and right-hand sides of an equation. In another, more analytical, sense it is a strategy for achieving middle-level explanations of phenomena that are less general than full-blown “laws,” if such exist, but more abstract and general than low-level narrative and description (Elster, 1998; Hedstrom and Swedberg, 1998a). A good example is a contagion process; another is a market that tends toward a price equilibrium. My own conception of mechanism acknowledges the existence of a causal element but downplays its explanatory function in favor of its practical function. In my rendering, that is, the mechanism is not interesting because it solves an explanatory puzzle of how to link some effect to its causes. Rather, it is interesting primarily for itself, as a method of actualizing some latent potential and converting it to any number of possible ends. The lever, for instance, a practical mechanism, is not interesting because it can “explain” how some particular body was moved or elevated from place A to place B. It is interesting because it has the potential to be applied in numerous and diverse settings for numerous and diverse ends.

as-teachers. The IPE takes advantage of the capacity of peers to function for one another as teachers, critics, coaches, and helpers.

With these two examples in hand, I want to say something more abstract about the idea of a “mechanism.” “Mechanism” is obviously a metaphor, and there is no escaping the use of metaphors when dealing with issues of ontology. I will add a second metaphor: the sort of basic mechanisms I have in mind tap into “reservoirs” of what might be thought of as energy, or potentiality, in “social nature.”<sup>7</sup> These reservoirs are sitting around free, nature’s gifts to us, like the energy stored in hydrocarbon molecules that can be extracted and then channeled into useable mechanical force. One such natural reservoir is the capacity of humans to learn more efficiently when moderately aroused. Another is the capacities and inclinations of students to educate one another, provided the social setting makes this possible. To extend the metaphor further, obviously, a mechanism that “taps natural reservoirs” must have a way of doing the tapping, some system of human contrivance that manages to draw on this reservoir of energies and channels them to some productive purpose. In the IPE example, this means, for the heightened-arousal mechanism, primarily a combination of time pressure, peer and faculty expectations, and isolation from the normal curriculum experience. For the peers-as-teachers mechanism, it means the constitution of project teams large enough to include individuals with diverse skills and knowledge but small enough to enable constructive interaction.

### *Cost-effectiveness*

What does a basic mechanism within the relevant practice do? It produces effects, of course, some sort of valued results. In this case, these are improved student capacities for public sector problem-solving. But in many cases, we are interested not only in effectiveness but in cost-effectiveness also. Typically, the question “How does it work?” is asked and answered in terms of effectiveness alone. But, because potential adopters of a practice generally do feel constrained by an implicit budget of money or personnel or time, explaining how it manages to be both effective and cost-effective is probably almost as important, if indeed the two can be separated at all. (I like to call apparently cost-effective practices “smart practices.”<sup>8</sup>) Now in the IPE case, it so happens that both these mechanisms are not only effective but cost-effective.

How so? It pays to remember that student time is one of the most constrained of inputs into the learning process. Heightened arousal is a way to make it more productive than it otherwise would be, and at relatively low cost to the rest of the curriculum. Of course, in a different ledger book, students’ learning hours are probably increased relative to what they might otherwise have been, and these hours must be

<sup>7</sup> How one ought to conceptualize and operationalize particular mechanisms is uncertain, at best. The contributors to the Hedstrom and Swedberg (1998a) volume focus on conceptual strategies likely to produce results analytically useful for social theory. In their introductory chapter, for instance, Hedstrom and Swedberg (1998b, p. 24) refer to four “core principles” for doing the conceptual work: action, precision, abstraction, and reduction. I do not disagree with these principles. However, my own approach conceives of mechanisms as ontologically real, and not merely analytical. Hence, reaching for an underlying, albeit invisible, reality is worth trying. Such a reality is approachable mainly through the use of metaphors. Agent-based simulation of complex processes could be another avenue too, and ultimately a preferable one since it is more precise. In some cases, mathematical models could also be helpful. In any event, it is essential to distinguish the verbal or mathematical or simulation model of the mechanism from the mechanism itself (Schelling, 1998).

<sup>8</sup> In general, the arguments in this paper reflect the arguments in my earlier publication (Bardach, 2000, pp. 71–85). I have made two significant semantic changes (“technology” became “mechanism” and “opportunity” became “reservoir”), however, and have amplified and, I hope, clarified a number of points.

drawn from students' leisure time. The psychological cost of these hours may be low, however, if the experience has what psychologist Mihalyi Csikszentmihalyi (1990) calls the "flow" quality that makes time disappear. As to the peers-as-teachers mechanism, the students obligingly charge their institutions nothing for their services and probably even enjoy the work.

### Contingent Features

Although the principal component of any practice that is a candidate for replication is its causal mechanism, this mechanism is surrounded by an institutional, political, economic, and interpersonal context. To the observer, the practice appears linked to a number of contingent features within this context that may or may not warrant extrapolation to the target site. For the most part, these somehow implement or support the basic mechanisms of the practice. One could say they are merely instrumental. As is the case for most instrumental functions, they can be performed in a variety of ways, depending on the available alternatives and on local costs, constraints, and preferences. Therefore, they are prime candidates for adaptation. The contingent features may be categorized as implementing features, optional features, and supportive features.

### Implementing Features

"Implementing features" are those that directly implement basic mechanisms. In the case of maintaining a high-arousal IPE environment, I would nominate tight deadlines, a well-chosen topic, the division of the class into project teams, and the exclusion of competing school-related activities. With regard to creating a capacity to use peers as teachers, the key feature is project teams that include a diversity of talents but are small enough to permit effective interaction.

In general, implementing features can be contingent. If, in our architectural example, view windows had been intended primarily as a mechanism for admitting light rather than scenery, this function could be performed by a skylight.<sup>9</sup> In the IPE case, maintaining a high-arousal environment could be performed as we do it at the Goldman School, in our "48-hour exercise." In this exercise each student writes an issue memo on some unfamiliar topic framed by the course instructors. Each topic is drawn from a hat containing as many topics as there are students. One of the rules is that if the student knows too much about the topic already, he or she must replace the topic and draw another. The atmospherics of topic selection followed by frenetic searching among one's classmates for leads and ideas are, in effect, a functional substitute for the more structured method of setting peer expectations in the IPE.<sup>10</sup>

<sup>9</sup> In more explicitly causal language, given some function to be performed, some combination of implementing features is sufficient but no particular combination is necessary, so long as one of a number of effective combinations is actualized. These are, therefore, what the philosopher John L. Mackie calls INUS conditions: "...the so-called cause is...an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result" (Mackie, 1993, p. 34). The idea of INUS conditions to some degree saves the language of "necessary and sufficient conditions" from being as sterile and misleading as it often is. However, it too is limited and often inappropriate (Brady, 2003, pp. 14–18).

<sup>10</sup> I will here warn the reader of a possible semantic problem. Implementing features sometimes seem to be the least contingent of contingent features, in that we often define a basic mechanism in terms of one or more of its implementing features. In the minds of some observers, we might not be referring to "the Intensive Policy Exercise" at all if the exercise in question did not exclude competing school-related activities, even though the functions performed by this exclusion could perhaps be accomplished by different means altogether.

Of course, variation in the more detailed implementing features is especially common. We usually see it taking place across sites implementing the same generic mechanism. At the Ford School the duration of the IPE is one week, and at the Kennedy School it is two; at the Ford School it is scheduled before the spring semester, whereas at the Kennedy School it interrupts classes in the middle of the semester.

### *Optional Features*

Optional features vary across sites depending on local preferences and constraints. Unlike implementing features, they do not play an essential functional role. In the IPE case, the optional features have to do with the particular policy topic and level of government, and the mode of presentation. I would also include here the nature of faculty participation in coaching and giving feedback to students. Such participation could be high or low, presumably with somewhat greater or lesser benefit to the students, but this variation would not go to the essence of the practice. Although the use of high-level officials as an audience seems to me like an optional feature, some might argue that it contributes so much to heightening arousal that it should be deemed essential.

### *Supportive Features*

Supportive features are primarily those resources used to bring the implementing features into being, for instance a budget and an institutional infrastructure. Other supportive features have a less directly instrumental role but may nevertheless be important, e.g., the culture of the organization or the broader political environment. Supporting features for the IPE would include faculty, course assistants, clerical help, and physical facilities. I would also include student enthusiasm, since I doubt it is possible to carry out the exercise well without active student cooperation.

### *Secondary Benefits and Costs*

No mechanism does only one thing. It has secondary effects, both intended and unintended, both beneficial and costly if not harmful. Extrapolation from a source site must take note of these secondary effects as well as the primary and intended effects. The most interesting secondary effect of the IPE is that it encourages bonding among the participants. This has value over and above any learning effects. These may be especially valued at the Ford School, where the IPE kicks off the spring semester.

### *Vulnerabilities and Failure Modes*

Even the smartest of smart practices has its intrinsic vulnerabilities. These can lead to failure, corruption, political embarrassment, and more. For instance, a money-saving shift to outsourcing might end up costing more and accomplishing less if contracts are not correctly designed, or if the competitive environment is imperfect.

In the case of the IPE, the important vulnerabilities seem to me to be these:

- The exercise overall could be seen as too artificial, and student enthusiasm could wane or turn to cynicism.
- The topic might prove to be either too complex or too simple for effective learning.

- Related to the foregoing, some work groups may find their piece of the work too simple or too complex.
- Poor group dynamics might lead some teams to fall apart, perhaps the very teams whose products were needed to make the whole effort succeed.

There are two main sorts of vulnerability. One concerns what the evaluation literature calls “mediating conditions.” Such conditions cause the basic mechanism to work unusually well or poorly, e.g., the education or income level or cultural orientation of the population affected. I was once told that any favorable findings regarding implementation of a management reform in Minnesota could not be generalized because Minnesota was full of “good-government types” of Scandinavian descent.<sup>11</sup> Unfortunately, mediating conditions are numerous, and they sometimes—though not always—pose threats to the internal validity that I earlier assumed away.

The second source of vulnerability concerns the way a practice is implemented. This is the main source of vulnerability for the IPE, I believe. As James Q. Wilson (1967) wrote nearly 40 years ago, (bureaucratic) talent is the resource in most limited supply. He might well have expanded the list of scarce resources to include such things as political tolerance for trial-and-error learning, protection from partisan sniping, stability in key personnel, micromanagement by the legislative branch, and so on. Smart practices differ in just how vulnerable they are to adverse implementability conditions.

It is worth observing that explicit statements about vulnerabilities rarely show up in the descriptions one finds in compendia of best practices. Perhaps they would be too discouraging. Or perhaps the compendium-maker’s strategy is merely to spark initial curiosity. In any event, when a policy advisor in the target site returns with a report from the source site, her employers will want to know about the downside as well as the upside.

Unfortunately, the biggest limitations of working with one or just a few source sites is that one learns too little about possible failure modes. Accounts of their historical struggles to initiate and improve the practice can be helpful, as is a strategy of seeking out source sites where poor results or failure has occurred. Fortunately, with regard to the public sector, theory and imagination may be able to fill in where data are lacking. For instance, one can turn to political economy theories holding that any vulnerability to rent-seeking will be exploited, or to generalizations in public management about the cumbersome apparatus of governmental oversight delaying projects and impairing effectiveness.

### Three Heuristics

What does all this imply for the policy analyst rummaging about in various source sites looking for inspiring practices that might somehow be adapted back at the target site? At one level, the answer is, “Try to ascertain the mechanisms that make for both effectiveness and cost-effectiveness because, once you understand these, much else will fall into place.” Alas, it is hard to do this, for two main reasons. First, “mechanisms” are real but elusive. They are not tangible, like the view windows of my earlier architectural example. In all the conversations I have had with students and colleagues about “how the Intensive Policy Exercise works,” no one has ever

<sup>11</sup> How one assesses such a warning must be a matter of experience and judgment, especially in the absence of good systematic information. I believe there is some merit in the warning, although not so much as to justify ignoring Minnesota and its ilk altogether.

tried to conceptualize basic mechanisms, never mind the particular ones I have sketched. The standard interpretation is, “The IPE works by trying to simulate the real world.” This could be true, but it begs the question of why simulating the real world in this particular way would be an effective and cost-effective way to produce learning. To be sure, my own way of looking at things could be way off base. But I believe it more likely that people find it difficult to think their way down to those reservoirs in social nature where “mechanisms” get their fuel.

The second barrier to identifying cost-effective mechanisms has to do with the idea of cost-effectiveness. Here we enter even rougher conceptual terrain than that of “mechanism.” We are comfortable with the question, “What makes this practice effective?” We even have a set of theoretical actions, typically bearing on motivation, ready to act as candidates for answering this question: restructuring the incentives, applying deterrence, mobilizing peer support, improving the information available, creating stronger signals, etc.

Answers like these may be appropriate to questions about effectiveness. And indeed, “being effective” is very often what people mean when they say that some practice or program “works.” But if I am right that guidance is often sought not just about effectiveness but also about *cost*-effectiveness, then this sort of answer does not go far enough. We need somehow to invoke theories of cost-effectiveness as well. But, “What makes this practice cost-effective?” almost certainly feels more slippery, for it implicitly asks you to compare it to some other practice, unspecified, which is less cost-effective. And it is often hard to say what the most illuminating comparison would be.

#### “...Takes Advantage of...”

Let me suggest a semantic trick to help us over this treacherous terrain. Just as complicated semantics can obscure simple realities, so simple semantic tricks can help to clarify complicated realities. I can recommend such a trick for probing complex practices to pick out basic cost-effectiveness mechanisms (at least, where they exist). It is this. Fill in the blanks in the sentence, “This practice aims to produce a lot of value for relatively little in the way of ... resources by taking advantage of...” In the IPE case the complete sentences would read:

- “This practice aims to produce a lot of value for relatively little in the way of student time resources by taking advantage of the capacity of heightened arousal to facilitate learning.”

And

- “This practice aims to produce a lot of value for relatively little in the way of budgeted dollars or faculty time by taking advantage of the capacity of peers to function for one another as teachers, critics, coaches, and helpers.”

#### Ideas to Have Ideas with

I have another heuristic too. Because perception and observation are mightily assisted by a mind primed as to what to look for, I offer in Table 1 a short catalogue of generic smart-practice mechanisms.<sup>12</sup>

<sup>12</sup> This is a modified and, I hope, improved version of a catalogue I provided in my *Practical Guide*, p. 74, and in (Bardach, 1998, pp. 50–51).

**Table 1.** Some generic smart-practice mechanisms.

- 
- *Operations Research strategies.* By means of sequencing, timing, prioritizing, matching, clustering, and other such rationalizing arrangements, it is possible to use a fixed stock of resources to achieve higher productivity than would be possible otherwise. For instance, provided that traffic flow conditions are within certain parameters, high-occupancy-vehicle (HOV) lanes can maximize vehicle throughput in a fixed section of roadway.
  - *Cost-based pricing.* Discrepancies between prices and real costs present an opportunity for enhancing social welfare by adjusting prices to better reflect the reality. For instance, introduce congestion tolls, or eliminate cross-subsidies for peak-period utilization of electricity, or remove rent controls.
  - *By-products of personal aspirations.* It is possible to structure new incentives or create new opportunities for personal advantage or satisfaction that can indirectly result in social benefit, e.g., offering to share the savings from cost-reducing innovations with public-sector employees who conceive them and implement them.
  - *Complementarity.* Two or more activities can potentially be joined so that each might make the other more productive, e.g., public works construction and combating unemployment.
  - *Input substitution.* The world abounds in opportunities to substitute less costly inputs in a current production process while achieving roughly equivalent results.
  - *Development.* A sequence of activities or operations may be arranged to take advantage of a developmental process, e.g., assessing welfare clients for employability and vocational interest before, rather than after, sending them out for job search.
  - *Exchange.* There are unrealized possibilities for exchange that would increase social value. We typically design policies to simulate marketlike arrangements, e.g., pollution permit auctions, and arrangements to reimburse an agency for services it renders to another agency's clients or customers.
  - *Multiple functions.* A system can be designed so that one feature has the potential to perform two or more functions, e.g., when a tax administrator dramatizes an enforcement case in such a way as both to deter potential violators and to reassure non-violators that they are not being made into suckers for their honesty.
  - *Nontraditional participants.* Line-level employees of public agencies often have knowledge of potential program improvements that could usefully be incorporated into the agencies' policies and operations. The same is true of the agencies' customers, clients, or the parties whom they regulate. The IPE case of peer-based teaching fits here.
  - *Underutilized capacity.* An example, in many communities, is school facilities that are utilized for relatively limited purposes for only part of the day and for only part of the year—although school officials would be quick to warn that tapping this capacity without harming school functions is not always easy.
- 

Once you start writing up a checklist of such mechanisms, they begin to seem familiar and perhaps obvious if not simple-minded. That, of course, is all to the good. It is also good to have such a list, both for the reason I have already mentioned and for one other. My colleague at the Goldman School, Michael O'Hare, would call a checklist like this "Ideas to have ideas with." It would, as I have said, help you have insight into what is happening at source sites. It could also furnish such capital-I "Ideas" independently of seeing any particular empirical instance of them at all, just as an architect reading a book about forms and materials could imagine building designs without ever having seen an actual example or anything close to it.

### Use “-ing” Verbs

I have distinguished between the features that implement a basic mechanism and the functions that they perform in doing so. But how should one conceptualize these functions? One way is to apply a pre-existing theoretical framework. From my studies of “social regulation” (such as worker safety enforcement), I approach any program that roughly fits this category with an eye toward finding features that perform three central functions: standard-setting, information-gathering, and sanctioning. I also keep a lookout for educating, counseling, and motivating functions.

But not all phenomena roughly match pre-existing templates; and not all pre-existing templates come with an easily conceptualized functional schema. In the more general case, then, the strategy is to look at each separate feature and ask what function it might perform with regard to implementing the basic mechanism. But this naming of functions is not so easy either. Another semantic trick is useful here, namely, the fashioning of verbs ending in “-ing” to describe the function. I have done this above with regard to the functions needed to implement a regulatory program. In the IPE case, in my discussion of the key implementing features I referred to the functions “maintaining a high-arousal environment” and “creating a capacity to use peers as teachers.”<sup>13</sup>

### “COULD WE MAKE IT WORK HERE?”

Besides ignoring it, there are four things you can do with somebody else’s good practice: replicate it, adapt it, experiment with it, or get some further ideas that are inspired by it. The course of action chosen in the target site will depend on local objectives, resources, and available alternatives. With regard to the IPE, faculty and students at the Goldman School considered it long and hard. In the end, we chose not to replicate or adapt it, because we were already running a semester-long project course.<sup>14</sup> That course experience does take a whole semester, but it does also offer the opportunity to work for a real client and to learn to do field work. As mentioned earlier, we also run a curtailed and individualized variant of the IPE in the form of the 48-hour project. We were very taken with the possible bonding effects of the IPE, though, and keep alive the possibility that we might experiment with some variant of the IPE principally for this benefit.

Now suppose we had wanted to replicate or adapt the IPE at the Goldman School. We would have needed, at a minimum, to consider whether our resources were adequate in light of the known and suspected vulnerabilities of the IPE. What guidance might we have extracted from the experiences at our source sites? More to the point, what generalizable method might we have used in doing so?

This is nothing more nor less than a problem of uncertainty reduction in the Bayesian framework. Once one has extracted ideas from one or more source sites, their most important guidance function is over. We enter the general realm of decision-making under uncertainty. Evidence from a source site can help to reduce the

<sup>13</sup> In my *Practical Guide*, p. 79, I had recommended even more explicitly using gerunds. Rossi and his colleagues give essentially the same advice as I do but allow more latitude: “...it is important for the evaluator to carefully identify each distinct program component, its functions, and the particular activities and operations associated with those functions. For this purpose, it is usually most instructive to view the program as a process rather than as an entity and describe it with verbs rather than nouns....” (Rossi, Freeman, and Lipsey, 1999, p. 169)

<sup>14</sup> We also chose not to experiment with it, though that remains an option for the future.

uncertainty about whether some version of the practice would work adequately at the target site and with what costs and risks of unacceptable failure. I shall limit the discussion here to evidence called for by two particular strategies of uncertainty reduction, *a fortiori* analysis and break-even analysis.<sup>15</sup>

### **A Fortiori Analysis**

The *a fortiori* procedure consists of comparing target site and source site with regard to any single resource or some (relatively small) combination of resources that might affect results and then asking:

- “How much trouble did they have, given their level of this (these) resource(s)?”
- “Are we better or worse off than they were with respect to this (these) resource(s)?”

If they had trouble and you are worse off, you have located a danger zone, and you should develop options for getting out of it or else be prepared to take risks.

In the case of the IPE, faculty interest and talent are the crucial resources. I am aware that Michigan has had trouble recruiting suitable faculty for the job, though I am not certain about Harvard. I do know that the Goldman School has a much smaller faculty than either of these institutions, and we would have had a hard time putting together a team adequate for mounting the IPE year after year.

*A fortiori* reasoning also helps in designing a search strategy. If you have the opportunity and the means to do so, go look at source sites that you expect might have more trouble than you would. If the practice “works” there, you can have greater confidence that it will work in the target site.<sup>16</sup>

### **Break-even Analysis**

Though it is greatly underutilized, break-even analysis, sometimes called threshold analysis, is one of the most useful tools in the policy analytic tool kit. It is very helpful when choosing which aspects of the source site experience to probe with special care.

Break-even analysis permits you to frame residual uncertainties about the wisdom of a course of action against two backdrops that are more certain. One such backdrop is the threshold level of success—for convenience, say it amounts to “benefits exceeding costs” or “breaking even”—that you require before going forward. The second is what you can reasonably assume about your situation that contributes to success (or failure) *before* taking into account your residual uncertainties. Your critical questions then become something like these: “What do we have to believe about the residual uncertainties in order to believe that we can achieve our threshold level of success?” and “How reasonable is it to believe these things?”<sup>17</sup> You can look to the experience of source sites for clues that help you answer both of these questions.

<sup>15</sup> I would simply like to note here the applicability of Monte Carlo simulation to the problems discussed here and the likelihood that it is not sufficiently appreciated (Caulkins, 2002).

<sup>16</sup> This is a variant of the proposition that, in general, the more heterogeneous the source sites in which the practice in question has been observed to “work,” the more confidence you are entitled to have about extrapolating to the target site (Shadish, Cook, and Campbell, 2001, pp. 23–24, 376–377).

<sup>17</sup> For a more extensive discussion of break-even analysis, see Bardach, 2000, pp. 29–32.

I will illustrate the procedure with respect to the possibility of the IPE displacing the Goldman School's 48-hour project. Assuming the required faculty resources could in fact be made available, we might estimate the net costs of a one-week IPE with "reasonable" certainty. The main resource costs would be three extra days of student time per student for each of, say, 75 students, and approximately 100 hours of faculty member time and 100 hours of teaching assistant time. (Evidence from the source sites could of course help with these estimates.) The main non-resource cost would be the learning benefits forgone by giving up the 48-hour project. The big residual uncertainty in this example concerns the benefits of the IPE. For the switch to the IPE to make sense—that is, to "break even"—the benefits would have to be big enough to offset these costs. How big would these benefits have to be?

The experiences at Harvard and Michigan would provide the best clues we could get. Our ability to calibrate the magnitude of the learning effects in the source sites with any precision is very limited. However, we do not need to be very precise. All we need to do is to estimate whether the expected magnitude exceeds the break-even level. With scarcely any discussion, Goldman School faculty and students concluded that that it would not. For us to have believed otherwise, we would have had to have observed extremely dramatic signs of support for the IPE from the source sites. We did not. Of course, we might have been wrong. But a combination of break-even analysis and minimal probing for clues at the source sites permitted us to draw what still seems to me a reasonable conclusion in short order.<sup>18</sup>

## SUMMARY AND CONCLUSIONS

Extrapolating from the experience of others is not literally "extrapolation." It is more like searching for interesting ideas about successful mechanisms that might be adapted at home, and then inferring from the experience of others plausible estimates about the chances of success or failure in implementing the adaptation.

I have focused in this paper mainly on the challenge of extracting ideas from what the source sites have done. I have argued that an observer should be aiming to understand the "basic mechanisms" underlying effectiveness and cost-effectiveness of what a source site has done—assuming, of course, that what they have done is indeed effective and cost-effective. The "contingent features" of the observed practice, those that are merely instrumental or optional, might be adapted in the target site without substantial risk. Furthermore, I have argued that, within limits, it is possible, and is certainly desirable, to analyze the vulnerabilities of the observed practice and possible adaptations of it to a variety of hazards.

I acknowledge that these arguments have been pretty dense, and would have been even denser were I doing them full justice. Please resist the temptation to be suspicious of them simply because they are dense. Compared with the actual complexity of the real world, the schema sketched here is very simple. Whether I've simplified the world to the right degree or in the right way are other matters entirely, of course, and surely very debatable. My desire is more to set such a debate going, not to settle it.

<sup>18</sup> Note that the evidentiary problems entailed by extrapolating to a specific target site are less severe than those of generalizing about the effectiveness of a practice in an unspecified population of target sites. This is so because we can make use of site-specific information about the target site to help bound the uncertainties about which we are seeking information from source sites. This point may not be sufficiently appreciated in the evaluation literature.

Another temptation to be resisted involves the lack of interest on the part of real-world decision makers in an analysis of others' practices that aims to go this deep. Referring to replication of the practices of the winners and finalists in the Innovation in American Government program, Gail Christopher (2003, p. 687) writes in the *JPAM* symposium on best practice compendia: "Results and simplicity in design have been and continue to be the catalysts for adoption by other jurisdictions... Decisionmakers seem to be less concerned with why a program works than with results."

True, Christopher is writing about replication, which doesn't require a very deep understanding of the innovative practice in question, but I suspect her point applies to a degree even when adaptation is more likely. Decision makers, when they send to learn "what they're doing over there that might make sense for us," don't usually want, by way of an answer, to hear metaphors about mechanisms and reservoirs of energy and nature's free gifts. And of course they ought not to hear such highfalutin' talk. Their policy analysis staff or consultants must translate their own more rarefied understanding back into ordinary language. We are used to this translation requirement when it comes to quantitative analysis. It should not be too surprising that good conceptual analysis would labor under a similar requirement.

But the main point is that there ought to be something worth translating. Remember the architect and the client on their home-and-garden tour. The client sees the surface reality while the architect sees that plus several deeper levels of the same reality. That is why she is being paid to come along on the tour. She is not usually being paid to explain the deeper levels of what she sees, much less to explain how it is that she manages to understand what she sees. But unless she could see and understand the deeper reality, she would not be worth much to the client. So too with policy analysts and decisionmakers. Analysts that insisted on trotting out the full and unmediated story of how they have learned from the experiences of others might not be very welcome. But the more they could do it if called upon to do so, the more valuable their products will be.

I am grateful to Nancy Bardach, Rebecca Bardach, Michael Barzelay, Tom Cook, Frank Levy, Rob MacCoun, Michael O'Hare, Eric Patashnik, Beryl Radin, and Buzz Breedlove for their helpful comments. I, of course, bear full responsibility for the final product.

*EUGENE BARDACH is Professor of Public Policy at the Goldman School of Public Policy at the University of California, Berkeley, and President of APPAM in 2003.*

## REFERENCES

- Bardach, E. (1998). *Getting agencies to work together: The practice and theory of managerial craftsmanship*. Washington, DC: Brookings Institution.
- Bardach, E. (2000). *A practical guide for policy analysis: The eightfold path to more effective problem solving*. Washington, DC: CQ Press.
- Bardach, E. (Ed.) (2003). Creating compendia of best practice. *Journal of Policy Analysis and Management*, 22, 661–688.
- Brady, H.E. (2003). *Models of causal inference: Going beyond the Neyman-Rubin-Holland theory*. Unpublished ms., Survey Research Center, Univ. California, Berkeley.
- Caulkins, J.P. (2002). Using models that incorporate uncertainty. *Journal of Policy Analysis and Management*, 21(2), 486–491.
- Christopher, G.C. (2003). Innovative government practice: Considerations for policy analysts and practitioners. *Journal of Policy Analysis and Management*, 22(4), 683–688.

- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
- Elster, J. (1998). A plea for mechanisms. In P. Hedstrom & R. Swedberg (Eds.), *Social mechanisms: An analytical approach to social theory* (pp. 45–73). New York: Cambridge University Press.
- Hedstrom, P., & Swedberg, R. (1998a). *Social mechanisms: An analytical approach to social theory*. New York: Cambridge University Press.
- Hedstrom, P., & Swedberg, R. (1998b). Social mechanisms: An introductory essay. In Hedstrom & Swedberg (Eds.), *Social mechanisms: An analytical approach to social theory* (pp. 1–31). New York: Cambridge University Press.
- Kruger, J.H. (2002). Best practices as found on the Internet. In E. Oyen (Ed.), *Best practices in poverty reduction* (pp. 108–129). London: Zed.
- Mackie, J.L. (1993). Causes and conditions. In E. Sosa & M. Tooley (Eds.), *Causation* (pp. 33–55). New York: Oxford University Press.
- Rossi, P.H., Freeman, H.E., & Lipsey, M.W. (1999). *Evaluation: A systematic approach*, 6th ed. Thousand Oaks, CA: Sage.
- Schelling, T.C. (1998). Social mechanisms and social dynamics. In P. Hedstrom & R. Swedberg (Eds.), *Social mechanisms: An analytical approach to social theory* (pp. 32–44). New York: Cambridge University Press.
- Shadish, W.R., Jr., Cook, T.D., & Campbell, D.T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Wilson, J.Q. (1967). The bureaucracy problem. *The Public Interest* (Winter), 3–9.