EVALUATING CRIMINAL JUSTICE REFORMS

JONATHAN D. CASPER*
DAVID BRERETON

Research on the implementation of criminal justice reforms, such as sentence guidelines, determinate sentence laws, and no plea bargaining policies, has proliferated in recent years. This research raises a variety of concerns about how one should evaluate the effectiveness of policy innovations. We deal here with recurring issues in the evaluation of criminal justice innovations, including the problem of specifying goals against which to measure "effectiveness," the need to interpret research findings in light of theories about how criminal courts operate, and the importance of choosing appropriate time periods in which to conduct implementation or evaluation studies.

I. INTRODUCTION

Research on the "implementation" of policy has proliferated in recent years. The study of implementation (sometimes called "evaluation" research) has focused upon a diverse and rich set of policy issues, including education (Murphy, 1971), employment (Pressman and Wildavsky, 1973), mental health (Bardach, 1977), decisions of the Supreme Court (Wald, 1967; Muir, 1973; Peltason, 1961), as well as a number of innovations in criminal justice policy to be discussed here.

The implementation literature has often assumed that policy failure or success can be ascertained with relative ease and that the major intellectual task is to identify the crucial intervening variables that shape policy effectiveness. In this respect, this work has retained a flavor reminiscent of the old policy-administration literature, in which it was assumed that goals, means, and ends could be neatly separated. However, recent more sophisticated approaches to policy analysis

* For helping us to think about these issues and for suggesting several of the ideas argued here, we are indebted to Kennette Benedict. Milton Heumann, Michael Gottfredson, and Peter Nardulli provided helpful criticisms of previous drafts. Richard Lempert made valuable substantive and editorial contributions. The research was supported by the National Institute of Justice (79-NI-AX-0042). All responsibility for the arguments and conclusions rests with the authors.

LAW & SOCIETY REVIEW, Volume 18, Number 1 (1984)

suggest that concepts such as policy success or failure are in fact highly problematic, not only because effects may be difficult to isolate and measure but also because the intent of the policy itself is often difficult to ascertain (Van Meter and Van Horn, 1975; Mazmanian and Sabatier, 1981). Insofar as this is the case, the consequences for policy-oriented research are significant. For one, much of this research focuses upon real, live programs in which substantial amounts of money and other resources are invested. How these programs are characterized by those analyzing and evaluating them may thus have significant consequences for the future deployment of these resources. More important, we need to develop useful theories about the conditions under which policy innovations are, or are not, likely to be effective. Without a clear and suitable way of characterizing concepts like effectiveness or ineffectiveness, success or failure, etc., such theoretical advancement is unlikely.

Our purpose here is to identify some of the major problems involved in talking about policy implementation in terms of success or failure and to suggest some ways in which these problems might be overcome. In the first section, we deal with the difficulties of selecting the referents against which innovations should be evaluated. Here we discuss the complexity and ambiguity that inheres in the notion of goals and indicate why a concern for goals or purposes makes the evaluation of reforms so difficult. In the next section we examine the problems involved in measuring and interpreting impact. First, we suggest that much of the research on criminal justice innovations has not taken its own theory seriously enough. Small changes in an expected direction are often interpreted as evidence that an innovation has "failed" when existing theory about criminal court behavior suggests that such changes may be quite consistent with "success." Next, we argue that much research in the criminal justice area has often failed to employ effective research designs. Two-point before-and-after research designs have predominated, even though the literature on research design is replete with warnings that such designs are likely to produce misleading results. We explore the implications of such designs, some reasons why they are so common, and some alternative strategies. In both sections we illustrate our arguments with examples from the literature on criminal justice reforms as well as from our own recently completed study of the implementation of the California Determinate Sentence Law (Casper *et al.*, 1982).

II. THE GOALS OF POLICY FORMULATORS

In order to assess the extent to which a policy has met its objectives, we must begin with some means of ascertaining what these objectives are. The standard approach to this problem, both in textbooks on policy evaluation and in much of the research literature, is to examine the policy document (i.e., the statute, administrative rule, court opinion, etc.) as well as any public statements made by key supporters of the initiative. It is assumed that the general purpose or intent of the policy can be derived from this examination, thus allowing the investigators to get on with the main business at hand, the identification and explanation of divergences between actual and intended effects. In practice, however, discerning intent is often not easy, especially where major innovations are involved.

Manifest v. Latent Goals

A basic problem is that there may be a significant gap between what can be termed latent and manifest intent, i.e., between the actual policy goals of those who help formulate or adopt an innovation and the goals that are "self-evident" from the face of the legal documents commanding the innovation or in the public discussion surrounding its adoption. Such a disjunction may result, for example, when policy-makers attempt to achieve their goals of modifying outcomes by a strategy of indirection, as suggested in Bardach's study (1977) of mental health policy in California. Bardach notes that implementors often engage in adaptive behavior designed to serve their own goals and institutional or personal needs. Policy-makers, he argues, should realize this and should engage in "scenario-writing"—imagining the possible responses of implementors to an innovation and attempting in advance to counter them in order to achieve desired results. But if they do this, the policies they convey to implementors may not reflect their true intentions.

The California Determinate Sentence Law (DSL) offers an example of such a divergence between manifest and latent intent. Law-enforcement supporters of the DSL wanted to increase the prison commitment rate, but they eschewed the direct route toward this goal, which would have been passage of a large-scale mandatory-minimum sentence law. Their reluctance stemmed in part from the likelihood that such bills would be strongly opposed by liberal elements in the Assembly as well as by judges who sought to retain discretion. Instead

those interested in imprisoning more defendants adopted a "gaming" strategy. Judges, they reasoned, were reluctant to sentence "marginal" defendants[1] to prison terms because of the apparently very long terms required by the Indeterminate Sentence Law (e.g., first-degree burglary required a term of one-year-to-life; second-degree burglary, a term of one-to-fourteen years). The scenario "written" by law-enforcement interests was that if sentences were made shorter and more certain, judges would respond to the change by sending more people to prison.[2] Thus, a statute that on its face may appear to be designed to promote equality in sentencing and ensure that criminals don't serve excessive sentences may, in the minds of those who were most instrumental in its enactment, be designed to get judges to sentence more criminals more severely, i.e., to prison rather than jail.

To the extent that policy-makers are proceeding by such strategies of indirection, it is potentially quite misleading to infer the intent of an innovation or of its supporters on the basis of the policy itself. From examination of the California Determinate Sentence Law, for example, one would expect that "successful" implementation has little or nothing to do with the prison commitment rate, for the law is virtually mute on this subject. Yet from the perspective of its law-enforcement supporters, this was the key to its success or failure.

Manifest and latent aims may also diverge when symbolic issues are paramount (see, generally, Arnold, 1935; Edelman, 1964). Often the rewards offered to constituents involve not substantive changes in the distribution of costs and benefits but largely symbolic reassurance that needs are being attended to, problems are being solved, help is on the way, etc. From the perspective of policy-makers, symbolic actions may increase popular support and quiet opposition, and these advantages may accrue from the adoption of the new policy almost regardless of its actual implementation.

Nowhere does symbolic politics seem so prevalent as in the area of criminal justice, where it has long infused our "bark"

---

[1] A "marginal" defendant is one who is viewed by a sentencing judge as on the borderline between a long jail term (typically near the maximum of twelve months) and a commitment to prison. For example, a defendant convicted of a property offense with a substantial prior record but no prior prison terms is often viewed as on the margin between jail and prison.

[2] The terms of the new DSL surely *looked* shorter than the old ISL terms: for second degree burglary, the one-to-fourteen-year term was replaced by a selection from sixteen months, two years, or three years; for strong-arm robbery, one-to-life was replaced by two, three, or four years. Whether in fact the actual time served was to differ markedly was not entirely clear at the time the law was passed.

and "bite" sentencing policy. The "bark" is the long nominal terms that legislators impose for illegal behavior, knowing full well that the actual "bite" will be substantially reduced by the activities of judges, parole authorities, and others. Legislators have thus been able to appear very tough on crime without having to face up either to the costs that the actual imposition of nominal sentences would entail in terms of prison construction or to the issue of whether the Draconian penalty structures that characterize our criminal law are just.

Issues of personal safety arouse considerable concern, and fear of victimization has grown in recent years (Hindelang *et al.*, 1980: 174). Moreover, the belief that crime can be dealt with by changes in punishment structures appears to be quite widespread. As a result, the potential for symbolic politics is very high. This seems particularly true of mandatory-minimum sentence innovations which require that defendants convicted of certain crimes be sent to prison. Statutes which impose special prison terms on those carrying firearms (Beha, 1977; Heumann and Loftin, 1979; Pierce and Bowers, 1981; Loftin *et al.*, 1983) or those selling drugs (Joint Committee on New York Drug Law Evaluation, 1977) or those judged to be "career criminals" are typically supported on a variety of grounds, including incapacitation of the dangerous and enhanced deterrence. The actual enforcement of such statutes—the incarceration of all or most who are convicted of carrying guns or selling drugs or with substantial criminal records—would be extremely costly given prison capacity and prior sentencing practices which have treated such offenses less seriously. Moreover, given what we think we know about the causes and prevention of crime, it is questionable whether mandatory-minimum sentence laws—or any activities of the criminal justice system alone—are going to have any appreciable effects on illegal behavior.

Nonetheless, the passage of such legislation provides something very important to constituents—a sense that policy-makers care about their fears and are taking steps to make society safer. Moreover, the adoption of such innovations will cost little in material terms if they are not carried out (a mandatory-minimum sentence law which is generally evaded by charge bargaining, for example, will not require increased prison construction) and offend no powerful counter-constituency.

Concerns of symbolic policy-making and the difficulties it creates for implementation research appear in other criminal

court reforms as well. Prosecutors or judges who urge the abolition of plea bargaining or claim to have achieved this result are often seeking public support based on the popular belief that such a reform will reduce leniency. "Voluntary" sentence guideline systems, in the same fashion, may be useful in promoting public support for apparently less discretionary and more even-handed sentencing patterns.

The pay-offs received by policy-makers for all of these innovations largely accrue from the *passage* of legislation or the *formal adoption* of a new policy rather than from its actual translation into behavioral change. The latter is often difficult to discern, and if it is discerned that the policy has not had its putative effect (e.g., reducing crime, incapacitating larger numbers of dangerous criminals, decreasing sentence disparity, etc.), it is always possible to adopt new measures. Most such innovation is not, for example, accompanied by the development of monitoring mechanisms or evaluation research to report back to policy-makers about its implementation. One must inquire carefully into the latent purposes of policy innovation and should not assume, without good evidence, that policies that are likely to be costly in terms of resource allocation were indeed *intended* to achieve their manifest goals.

Coalition-Building and Lack of Clarity in Goals

In other instances, the problem may not be so much that latent and manifest intent diverge but that intent at either level is difficult, if not impossible, to discern. Thus, Nakamura and Smallwood (1980) suggest various reasons why policies may lack the clarity necessary for implementors to be sure what they are "supposed" to do or for evaluators to discern whether programs have been carried out "successfully." They discuss technical limitations (e.g., the fact that policy-makers may not know how to achieve agreed-upon goals); conceptual complexity (largely a result of conflicting goals among supporters of a policy); and coalition-building (which tends to produce vague policies in order to broaden the base of support).

The adoption of the California DSL illustrates several of these issues about the character of goals and their use in implementation research. The coalition structure supporting movement from indeterminate to determinate sentences brought together individuals and interests that generally disagree about criminal justice policy. On one side were law-enforcement interests (e.g., associations of district attorneys,

police officers, etc.) and their legislative allies, who generally favor more punitive sentencing policy, increased certainty of punishment, and fewer procedural rights for those accused of crime. These interests had traditionally supported the Indeterminate Sentence Law (ISL) because its nominally high maximum terms gave extensive control over prison inmates to the parole authorities, favored such reforms as mandatory-minimum sentence laws, and opposed decriminalization of consensual crimes. On the other side was a loose-knit coalition of interests (including the ACLU, prisoner support groups, etc.) that had traditionally controlled the California Assembly Criminal Justice Committee. These groups had favored the ISL because of its association with the goal of rehabilitation in prisons, supported decriminalization of consensual crimes, and opposed extensive use of mandatory-minimum sentence laws.

During the early 1970s, both the law-enforcement and civil rights interests became persuaded that a move to determinate sentencing was desirable. Law-enforcement interests desired greater incarceration rates and believed that if judges were provided with what appeared to be shorter, determinate sentences, they would send more "marginal" defendants to prison. Liberal due process interests supported the move to determinate sentencing for quite different reasons. They had been persuaded by the "nothing works" literature that prison programs could not be expected to rehabilitate (Martinson, 1974). Without the rationale of rehabilitation, the medical model of indefinite sentencing and parole release lost most of their attraction, and negative features of the system such as the allegedly arbitrary exercise of discretion by the parole authorities grew more salient. Thus, the innovation of determinate sentencing was for liberals and conservatives a "solution" to quite differently conceived problems.

Given the coalition structure that emerged, the DSL was passed with relative ease in 1976. But the coalition structure means that determining what the DSL was "supposed" to achieve or evaluating its effects in terms of the goals of its supporters is difficult. If the DSL were to produce a higher prison commitment rate or longer terms, it would clearly be judged a success by law-enforcement interests, even though its liberal due process supporters would not count such outcomes as desirable. By the same token if it were only to decrease sentence disparity, its law-enforcement supporters would not hail its accomplishments. Thus, those who cooperated to design and pass the DSL anticipated and desired a variety of

effects which, if not mutually exclusive, were neither mutually entailed nor part of a coherent, shared agenda.

California is not unique in this respect. Looking at Minnesota, Martin (1983: 271) notes coalitions that brought together groups with similarly disparate goals and expectations about what the innovation might produce:

> The [Minnesota sentence] guidelines alternative appeared to offer a rare acceptable compromise between fiscal conservatives and corrections liberals. It promised the most important changes or provisions that several key interest groups had sought, met other goals of those interest groups or offered them a share in decision-making. . . . Police and prosecutors had sought and won greater influence in shaping the sentencing decision and more predictable sentences for the "worst" offenders. The judiciary got structured discretion over sentence lengths rather than no discretion over them. To the corrections bureaucracy and the defense bar, less concerned with discretionary authority than with warding off increased severity, the guidelines seemed to offer a better prospect than legislatively set flat-time sentences.[3]

The examples of California and Minnesota, as well as others in the literature on sentence reform, show how the coalition-building process may mean that supporters of a policy have disparate or even conflicting manifest and latent goals. When this occurs, assessing the policy's effect in terms of its purposes is fraught with difficulty. Canvassing statutory language or legislative history in search of the policy-makers' goals does not resolve the problem. These sources, for reasons we have identified, can give a distorted sense of what purposes were important to those who participated in the adoption of the policy. The problem is compounded when there are multiple stages of formulation—as when a legislature delegates the construction of sentence guidelines to a commission—and it is not clear *whose* goals should be examined.

When the policies underlying a reform are conflicting or ambiguous, this is often a signal that the coalition that

[3] The development of the Minnesota sentence guidelines also illustrates another problem in using intent or goals as benchmarks for assessing implementation. In Minnesota, the crucial issue of formulating the guidelines (deciding what criteria to use and how they were to be weighted) was delegated by the legislature to an appointed commission. Assessing the "success" or "effectiveness" of the eventual guidelines requires one to confront the intentions not only of participants in the legislative process but also of those members and staff of the commission who formulated the actual guidelines. Such delegation—and the attendant issue of multiple stages and participants whose goals must be taken into account—is a common phenomenon.

supported the reform is weak. The weakness may be so pronounced that the coalition does not survive until the time of the evaluation. To the extent that this happens, research on implementation must be sensitive to changes both in the innovation itself and in the political climate surrounding its implementation.

In California, for example, the coalition that supported the DSL was well on the way to disintegration by the time the law went into effect. The law was passed and signed in September, 1976, to go into effect on July 1, 1977. By the time of its effective date, several amendments had already been passed to increase the statutory penalties. Within a couple of years additional bills had been passed raising prison terms, and it was clear that the law-enforcement interests had gotten control of the policy-making process. Most legislators may have preferred a definite five-year term for robbery even though the pre-existing ISL had imposed a term of one-year-to-life. But once the principle of determinate sentencing was adopted, given popular concerns, it was difficult to resist the claim that a definite term of ten years in prison was preferable to a definite five. After only a few years of determinate sentencing, the due process liberals were looking for a way to modify sharply the innovation they had so warmly supported a short while before. From their point of view the "intended" effects of the legislation had, via the amendment process, already been greatly changed.

Examination of original coalition structures might indicate when such shifts are a likely occurrence, and can sensitize researchers to short-run changes that are important in understanding and evaluating the implementation process. More generally, analysts must often choose the point in time at which an evaluation should identify intent. This choice more than anything else may determine whether an innovation is regarded as successful, since at both manifest and latent levels there may be continuing redefinitions of what an innovation is "supposed" to achieve. (See Pressman and Wildavsky, 1973: Ch. 9.)

The burden of this discussion is that evaluating the effectiveness of a policy innovation is a difficult enterprise. Given the possibility of manifest and latent purposes and changes in goals over time, and the fundamental difficulties in discerning the purposes of innovations, evaluators must proceed with great care.[4] At a minimum, it generally makes

[4] A conceptually different and curious type of innovation for which implementation research is difficult involves policies which are quite clear but

sense to evaluate the effectiveness of a policy in terms of the expectations of its supporters rather than on the basis of the language of the formal rule or law. Moreover, where the supporting coalition is characterized by competing objectives, the appropriate procedure may be to ascertain effectiveness vis-à-vis each set of objectives. In such circumstances, all that can be ascertained is whether particular supporters were successful or not, and the interesting question is why one set of interests came to dominate. Finally, a case can be made that, where intent is obscure, the best approach is to evaluate the innovation by the use of independently derived criteria. Here researchers decide what they think a policy *should* be doing and then seek to judge how well it does it. For example, they may choose to assess a policy by its impact on crime reduction, whether or not this was the original purpose of the policy. Such an approach requires researchers to make clear that it is their criteria, not those of the policy-makers, that are providing the benchmark. This last strategy may, of course, be used even when the policy-makers' goals are clear and consistent, but in these circumstances one will almost always want to use the policy-makers' goals as an initial benchmark.

III. ASSESSING IMPACT: TAKING THEORY SERIOUSLY

Assuming that it is possible to select some kind of referent against which to judge the extent to which an innovation has succeeded, the researcher must then deal with the problems of measurement and characterization. The measurement problem is the technical one of assessing the amount and type of change which can be attributed to an innovation. Characterization involves deciding on the basis of the evidence whether the policy can be spoken of as a success or failure, as worthwhile or not worthwhile, and so on. In the discussion to follow, we

---

do not impose many restrictions upon behavior. In the Massachusetts sentencing guidelines, for example, the ranges in permitted sentences were on the order of ± 50%. Thus, a guidelines sentence might be 10-30 months in prison. Such a policy is clear, for the judge is unambiguously instructed that certain classes of defendants shall receive sentences within specified ranges. But because the range of behavior specified is so broad, "compliance" will almost inevitably occur, and the policy is very likely to be implemented, even though it may not discernibly change prior behavior. Similar problems are posed by sentencing innovations like the one in Pennsylvania that specify narrower ranges yet allow explicitly for departures based on a variety of factors. Disagreement among members of a supporting coalition may produce such specific-yet-empty policy innovations. In the guidelines area, for example, reforms like these may result from a "compromise" between those who prize judicial discretion and those who favor some structuring of this discretion. Because each side gains something it can point to, both are "satisfied," though the policy may do little to change prior sentencing practices.

will focus on the latter issue and then briefly examine some of the problems involved in measuring impact.

One of the most intriguing aspects of implementation research in the criminal justice area is the way in which research results are related or not related to existing theories about the operations of criminal courts. Much implementation research slights its own theoretical underpinnings.

Most recent investigations of criminal courts find that they consist of relatively stable workgroups whose members—judges, prosecutors, defense attorneys, probation officers—work regularly with one another and develop patterns of behavior that serve both individual and institutional needs. Among the most important kinds of patterned behavior are going rates (shared beliefs about appropriate sentence levels for defendants charged with given crimes who possess similar records) and the negotiation routines that lead to guilty pleas in most cases. Standard operating procedures which regularize sentences and routinize plea bargaining may develop for a variety of reasons. Once developed, they continue because participants *believe* they are desirable or necessary, and they become deeply embedded in the beliefs and values that make up the courtroom workgroup culture (Heumann, 1978; Mather, 1979).

This workgroup theory typically informs most implementation research that looks at criminal courts. Implementation of an innovation involves an "outsider"—e.g., a legislature, a supervisory judge or prosecutor—attempting to alter the behavior of members of the courtroom workgroup. A mandatory-minimum sentence law, for example, tells the workgroup that a class of offenders who in the past may have only sometimes been sent to prison shall now *all* receive terms in prison. A determinate sentence law may specify terms for certain types of defendants that are longer or shorter than they were in the past. A "ban" on plea bargaining may forbid charge or sentence concessions that were common in the past.

Most evaluations of such innovations conclude that they are not particularly effective.[5] Usually we are told that the innovation has been ignored or that the participants developed adaptive strategies that enabled them to comply in formal terms without substantially changing their behavior. The

[5] The major exception is the Alaska experience in abolishing plea bargaining. See Rubinstein and White, 1979. While other studies may underestimate the effects of innovations, the Alaska research probably errs in the other direction. For a critique, see Cohen and Tonry, 1983.

moral drawn is that it is difficult (sometimes the suggestion is that it verges on the impossible) for outsiders to affect markedly the behavior of court participants.

This conclusion may often be misleading. When an implementation study examines behavior for a (typically) short time after an innovation and finds only slight changes of the kind the innovators appear to have desired, the minimal movement is often characterized as resistance to change or evasion. Evaluators generally ignore the possibility that small observed changes may be the first stage of what is likely to become over time a significant change in behavior. Small initial changes might well be taken as a sign of innovative failure if theory led us to expect immediate and dramatic changes when courtroom participants adapt to new laws or authoritative orders. But when our theory predicts that workgroup members will *resist* change, small changes in short periods are surprising and may reflect a substantial force for change. The question is whether the force is exhausted in the small change or whether it will eventually transform the system in a way that accords with the intent of the policy-makers. The latter possibility is seldom appreciated, much less addressed. To exaggerate a metaphor, it is a matter of whether to view the glass as 19/20 empty or 1/20 full. Too often, the glass-is-empty conclusion seems to be reached when the theory guiding the research suggests that the moisture on the bottom of the glass is the unexpected finding. To be sure, those policy-makers who expected swift and large-scale behavioral change will be inclined to view such initial small changes as inadequate and indications of evasion and implementation "failure." Analysts who take their own theory seriously, though, might do well to stress the change they *do* observe.

For example, Church (1976) studied a plea bargaining ban in "Hampton" County in which a new law-and-order prosecutor forbade his deputies to engage in charge or sentence bargaining in drug cases. He found that in the year or so after the ban was imposed, the number of drug defendants pleading guilty to reduced charges fell virtually to zero, indicating formal implementation of the ban. Yet he also observed substantial adaptive behavior designed to maintain the rate of guilty pleas and the going rate for drug cases. There was evidence of earlier screening and the rapid dismissal of trivial cases. More important was the resurrection by some judges of sentence bargaining. Since the prosecutor could not offer charge concessions, some of the judges took up the slack and offered

cues to defense attorneys about likely sentences if the defendant pled without a charge reduction. Church interprets his results as indicating that plea bargaining is, indeed, difficult to abolish given the incentives to engage in it and the availability of adaptive strategies:

> Given equally "resourceful" attorneys, prosecutors, and judges elsewhere, it is unclear how any fundamental shift away from bargain justice could occur without an even more fundamental change in the incentive structures of the participants (1976: 400).

Examination of Church's data suggests the possibility of a somewhat different conclusion. His data on the adaptation involving judicial sentence bargaining is anecdotal, and he notes that only about "half" of the judges engaged in the practice. Moreover, he asserts that judges who were "strict" sentencers experienced substantially greater increases in trial rates than those who were lenient sentencers (a finding which Church interprets as indicating that strict judges were unwilling to offer sentence concessions). Thus, some judges apparently did not begin sentence bargaining; moreover, Church's quantitative data show that charge bargains disappeared in the "after" period. Taken together, the evidence suggests that the ban *did* reduce the incidence of plea bargaining. Clearly it did not eliminate the practice, and Church's analysis provides a neat example of the adaptive process. The issue is one of perspective. If one begins with the hypothesis that plea bargaining is an unimportant aspect of the case disposition process, one might expect that an order to "ban" it will produce quick compliance. If, on the other hand, one begins with the hypothesis that most courtroom participants regard plea bargaining as a crucial aspect of keeping up with heavy caseloads and believe that "just" sentences are arrived at by negotiation, one would not expect an "outsider's" order to cease the practice to result in much short-run change. Since the latter perspective is generally accepted by students of criminal courts, the results observed by Church are consistent with what our theory of court behavior suggests might occur in the early stages of a behavioral change that culminates in the demise of plea bargaining. One must await further development before finally pronouncing on the efficacy of the change.

A similar pattern of analysis characterizes Heumann and Loftin's (1979) perceptive study of the implementation of the Michigan Felony Firearms Statute. This statute instructed judges that all defendants who were found to have carried a

firearm in the course of a felony (a) were ineligible for probation and (b) should receive two years in prison in addition to whatever term the judge chose to impose on the underlying count. Thus, an armed robber who before the innovation would have received five years should following the "gun law" serve seven years, five for armed robbery plus two years for carrying a firearm. In Detroit, the jurisdiction studied, the district attorney followed through on the legislature's intent by announcing that all defendants who carried guns were to be charged under the law and that plea bargaining which involved dropping the gun allegation was forbidden. Heumann and Loftin argue persuasively that these policies were followed by deputy prosecutors.

In a first and avowedly tentative evaluation of the implementation of the new law, they compare dispositions in the six months before and six months after the law went into effect. Their conclusion, based on statistical evidence as well as interviews with participants, is similar to Church's:

> [T]he experience with cases completed during the six months after the intervention of the Gun Law indicates that there has been only a slight upward shift in the average sentence. Clearly there has been no massive increase in the number of cases that receive a sentence of two years or more (1979: 415-16).
>
> . . . [T]he system managed to digest the two policy innovations without a radical alteration in its disposition patterns. Court personnel suspected as much: time and again in the interviews, they indicated that somehow the system would accommodate itself, that things would work themselves out without any major departures from past practice (1979: 426).

Heumann and Loftin not only examine the implementation of the law but also provide a perceptive account of two of the central themes of the criminal court literature in general: the notion of going rates and the crucial role of plea bargaining in facilitating the institutional and personal goals of participants. Their discussion of adaptive strategies in Detroit—which like Hampton County involved both earlier screening of cases as well as a form of judicial sentence bargaining—provides a framework for the expectation that an innovation like the Felony Firearms Statute would be likely to encounter resistance from courtroom participants.

Resistance was to be expected because the law touched upon things that participants held dear: their ability to keep up with caseloads through settled patterns of sentencing and their sense of what was "just." For the most common type of case

Table 1. Prison Commitment Rates Before and After Implementation of Michigan Felony Firearms Statute*

Proportion of Convicted Defendants Receiving Prison Terms

	Felonious Assault	Other	Armed Robbery	Total
Pre-Law	31% (65)	57% (155)	95% (321)	76% (541)
Post-Law	37% (19)	82% (21)	98% (81)	85% (129)

*Proportion of All Defendants Charged Under Law Receiving Prison Terms***

	Felonious Assault	Other	Armed Robbery	Total
Pre-Law	14% (145)	37% (240)	64% (471)	48% (856)
Post-Law	18% (39)	42% (53)	60% (136)	48% (228)

*Table derived from Heumann and Loftin (1979) Table 3.
**Includes defendants who received acquittals and dismissals.

affected by the law, armed robbery, relatively little adaptation was required, for prison terms were typically imposed prior to the innovation. In some cases though—particularly felonious assaults, which were often neighborhood disputes in which a gun was discharged but no one was injured—prison was uncommon under the old law. Compliance with the Gun Law in these cases would require substantially harsher sentences and perhaps reduce the number of guilty pleas.

Table 1 summarizes the data Heumann and Loftin present on sentencing behavior before and after implementation of the new law. Among those charged with armed robbery, the overall prison rate actually went down slightly (because of a small increase in dismissals), and among those convicted (i.e., those the law directly affected) it went up only slightly. For felonious assaults, the overall prison rate (including cases resulting in dismissals) went up 4 percent, and the rate for those convicted went up 6 percent. For crimes other than felonious assault or armed robbery, the incarceration rate went up markedly among convicted defendants but only slightly if dismissals were included. Overall, the rate remained constant for all cases and went up around 9 percent for those convicted. Heumann and Loftin interpret their data as indicating the ineffectiveness of the new law, a "slight upward shift" but no "massive increase in the number of cases that receive a sentence of two years or more" (1979: 416).

Although this characterization is correct, the data support another interpretation. Given the expectation, which the authors share, that courts will strive to resist external pressures to change their ways of doing business, it would have been truly surprising had there been a "massive shift" in sentencing practice in a period of six months. Taking all cases into account, one might argue that the law had, as one would expect, a slight marginal impact on sentencing in the initial six months of its implementation, since sentencing patterns were nudged in the direction one might expect given the manifest purpose of the statute. The direction and magnitude of the change is consistent with a theory which says that courts have settled ways of doing business and will be resistant but not impervious to change. While the law did not produce the instant results some of its sponsors may have envisaged (and to this extent, by their standards, it may be properly termed a short-term failure), this does not mean that it has had no effects or that its effects over time will not be great.[6]

An example of the potential differences between short-run (which can be a matter of years) and long-run impacts is found in changing patterns of compliance with the *Miranda* decision. Early studies show that the initial response (which lasted for a matter of *years* in some of the jurisdictions studied) was in many areas not to give the full warnings, or to use evasive strategies, or to give the warnings in bureaucratic or threatening tones which suggested that they were not to be taken seriously (Wald, 1967; Medalie *et al.*, 1968). Over time, however, the warnings have been integrated into police practice. An implementation study of the period six months or a year after the decision might have argued—if it ignored its theory about how the police culture resists interference by outsiders—that the small numbers of cases in which all the warnings were given provided evidence that the decision would never be implemented. But small shifts in the desired direction were probably the most that could be expected given the novelty of the Supreme Court's command and the hostility of its audience.

The important lesson of *Miranda* is that an initial small effect may be part of a ratcheting process. A "going rate" can gradually move upward in response to a mandatory-minimum

[6] Indeed, focusing on all cases may understate the law's short-term impact since Heumann and Loftin's sample is numerically dominated by cases of armed robbery. In these cases, as we have pointed out, compliance with the new law would not be expected to increase substantially the incarceration rate.

sentence law, or certain types of plea concessions may become increasingly less likely in certain types of cases. Over time, new participants in the courtroom culture are socialized in an environment with a "new" going rate and a trend towards an even harsher one. Thus, "successful" implementation may involve gradual behavior change.[7]

To be sure, gradual processes are different from sudden ones. One important difference is that gradual processes often get short-circuited at some point. Prisons become overcrowded or caseloads mount and cues are received that sentencing ought to be moderated. A new prosecutor takes office and a plea bargaining ban is renounced formally or the word goes out that it is no longer a matter of serious concern. What this means is that small responses in the "expected" direction are simply difficult to interpret. They may indicate resistance that will never be overcome, they may reflect a small "blip" of compliance with the new policy which quickly decays to the old rate, or they may be the first stages in a process of major behavioral change. Doing implementation research and characterizing outcomes requires taking seriously those theories that explain how target populations work. It also requires patience and a willingness to return on several occasions to the scene being studied.

Heumann and Loftin's recent work nicely illustrates such a strategy, for they have published a further account of the implementation of the Michigan Felony Firearms Statute which takes advantage of substantially longer pre- and post-innovation periods (Loftin *et al.*, 1983). Their more detailed data tend to support the conclusions they reported earlier: overall the Gun Law has had no striking impact on the proportion of defendants imprisoned or upon the length of terms imposed during the first two years of its implementation. The later data also support the proposition that in the case of serious crimes plea bargaining and parole practices were able to absorb the additional two-year sentences required by the law (although they report anecdotal evidence of a possible increase in terms in the most serious cases). In the least serious cases (felonious assault, in which the prior going rate did not include prison) the previously-observed pattern of increased dismissals, bench trials with a misdemeanor conviction, etc. also continues. But at the same time there is some increase in the average length of incarceration given

[7] Heumann and Loftin make reference to a similar idea, calling it a "trickle-up" effect (1979: 424).

conviction, which is consistent with a ratcheting or trickle-up effect.

Heumann and Loftin's strategy of data gathering over a substantial period of time lends credence to their findings in addition to providing sufficient data to support their assertions about the *process* by which adaptation occurs. They are also aware that longer time periods may bring further changes in the "expected" direction, as the trickle-up process may continue to work.

Thus, taking theory seriously not only has implications for research strategy (e.g., issues of waiting "long enough" and the use of appropriate research designs, discussed below) but also for the development of hypotheses about *how* implementation may proceed. The deeply embedded disposition and sentence routines encountered by attempts to change the behavior of courtroom participants mean that small changes in the short run *may* be highly significant. They are, to be sure, small and may disappoint policy-makers bent upon rapid reform. But they may, indeed, be the most that can be expected if courts and other operating agencies are as dominated by inertial forces as much of our theory suggests.

IV. APPROPRIATE TIME PERSPECTIVE IN IMPLEMENTATION RESEARCH

The issue raised in the last section suggests a more general point about implementation research. It may take a substantial amount of time for an innovation to work its way through the implementation process and for an observer to be sure that it had an effect, was modified, or made no difference. The importance of an appropriate time frame, although often discussed in the literature, is seldom taken with the seriousness it deserves. Choosing an inappropriate "before" period or not waiting long enough to be sure some normalization has occurred can cause a variety of misunderstandings, ranging from the incorrect assertion that an innovation was not implemented or made no difference to the assertion that it made a substantial difference when it did not.

The optimal research strategy for assessing the effect of an innovation is the true experiment in which subjects (defendants, for example) are assigned randomly to treatment and control groups. With this design the effects of the innovation can be assessed while randomization controls for the effects of confounding factors. Such experimentation,

however, is rendered difficult by legal and ethical constraints.[8] The "next best" research strategy is to treat innovation as a "quasi-experiment" by careful analysis of pre- and post-innovation data. The application of a variety of designs to legal innovations is discussed usefully by Lempert (1966). He notes that the general Campbell-Stanley strategy suggests a number of ways of gathering data on "natural" experiments, ranging from simple two-point before-and-after comparisons to more sophisticated time-series analysis combining multiple observations with attention to relevant control groups.

Most of the research on implementation has relied on the simplest design, a before-and-after look at the results of a particular innovation in a single jurisdiction. Such simple two-point comparisons have a variety of shortcomings. One difficulty is that they mask secular trends so that what appears to be a change following implementation of a reform may simply be a continuation of some trend which was underway well before the innovation took place. They may also confuse short-term variation with longer term change. Here, the apparent impact may turn out to be variation which is either "random" in character or the temporary result of a shock to the system. Ross (1981) has found such patterns in a series of studies of traffic law enforcement. He could not have done so had he used simple before-and-after designs.

Most of the more sophisticated attempts to measure impact of which we are aware improve upon the before-and-after approach only insofar as they take multiple observations in the two time periods and attempt to control statistically for other factors that might affect the dependent variable, as for example, when a time-series evaluation of a crackdown on speeders controls for gasoline tax revenues as an index of miles driven, or prior record of defendants is controlled in a time-series developed to assess the impact of a new sentence law. This design, despite its advance on the simple before-and-after approach, is still a relatively simple one. The danger is that the observed change may be wrongly attributed to the innovation itself, rather than to some other factor which was also present. For example, in our study of the California DSL we found, as some observers predicted, that prison rates rose after implementation of the law. However, there were

[8] "True" experiments based on random assignment are not always impossible. Two current studies of innovation in criminal justice policy are based on such random assignment procedures, a study by Goldkamp and Gottfredson of bail guidelines in Philadelphia and an evaluation by the URSA Institute of a system for early assignment of defense counsel in three cities.

indications that a "law-and-order" trend was underway prior to passage of the DSL. For example, there had been an increase in public concern over crime, an increase in the number of contested judicial elections, and the prison commitment rate had been rising. This heightened concern about crime not only contributed to the passage of the law but arguably would have caused increased prison rates of the magnitude seen *after* the DSL was implemented had that law never been passed. Lempert notes this general type of problem, discussing it as a "history-selection interaction":

> [Designs without control groups are] quite weak in the area of history-selection interaction. Using [them] does not help the investigator decide whether a particular law caused a perceived behavioral change or whether the change would have occurred anyway with the law being merely an expression of an intent that such a change should occur (1966: 124).

Examination of other states that did not pass determinate sentence laws reveals that many also experienced increased prison rates, suggesting the importance of examining control groups when possible.

As noted above, the importance of the proper time frame in evaluating the effects of an innovation should be well known, for it is widely discussed and the point is not hard to understand (Cook and Campbell, 1979; Lempert, 1966). Yet a good deal of the implementation literature, particularly that involving criminal justice system innovations, does not heed the available advice. Two-point comparisons with pre- and post-innovation data collection periods of short duration (e.g., six months or a year before and after) are common.[9]

A major reason why such simple research designs still predominate is that data collection is expensive, and gathering less data requires less money. Designs involving short before and after periods are cheaper to implement and will remain so. With limited resources, it may be wiser, however, to sample

[9] Heumann and Loftin offer an argument in support of such a strategy. They suggest that the period immediately after passage is the best time to examine the implementation of an innovation because it is highly publicized, fresh in people's minds, and one has to worry less about the effects of other explanatory factors that may develop during a longer post-innovation period. The argument is plausible, but the power of inertial forces in criminal courts is such that the expectation of substantial change in a short period of time is implausible. Moreover, observed change over a short period may be deceptive, for it may be a short-term response that decays over time. Although a longer post-innovation period requires more care in controlling for rival factors that may emerge in the post period, it seems a better research strategy if an innovation can only be studied once. Ideally, of course, the evaluation researcher would periodically assess the effects of the innovation under study.

several short time periods during longer before and after periods than to invest everything in a continuous sample just before and just after the innovation. The latter, more common strategy is likely to generate larger numbers of cases and permit better control for some rival hypotheses (e.g., in the case of criminal courts, changes in defendant characteristics like prior record or crime type). On the other hand, it is likely to miss short-term cycles or secular trends in which an innovation may be embedded, and it may not allow enough time for an innovation to normalize and become integrated into the behavior of those participants whose behavior is of interest. A series of data snapshots over a longer period of time runs into problems that stem from small numbers of cases (and hence more difficulty in controlling for changes in the population studied) as well as the effects of history since the number of relevant events overlapping with the period over which the innovation is studied will increase. Clearly, one cannot in the abstract prescribe an optimal strategy, but the common reliance on two-point comparisons requires careful scrutiny and justification.

The tendency to rely upon relatively short periods for implementation research is not only a product of economic constraints. It is also related to the interests of funding sources and fads in public and academic interest. The money for implementation-related research usually comes from sources interested in seeing whether some program or other works. The time horizon of such funding sources is often short. They are interested in knowing whether the latest innovation is going to achieve some set of purposes, and they want to know soon so that they can decide whether to spend more money on it or to try something else. This perspective leads to a desire for quick answers, which usually means that funding sources want the evaluation to begin almost as soon as the innovation has begun. Again, the California DSL provides a good example. It received considerable attention from the practitioners, researchers, and officials and politicians in other states considering similar reforms. To satisfy these interests, a good deal of research, almost all funded by the National Institute of Justice, was undertaken shortly after passage of the legislation.[10] Ironically, the great early interest in the DSL and the willingness to commit substantial funds to study it may have made more revealing medium-term studies less likely, for

[10] By 1981 at least seven studies of the implementation of the California DSL had been completed (Cohen and Tonry, 1983).

there is a natural tendency not to fund further research in an area until the results of the first wave of funding are in. As a result of this funding pattern, much of what we "know" about California comes from work that focuses upon what happened in 1978, the first full year of the act. Because information on the California DSL was wanted quickly, studies of it, including our own, lack data sufficient to warrant the assertion that what happened after implementation is attributable to the law itself. Those studies which observed changes (overall or for crime types) in their post-innovation periods may simply be reporting temporary adjustments which decayed back to prior practice; those which report little change may not have followed the law long enough for it to work its way into court disposition processes. Moreover, many of the studies lack multiple observations in the pre-innovation period and thus miss the pre-existing trends towards increased prison commitment rates. This, of course, is not entailed by the fact that an evaluation is commissioned shortly after an innovation, but it may be a function of the pressures to report on the innovation as soon as possible. Collecting and analyzing extensive pretest data takes time, and desirable complexity may be sacrificed when early answers are wanted.

Not only are funding sources likely to be interested in what is new and different, but so are investigators. Matters that have been little studied are likely to attract considerable interest. Thus, the tendency toward relatively rapid investigation of the implementation of innovations and its concomitant—short post-innovation periods of data collection and analysis—cannot be attributed entirely to short-sighted funding sources. To some extent the incentives of fund sources and investigators converge. To assert this is neither to condemn it nor to offer a solution. Rather it is to emphasize that effective work on implementation requires substantially longer time perspectives and more extensive data collection than the exigencies of the research situation tend to prescribe. For those who are doing and evaluating implementation research, these issues require careful attention.

V. CONCLUSION

The implications of this discussion are straightforward. Analysis of the implementation process requires careful attention to the notion of "goals." Goals are vague, slippery, sometimes ephemeral, and often quite different from what pre-adoption rhetoric or manifest policy content may suggest. By

the same token, attempts to assess the impact of policy innovations require the passage of sufficient time to allow one to make sense of the complex interaction of motives and behavior that implementation encompasses. Finally, in interpreting whatever alterations in behavior have occurred, we should take our own theories seriously. "Small" initial changes can reflect a substantial motivating force if there is reason to believe that tendencies to resist change are powerful, and small changes may presage marked change over time. These adjurations have been offered by others before. Our purpose here has been to suggest by argument and example that they are, indeed, crucial to understanding and analyzing the implementation process, as well as to appreciating what it means to evaluate the effectiveness of a policy innovation.

REFERENCES

ARNOLD, Thurman W. (1935) *The Symbols of Government*. New Haven: Yale University Press.

BARDACH, Eugene (1977) *The Implementation Game: What Happens After a Bill Becomes a Law*. Cambridge, MA: MIT Press.

BEHA, James A. (1977) "And Nobody Can Get You Out," 57 *Boston University Law Review* 96 (Part I), 289 (Part II).

CASPER, Jonathan D., David BRERETON and David NEAL (1982) *The Implementation of the California Determinate Sentencing Law*. U.S. Department of Justice, Washington, D.C.: U.S. Government Printing Office.

CHURCH, Thomas W., Jr. (1976) "Plea-Bargains, Concessions, and the Courts: Analysis of a Quasi-Experiment," 10 *Law & Society Review* 377.

COHEN, Jacqueline and Michael TONRY (1983) "Sentencing Reforms and their Impacts," in A. Blumstein, J. Cohen, S.E. Martin and M.H. Tonry (eds.), *Research on Sentencing*, Vol. II. Washington, D.C.: National Academy Press.

COOK, Thomas D. and Donald T. CAMPBELL (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

EDELMAN, Murray (1964) *The Symbolic Uses of Politics*. Urbana, IL: University of Illinois Press.

HEUMANN, Milton (1978) *Plea-Bargaining: The Experience of Prosecutors, Judges and Defense Attorneys*. Chicago: University of Chicago Press.

HEUMANN, Milton and Colin LOFTIN (1979) "Mandatory Sentencing and the Abolition of Plea-Bargaining: The Michigan Felony Firearms Statute," 13 *Law & Society Review* 393.

HINDELANG, Michael, Michael GOTTFREDSON and Timothy FLANAGAN (eds.) (1980) *Sourcebook of Criminal Justice Statistics*. Washington, D.C.: U.S. Government Printing Office.

JOINT COMMITTEE ON NEW YORK DRUG LAW EVALUATION (1977) *The Nation's Toughest Drug Law: Evaluating the New York Experience*. New York: Association of the Bar of the City of New York.

LEMPERT, Richard (1966) "Strategies of Research Design in the Legal Impact Study," 1 *Law & Society Review* 111.

LOFTIN, Colin, Milton HEUMANN and David McDOWALL (1983) "Mandatory Sentencing and Firearms Violence: Evaluating an Alternative to Gun Control," 17 *Law & Society Review* 287.

MARTIN, Susan (1983) "The Politics of Sentencing Reform," in A. Blumstein, J. Cohen, S.E. Martin and M.H. Tonry (eds.), *Research on Sentencing*, Vol. II. Washington, D.C.: National Academy Press.

MARTINSON, Robert (1974) "What Works? — Questions and Answers about Prison Reform," 35 *Public Interest* 22.
MATHER, Lynn M. (1979) *Plea-Bargaining or Trial? The Process of Criminal Case Disposition*. Lexington, MA: D.C. Heath.
MAZMANIAN, Daniel A. and Paul A. SABATIER (eds.) (1981) *Effective Policy Implementation*. Lexington, MA: Lexington Books.
MEDALIE, Richard J., Leonard ZEITZ and Paul ALEXANDER (1968) "Custodial Police Interrogation in Our Nation's Capital: The Attempt to Implement Miranda," 66 *Michigan Law Review* 1347.
MUIR, William K. (1973) *Law and Attitude Change*. Chicago: University of Chicago Press.
MURPHY, Jerome T. (1971) "Title I of ESEA; The Politics of Implementing Federal Educational Reform," 41 *Harvard Educational Review* 35.
NAKAMURA, Robert T. and Frank SMALLWOOD (1980) *The Politics of Policy Implementation*. New York: St. Martin's Press.
OAKS, Dallin H. (1970) "Studying the Exclusionary Rule in Search and Seizure," 37 *University of Chicago Law Review* 665.
PELTASON, Jack W. (1961) *Fifty-Eight Lonely Men: Southern Federal Judges and School Desegregation*. New York: Harcourt, Brace.
PIERCE, Glenn L. and William J. BOWERS (1981) "The Bartley-Fox Law's Short-Term Impact on Crime in Boston," 455 *Annals* 120.
PRESSMAN, Jeffrey L. and Aaron WILDAVSKY (1973) *Implementation*. Berkeley: University of California Press.
ROSS, H. Laurence (1981) *Deterrence of the Drinking Driver: An International Survey*. Washington, D.C.: United States Department of Transportation, National Highway Traffic Safety Administration, Technical Report DOT HS-805 820.
RUBINSTEIN, Michael L. and Teresa J. WHITE (1979) "Alaska's Ban on Plea Bargaining," 13 *Law & Society Review* 367.
VAN METER, Donald S. and Carl E. VAN HORN (1975) "The Policy Implementation Process: A Conceptual Framework," 6 *Administration and Society* 445.
WALD, Michael (1967) "Interrogations in New Haven: The Impact of Miranda," 76 *Yale Law Journal* 1519.