1973

Toward a Science of Impartial Judicial Behavior

Harry Kalven Jr.

TOWARD A SCIENCE OF IMPARTIAL JUDICIAL BEHAVIOR

*Harry Kalven, Jr.**

One of the strongest aspects of these studies by Professors Cook and Goldman is that they represent continuities in research. They thus give promise that social research in law may be ready to emulate its neighbor, the natural sciences, in the patient, incremental adding to knowledge by taking its point of departure from prior studies. In this instance both studies, as their notes amply document, are placed in a general area—the empirical study by quantitative techniques of judicial decision-making—in which considerable work has already been mounted. But the continuities are more specific and more emphatic.

Professor Goldman is extending techniques for analyzing collegial judicial behavior to groups of judges who do not, like the United States Supreme Court, act en banc, but rather in panels of three. Since there is considerable business in common between the Supreme Court and the courts of appeals, he appears to have enlarged significantly the arena in which judicial behavior can be systematically scrutinized to detect voting blocs and attitudinal syndromes.[1] Professor Cook, as her repeated use of the prior Markham study shows, is building directly not merely on studies of sentencing—something of a vogue at the moment—but on studies on her precise topic, sentencing in draft cases.

In another positive aspect, these studies show once again that the inability to conduct controlled experiments need not put an end to the law's aspirations toward empirical inquiry.[2] And they underscore how enormously useful the sensible collection of statistics by the Administrative Office of the United States Courts may prove to be; it is providing a stockpile of data for endless

* Harry A. Bigelow Professor of Law, University of Chicago. A.B., Chicago, 1935; J.D., 1938.

1 For an amateur struggle with counting votes and detecting blocs at the Supreme Court level, see Kalven, *Foreword: Even When a Nation Is at War*, 85 HARV. L. REV. 3, 6–10 (1971).

2 *See* Kalven, *The Quest for the Middle Range: Empirical Inquiry and Legal Policy* in LAW IN A CHANGING AMERICA 56, 69–71 (G. Hazard ed. 1968); Zeisel, *Reflections on Experimental Techniques in the Law*, 2 J. LEGAL STUDIES 107 (1973).

secondary analyses by imaginative social scientists.[8] What is most promising is that both studies have been able by statistical techniques to exploit routine descriptive judicial statistics as a kind of circumstantial evidence of the factors underlying judicial behavior.

I found the Cook study, perhaps because its topic is more readily accessible, particularly intriguing as to the possibilities in creative cross-tabulation. She does not limit herself to cross-tabu lating the severity of draft sentencing against the familiar demographic variables, but introduces three variables that carry a promise of theory: judicial environment, judicial reference groups, and judicial structure. She is thus able to probe subtle relationships one would have thought only in depth interviewing could detect. Several points invite special coment.

One cannot but admire the ambitious effort to trace American Legion pressure on draft sentencing. The technique was to use census figures on the percentage of veterans in each state who belonged to the Legion. Professor Cook thus derived a Legion concentration index to run against her judicial severity index to test whether severity corresponded to Legion density, and hence, by hypothesis, to Legion pressure. One can easily think of reasons, as I am sure Professor Cook did, why these measures might be too crude to detect such pressures if they did exist. But such common sense methodological criticism misses the point. What is arresting and welcome is that a study utilizing only judicial statistics could nevertheless pursue so lively a quarry.

On at least three occasions it seems to me Professor Cook turns up modest correlations which trigger reflection on the well springs of sentiment. Consider the point that severity correlated inversely with judicial workload: the more draft cases a judge handled, the more lenient his sentencing of draft violators. Repeated exposure to the dilemma of the young draft violator apparently made the judge more understanding and more sympathetic. A classic

[8] It is an important point in designing social research whether it can proceed by re-analyzing existing records, that is, by secondary analysis, or whether in addition to the burden of analysis, it must also carry the burden of generating its own data. In the Jury Studies, the study of court congestion, H. ZEISEL, H. KALVEN & B. BUCHHOLZ, DELAY IN THE COURT (1959), was an instance of secondary analysis; the study of judge-jury differences in deciding the same case, H. KALVEN & H. ZEISEL, THE AMERICAN JURY (1966), and the study of mock juries, R. SIMON, THE JURY AND THE DEFENSE OF INSANITY (1967), were instances where it was necessary first to create the data.

defense of the reliance on the jury in criminal cases has been founded on the need to avoid a tendency toward stereotyping of criminals by weary, disheartened, experienced judges; the jury was always new and freed thereby to take a fresh look. The distinctive puzzles generated by draft violations during the Vietnam War produced in the judge the fresh reaction.

Or consider the suggestions that judges with sons of draft age tended to be more severe, while judges with military service backgrounds of their own tended to be more lenient. The correlations are not powerful, but the results, while at first surprising, are on reflection plausible and rewarding. They honor the complexity of the policy issue posed for the judge in dealing with a draft violator during the Vietnam War. A tiny step perhaps, but a step on the long road toward more general theories of judicial decision-making.

These are, then, good examples of a genre of behavioral research, and we are glad to have them added. One nevertheless cannot suppress some ambivalence about the genre itself.[4] I suspect that ambivalence arises not because we are persuaded such efforts can tell us nothing useful about judicial behavior, but rather because we are so anxious to have them tell us more than they do. Perhaps I misunderstand, but they seem to be content with documenting the fact that judges are human and imperfect, and that all sorts of factors impinge on their decision-making. They thus risk the appearance of telling us, after empirical study, that there is no Santa Claus. Or to vary the unkind joke, they bring to mind a comment some 20 years ago by the atomic scientist, Leo Szillard, about American policies re atomic secrets. There is, said Szillard, no secret and we intend to keep it. Apropos to the students of judicial behavior might be that there is no secret and they intend to tell it.

I am left puzzled about what they conceive to be the counter expectation which is impeached by their study. In *The American Jury*, Hans Zeisel and I were able to array the decisions of juries

[4] There is ambivalence too about the special burdens of exposition in such studies. It is a formidable task to write English while seriously utilizing statistical data. The need to set forth the apparatus of inquiry tends to make reports top-heavy and too little like essays. Further, the need to measure neatly often requires the introduction of intermediate concepts which are themselves artifacts and at least one remove from experience. Thus the severity index in the Cook study conceals whether the sentence included prison.

against the decisions of judges and thus provide a baseline against which to appraise the jury's performance. How much and when did the jury depart from the judge? What is the analogous baseline for these two studies? In Professor Cook's study, for example, some 61 percent of all cases and 71 percent of all judges had sentences for draft violations which fell within the 10–29 brackets on a scale that ran from 1 to 99. This could be read as a notable degree of consensus and regularity among judges in a highly unstructured situation. In Professor Goldman's study the courts of appeal were unanimous over a six-year period in some 94 percent of their cases, and this despite the evidence of voting blocs.

The slogan "a government of laws not men" has always been understood to mean, in operation, a process of decision-making *by men*. The fundamental premise in the idea of impartial judges and rules of law is that certain kinds of decision-making, for example, by judges, can by institutional arrangements and role discipline be made to show less variance and less correlation to personal factors than other kinds of decision-making, for example voting and consumer preferences.[5] Beyond this there is the question of whether different arrangements will produce different degrees of impersonality in decision-making. In Professor Cook's instance, would, for example, exposure to sentencing seminars or official guidelines for draft cases alter the distribution of results on her severity index?

Watergate, among its many effects, has perhaps lent new significance to the question of impartiality in decision and made clear that it is not simply a myth sponsored by lawyers. The vivid example concerns not so much the judges involved, but the role of the independent prosecutor. What is it we expect of him? Certainly in this instance everyone has in mind some achievement of impartiality and even-handedness, some independence not only from obvious political pressures but from personal bias as well. Assume that in some way recurrent decisions were presented to, say, a dozen different independent prosecutors, and we later made a study of their responses. If they each had fulfilled our current aspirations for an independent prosecutor, what would the statistical pattern of results be like? Since prosecutorial decisions will necessarily involve some close judgmental points, can we ex-

[5] *See* Blum & Kalven, *The Art of Opinion Research: A Lawyer's Appraisal of an Emerging Science*, 24 U. CHI. L. REV. 1 (1956).

pect them to agree all the time? If they disagree, what would be the statistical evidence that they had remained independent?

The burden of making sense out of the idea of the independent prosecutor or the impartial judge does not rest alone on the empirical student of judicial behavior from political science. It rests equally on the student of jurisprudence from the legal profession. We found that empirical study of the jury increased sensitivity to theorizing about the jury as an institution. Empirical study of judicial behavior must stimulate and sharpen theory about the aspiration to have a human being who as a judge is tolerably impartial and tolerably independent. What seems to be needed is to have the student of jurisprudence join the student of judicial behavior in first translating the normative expectations into statistical expectations, and then in devising ways to "gradually, very gradually"[6] test their implementation.

Writing a few years ago on the relationships between law and social science, I concluded my essay with a slogan. It will be clear now that I am slow to have a new idea. The slogan was: "let us 'empiricize' jurisprudence and intellectualize fact finding."[7]

[6] Hutchins, *The Autobiography of an Ex-Law Student*, 1 U. CHI. L. REV. 511, 518 (1934).

[7] Kalven, *supra* note 2, at 72.