Law and Human Behavior, Vol. 3, Nos. 1/2, 1979

External Validity of Research in Legal Psychology

Vladimir J. Konečni* and Ebbe B. Ebbesen*

INTRODUCTION

The growth of legal psychology in the past ten or so years has been reflected in a relatively large number of edited volumes and symposia at various conventions, in the far greater visibility of psychologists in various interdisciplinary forums (such as the meetings of the American Psychology–Law Society), in the increasing frequency with which psychologists appear as expert witnesses (whose testimony addresses an ever broader range of issues), and, finally, in the founding of this journal. The main purpose of the discipline, presumably, is to explore various aspects of the "interface" between psychology and the law, and, more specifically, to enhance the understanding of the operation of the legal system by using psychological research methods and by testing the validity of psychological assumptions contained in legal statutes or else made by legal practitioners on an *ad hoc* basis.

It appears self-evident that legal psychology has a strong applied orientation. Many psychologists do research in it primarily because the results of their theoretical and empirical efforts can be applied in an obviously important social domain. Many lawyers take an interest in it because they have grudgingly begun to believe that psychologists can make a practical contribution to the judicial process. Furthermore, the research emphasis has been on the various practical aspects of criminal procedure, rather than, for example, on lofty speculations about the role of psychological principles in legal doctrines; namely, a close examination of the literature in legal psychology shows that a very large proportion of all research studies falls into the

*Department of Psychology, University of California, San Diego.

0147-7307/79/0600-0039$03.00/0

categories of jury decision making (e.g., Efran, 1974; Kalven & Zeisel, 1966; Kerr, Atkin, Stasser, Meek, Holt, & Davis, 1976; Landy & Aronson, 1969; Mitchell & Byrne, 1973; Nemeth & Sosis, 1973; Sigall & Ostrove, 1975; Vidmar, 1972), eyewitness identification (e.g., Buckhout, 1974; Buckhout, Alper, Chern, Silverberg, & Slomovits, 1974; Doob & Kirshenbaum, 1973; Egan, Pittner, & Goldstein, 1977; Levine & Tapp, 1973; Loftus, 1975; Loftus, Altman, & Geballe, 1975), and procedural justice (e.g., Doob, 1976; Farmer, Williams, Lee, Cundick, Howell, & Rooker, 1976; Lawson, 1970; Thibaut & Walker, 1975; Walker, LaTour, Lind, & Thibaut, 1974). In short, the purpose, origins, research orientation, and fast growth of legal psychology all seem to be largely attributable to an applied, practical emphasis.

Research in applied disciplines must be concerned with issues of external validity and generalizability to an unusually high degree. The criteria for what is a good experiment, when a certain methodology appears sound, and which results are to be trusted, must necessarily be different and more stringent when sweeping, costly, and far-reaching changes in public policy and—quite literally in the legal domain—people's futures and lives, depend on inferences from research results. Starting with the seminal work of Brunswik (1956), many aspects of the problem of external validity and related matters have been discussed at length by psychologists (Campbell & Stanley, 1963; Cronbach, 1976; Rosenthal & Rosnow, 1969; Slovic, Fischhoff, & Lichtenstein, 1977; Webb, Campbell, Schwartz, & Sechrest, 1966). Many of these authors stressed, among other things, the importance of the representativeness of research samples, treatment combinations, and outcome measures, and, in fact, made numerous, very specific recommendations regarding ways to increase the external validity of causal hypotheses.

What is surprising is the extent to which most studies in legal psychology have routinely ignored external-validity problems, despite the obvious applied nature of legal psychology and the airing that the notion of external validity has received. For example, of the various lines of research mentioned above, simulated-jury experiments have been by far the most numerous (cf. Davis, Bray, & Holt, 1977; Tapp, 1976), despite the fact that the very nature of this research problem necessitated that most of the studies be conducted in the laboratory (cf. Bermant, McGuire, McKinley, & Salo, 1974; Davis et al., 1977; Konečni, Mulcahy, & Ebbesen, in press; Tapp, 1976)—which is precisely a research setting that is most suspect from the external-validity point of view. Even those research problems in legal psychology that are not of necessity confined to laboratory settings nevertheless seem to be investigated there most of the time.

Criticisms of the external validity of laboratory simulations in legal psychology have not been lacking. For example, although it is agreed that college students are not representative of the participants in the real-world legal system (e.g., Miller, Fontes, Boster, & Sunnafrank, Note 1), student "juries" continue to be given two or three bits of information about a "case" and are asked to make guilt determinations on the basis of this information, even though it is clear that subject variables and quite trivial features, such as the differential number of words necessary to describe various factors (or levels of factors), may well decide which of the factors (or their levels) will produce greater effects. Often the decision required of students and the response mode are far removed from that typical of real-world juries (e.g., Bray, Note 2; Ebbesen & Konečni, Note 3; Konečni et al., in press). Besides, the fact that the information is presented in a decomposed form eliminates the task of extracting information

embedded in context (often the "context" consists of events that last for many days or weeks), which is precisely one of the more important tasks that the real-world legal decision makers have to face [Ebbesen & Konečni, in press (a); Gerbasi, Zuckerman, & Reis, 1977]. Student jurors continue to find themselves in many cells of repeated-measures experimental designs, forced to make numerous guilty/not guilty verdicts with no important consequences [e.g., Ebbesen & Konečni, in press (a); Wilson & Donnerstein, 1977], in short periods of time, taking into account curious "facts of the case," such as, for example, the "attractiveness of the defendant" (manipulated by photos from high school yearbooks), which may have effects only because so little additional information is presented. Furthermore, these decisions are often made in the absence of key procedural features common to juries (e.g., Izzett & Leginski, 1974; Myers & Kaplan, 1976; Vidmar, Note 4). (It should be noted that even robust, repeatedly obtained laboratory effects are not necessarily informative about the analogous real-world relationships.) In short, a close examination of the run-of-the-mill studies in legal psychology shows that most offer only lip service, if any, to external validity and have generally been impervious to mounting criticism concerning poor generalizability.

One is led to the conclusion that an implicit assumption behind these studies is that a correct understanding of the functioning of the legal system and a reliable description and predictions of the decision-making behavior of the participants in the system can be obtained regardless of the characteristics of subjects, experimental materials, treatment conditions, dependent measures, and so on.

Such an optimistic view does not seem warranted. A close examination of the decision-making literature reveals that subjects' decision strategies do not typically generalize across tasks, that subjects are often responsive to task features that are incidental, rather than central, to prior theoretical conceptions, that researchers typically do not know ahead of time when the subjects' "oversensitivities" to such incidental aspects of the tasks will occur, and so on; in short, subjects appear to create decision rules and ways of dealing with, and combining, information presented to them for each particular task they are working on [Ebbesen & Konečni, in press (a); Olson, 1976]. If subjects indeed create decision strategies in response to various, and often trivial, elements of a laboratory task, how informative about the operation of the real-world legal system can the ubiquitous laboratory simulations be?

A perhaps more general way of phrasing our concern is that it seems that the answer one gets from subjects and the usefulness of that answer depend at least as much on how a question is asked as on its content. By a "question" we mean not just the specifics of the particular experimental task given to the subjects, but also the general research method that is employed (to the extent that tasks are method-specific, as many are, the above two points become one). Very different conclusions may well be reached about a particular research issue in legal psychology depending on whether one or another experimental task is used, and on whether one uses interviews, questionnaires, laboratory simulations, naturalistic experiments, or archival data-collection methodologies—a diversity of conclusions that would almost certainly be further increased by using different types of subjects.

In order to buttress our critique of laboratory simulations in legal psychology and substantiate our task-and-method-specificity argument, we will describe in the next section of the article some of our work on two legal decisions, namely, bail setting and sentencing. Each type of decision was investigated using different research

methods (ranging from unstructured interviews to experimental simulations to archival data collection), using very different types of subjects (ranging from college students and probation officers to defense attorneys and judges), and occasionally studying the same type of subjects both with and without their knowledge that they were being studied.

Generally speaking, in the case of both types of decisions, each type-of-subjects/type-of-method combination produced different results. In other words, very different conclusions about how the decisions are made and how the legal system operates would have been reached on the basis of studies involving different subject/method combinations. This simultaneously means that if the purpose of the research had been to feed the information back to the system and to make practical recommendations—as should be the case for research in an applied discipline—quite different suggestions would have been made on the basis of one as opposed to another type of study.

Given that different methods led to different conclusions, which one should be trusted? In the final section of the article, we argue that unlike many other research problems and settings, in the case when the legal system is studied, there are certain logical and practical criteria that would lead one to trust the conclusions reached by one method over those reached by another on *a priori* grounds, with the important proviso that the researcher is interested in how the system actually operates, rather than in the phenomenology of the decision makers.

A COMPARISON OF SOME SIMULATED AND NATURALISTIC LEGAL DECISIONS

Setting of Bail

The work on the setting of bail [Ebbesen & Konečni, 1975; in press (b)] initiated our long-term research project on legal decision making, which has so far included studies of sentencing, police charging decisions, district attorneys' complaint-filing decisions and plea-bargaining behavior, judicial and psychiatric decisions concerning the disposition of persons suspected of being mentally disordered sex offenders, judicial decisions in personal-injury and child-support cases, etc. Because it was an early study, the bail-setting research investigated the behavior of only one type of subject, using only two data-collection methods.

Simulated Bail Setting

The subjects were San Diego County judges who had first-hand experience with bail-setting. In the first study, our assistants, who introduced themselves as college students interested in the setting of bail, saw the judges in their chambers. The judges were presented with sheets of paper, each of which presented a different factorial combination of cues that they were to use in reaching their simulated bail decisions. In order to make the experimental design manageable with regard to size, all cases involved the crime of robbery. The manipulated factors were: (a) *prior record* of the defendant (the fictitious accused was said either to have no prior record or to have a prior felony record and to be on probation at the time of the bail hearing; these two levels of the prior record factors simulated 93% of the actual robbery cases we examined); (b) *"local ties"* (the accused was described as having either strong local ties,

that is, that he had lived in San Diego for 4–6 years, that he was employed, and that his family also was living in San Diego, or that he had weak ties, in which case it was noted that the accused had been living in San Diego for only 1–2 months, that he was unemployed, and that his family lived in Northern California); (c) *defense attorney bail recommendation* (low: $0; moderate: $500–$600; high: $1000–$1200); and (d) *district attorney bail recommendation* (low: $1500–$1700; moderate: $2000–$2500; high: $5000–$7500).

These four factors were embedded in a "case record summary." The judges were told that the cases were selected from the trial archives in San Diego and that they were all concerned with the crime of robbery (thus, the severity of crime was kept constant in this study). They were further told that other factors, such as the age and sex of the accused would not vary in these cases. In the "case record summaries" that were presented to the judges, the age of the accused was always set between 21 and 25 years, which is quite realistic given that 60% of the felony cases in San Diego involve people in this age range. The accused was further described as an unmarried male Caucasian, who had been charged with robbery and had pleaded "not guilty." The details of the robbery were varied slightly across the cases, but the value of the stolen property was always set between $850 and $950. The bail recommendations by the district and defense attorneys closely corresponded, with regard to range and mean, to bail recommendations made by these people in the actual court cases we examined. Moreover, the high bail recommendation by the defense attorney was nevertheless lower than the low recommendation by the district attorney—another realistic feature.

In short, in the choice of subjects (actual judges), the choice of independent variables (with the exception of the severity-of-crime factor, which was excluded for the reason mentioned above, few, if any, other types of information are presented in the *actual* bail hearings), and in the range of the levels of the independent variables the study was an honest attempt to produce as good a simulation of the "real thing" as possible. This was also true for the dependent measure: The judges gave bail recommendations in dollars.

The independent variables were presented to 18 judges in such a way that the resultant judgments yielded a $3 \times 3 \times 2 \times 2$ factorial design with four observations per cell. Individual judges were randomly assigned to a group of eight case records selected from the entire factorial matrix in such a way that no judge was exposed to the same level of any independent variables more than twice.

An analysis of variance of the data yielded three significant main effects, all of which were in the expected direction (judges set higher amounts of bail in cases with longer prior records, weaker local ties, and higher district attorney's recommendations); the main effect for the defense attorney's recommendation, as well as all of the interaction effects, were not significant. The proportion of variance that each significant factor accounted for is of considerable applied interest because the levels of the factors simulated the most frequently occurring levels in actual bail hearings. The proportion of variance accounted for by a factor therefore should provide information about the relative importance of these variables (given the studied levels) in the actual decision-making process. From this point of view, the local ties factor [$F(1,108) = 22.38$] accounted for the most, by far, between-cell variation; the prior record and D.A. recommendation factors accounted for far less variance and were

about equal to each other [prior record $F(1,108) = 5.91$; D.A. recommendation $F(2,108) = 7.05$].

If this were the only study of bail setting we had done, the conclusions would have been that the judges tend to adopt a reasonable decision-making scheme in the setting of bail, given the "innocent until proven guilty" doctrine. The judges' decisions required people who had little reason to remain in the area while awaiting trial (weak ties) to post more bail than people who had many reasons for remaining in the area (strong ties). According to the Vera Foundation (Goldfarb, 1965), this strategy of bail setting has met with great success in New York City. Moreover, an emphasis on local ties (and thus the probability of appearance at trial) is a factor that is both formally and informally stressed by judges and included in many state guidelines across the nation. Thus, the inevitable conclusions would have been that: (a) the San Diego judges have adopted a reasonable decision-making strategy in setting bail; and (b) they operate according to the widely accepted Vera Foundation rules.

Courtroom Bail Setting

Although the use of real judges as subjects in the simulation increased the external validity of the study, it did not eliminate all the objections that might be raised to the conclusions drawn from it. In particular, the judges were, after all, aware that they were being tested and they may well have attempted to present themselves as reasonable rule-following and unbiased decision makers by attaching more importance to the local ties variable than they would normally. Furthermore, unknowingly echoing Brunswik, some of the judges reported that certain combinations of the four factors were difficult to respond to because these factorial combinations would seldom, if ever, occur in real bail hearings. In other words, certain treatment combinations in the full factorial design may have produced means that were unrepresentative of the real-world bail hearings, thereby reducing the external validity of the study. In addition, the decision-making strategy in *other* treatment combinations may have been affected by the "unnatural" cells.

For these reasons, we decided to obtain information in actual bail hearings. Data were obtained by having our assistants unobtrusively record the actual courtroom bail-setting behavior of five of the 18 judges who served as subjects in the simulation (only these five served on the bench in bail hearings while the study was conducted). Our trained assistants sat in the courtroom with specially prepared data sheets. For each case, they recorded the following information: sex of the accused, his/her approximate age, the type of crime presumed to have been committed, the plea of the accused (guilty or not guilty), whether or not a defense attorney was present, the amount of bail that the defense attorney recommended, the amount of bail that the district attorney recommended, the extent of the accused's prior criminal record, whether the accused was employed, whether he/she lived in the San Diego area and for how long, whether the accused had relatives living in the San Diego area, and the amount of bail finally set by the presiding judge. Since 96% of the accused individuals were males between the ages of 18 and 30, age and sex were ignored in the final analyses. Because the guilty plea was entered in only one case, this variable was also ignored.

Statistical analyses were conducted using five independent variables to predict the judges' bail-setting decisions: *Severity of the crime* [14 different crime categories were initially identified, ranging in severity from absent-without-leave to homicide; these were subsequently collapsed into seven categories, ranging from victimless crimes to homicide, and matching the ordering established in the previous research by

Coombs (1967), and Leon, Oden, and Anderson (1973)]; *severity of the prior record* of the accused (there were four levels of prior criminal record: None, a minor record consisting only of traffic violations, a moderate record consisting of no more than one felony conviction, or at least one violent-crime felony conviction); *strength of the local ties* of the accused (there were three levels, including weak ties for those cases where the accused had not lived in the San Diego area for more than a month, moderate ties, where the accused had lived in the area for over a year, but was unemployed at the time of arrest, and strong ties, where the accused lived in the area for over a year, was currently employed, and also had relatives living in the area); *defense attorney's recommendation* (this was the actual monetary recommendation made by the defense attorney in the hearing; the range of values was from $0 to $25,000); and finally, *district attorney's recommendation* (this was the actual monetary recommendation made by the district attorney in the hearing; the range was from $0 to $100,000). In all the cases we studied, the judge announced the bail that was to be set so that it could be recorded by the court stenographer; this oral statement was used as the actual bail set.

A multiple regression analysis was used to examine the data obtained in over a hundred bail-setting cases. The cases included four homicides for which special bail-setting procedures seemed to be used. (Even though most murders are committed by people who live in the area in which the murder was committed and who have no prior criminal record, very large bail bonds are usually recommended by the district attorney, possibly for public relations reasons). Therefore, the discussion that follows is based on the amounts of bail set in nonhomicide cases only.

A model with five predictors (the five factors described above) accounted for 94% of the variance, whereas a model with 15 predictors (the five major factors and their two-way interactions) accounted for 96% of the variance. It was found that in terms of the additional, or noncommon, sum of squares that a particular factor contributed to the accuracy of the model based on 15 predictors, severity of the crime [$F(1,90) = 4.26$; $p < .05$], defense attorney's recommendation [$F(1,90) = 23.63$; $p < .001$], and district attorney's recommendation [$F(1,90) = 523.81$; $p < .001$], were all significant, as were two two-way interactions (these will not be discussed further for the sake of simplicity). In contrast, neither the prior record factor, nor, more importantly, the local ties factor, contributed to the accuracy of the model (i.e., they were not significant).

It is immediately obvious that the results of the naturalistic study are entirely different from those obtained in the simulation. Among other things, the local ties factor, which was highly significant in the simulation, now had an F of less than 1.00. In contrast, the judges appeared to rely very heavily on the recommendations of the two attorneys, especially the district attorney.

One possible explanation for these findings is that the district attorneys and/or defense attorneys take the local ties variable into account in an appropriate way when *they* recommend bail, and that the judges know this to be the case. In order to obtain evidence relevant to this issue, we performed regression analyses directly on the district and defense attorneys' bail recommendations. As it turned out, in nonhomicide cases, the only significant predictor of both the district attorney's and the defense attorney's bail recommendations was the severity-of-crime factor. The local ties factor had an F of close to 1.00 in both cases. In fact, in the case of the district attorney, the local-ties factor was involved in a significant two-way interaction which indicated that the district attorneys recommended *more* bail in four out of five crime categories

when the ties were strong or moderate than when they were weak. Thus, as far as the district attorneys were concerned being tied to the area was actually detrimental to the defendant in the more severe crime categories!

In a further attempt to understand the differences between the results of the simulation and of the naturalistic study, we carried out a dummy-variable multiple regression analysis on the data from 63 naturalistic cases selected in such a way that the amount of bail recommended by the district and defense attorneys (three levels of each), the extent of the prior record (two levels), and the type of the local ties of the accused (two levels), exactly matched the levels of these variables in the simulation. This analysis revealed—quite unlike the simulation—that the recommendation of the district attorney was the only significant predictor.

It seems, then, that the results from the simulation and the naturalistic study are very different from each other no matter how one looks at them. To the extent that the reliability of coding in the naturalistic study was high (which it was, see Ebbesen & Konečni, 1975, p. 812), the results of the simulation are totally misleading. The way that the San Diego judges set bail in the courtroom is a far cry from what they appear to believe they do, or, at least, what they would like the researchers to believe they do (as judged by their responses in the simulation). Instead of focusing on local ties and following the Vera Foundation recommendations to which they pay lip service, in the courtroom the judges rely mostly on the district attorney's recommendation and, via this recommendation, on the severity of the crime.

Note that we are not making a value judgment as to whether what the judges actually do is "better" or "worse" than what they say they do. The point is simply that the results are *different* and that the results of the simulation are useless and misleading. (After all, there seems to be little *intrinsic* interest in studying the quasilegal behavior of real judges if this behavior does not successfully mimic their behavior in the real world.) From an applied point of view, whether the results of one or the other of the two studies were to be "fed back" to the participants in the system would have presumably resulted in very different policy changes and recommendations.

Sentencing

Our work on sentencing of adult felons was considerably more ambitious than that on the setting of bail. It involved six different data-collection methods, including *interview* (with San Diego County Superior Court judges as subjects), *questionnaire* (with judges as subjects), *rating-scale responses* (with judges, defense attorneys, and college students as subjects), *experimental simulation* (which used judges, probation officers, and students as subjects, in the latter case in either a between-subjects or a within-subjects experimental design), *observation of sentencing hearings* (coding of the content of the live sentencing hearings), and *archival analysis* coding of court files after the sentence had been passed). Our purpose in using the different methods, and the results obtained in the different studies, will now be presented in turn.

Interview (the "Journalistic" Approach)

One way of studying the factors that affect the sentencing decisions is simply to interview the judges who make those decisions. Although the various problems associated with this data-collection method are well known, it is undoubtedly one of the most popular research tools in the social sciences. The proponents of this impressionistic and intuitive method would claim that its advantages are a considerable

amount of flexibility and an unmatched opportunity to tap the rich phenomenology of the sentencing process, provided that the interview is unstructured enough and conducted well. In a sense, interviewing judges about the factors that they feel influence their sentencing decisions is analogous to the way in which anthropologists and linguists use native "informants" to discover the otherwise inaccessible aspects of a foreign culture or language.

We decided to use the interview as one of the methods of studying sentencing for two principal reasons. One was economy: Interviewing is a relatively inexpensive data-collection method when the subject population is fairly small. If we were to find that the conclusions one would reach on the basis of the interview data were not different from those reached by other methods, we would then be inclined to rely heavily on the interview in our other projects. There was, however, a far more fundamental reason for including the interview in the project. Social scientists are by no means the only seekers of information who rely—via the interview—on people's self-reports as accounts of the causes of their behavior. Far-reaching reforms with a considerable social, political, and economic impact—including reforms of the criminal justice system—are frequently undertaken and passed by legislators in part because they are "supported" by lengthy compilations of alleged "facts," yet these support documents often contain mostly self-report data provided by the participants in the system that is being reformed. For example, the change from an indeterminate to a determinate sentencing system in California in 1978—an event of considerable social and economic importance—was to a large extent caused by the opinions, impressions, and intuitions of the participants in the system (including the judges)—precisely the self-report type of material that one might elicit in an interview. (The same is true of recommendations based on Congressional and Senate Committee hearings.)

Furthermore, it could be argued convincingly that much of what the tax-paying general public believes about the functioning of a (very expensive) social institution, such as the criminal justice system, and about the way that decisions are reached within it, such as the sentencing decisions in felony cases, is shaped to a large extent by one or another type of self-report provided by the participants in the system (such as the judges). This occurs mainly because the media that play the chief informational role rely to such a high extent in *their* information-gathering efforts on various types of interviews and self-reports by the participants in a system.

It was this latter concern that to a large extent determined exactly how our interview study of sentencing was done. Let us first describe how the study was carried out and then give reasons for the various details. The study was conducted by three undergraduates from the University of California at San Diego. These people were enrolled in a course in field research methods in social psychology and were unaware of our prior work in the criminal justice system, of the purpose of the study in which they were involved, and of the nature of other studies within the sentencing project, including those carried out by other students in the class (we met with the various research groups individually). The students were told that their work in the class would consist of becoming familiar with interviewing techniques (by reading books about interviewing) and of applying this knowledge in interviews with the San Diego County Superior Court judges, in order to find out about the sentencing process. Students were also encouraged to consult people who were familiar both with interviewing and with the law and the operation of the courts, namely, the reporters of the

San Diego newspapers assigned to cover the events at the County Courthouse. They were further told that the general approach they should take was a "journalistic" one, which meant that they should (a) use journalistic interviewing techniques as much as possible, (b) be thoughtful, thorough, and responsible, and (c) eventually write a report that could conceivably be published in a San Diego newspaper.

After these instructions, the students were more or less on their own. They read texts on interviewing, familiarized themselves with the criminal justice system in California and with the sentencing process, and, above all, spent many hours with various newspaper reporters assigned to the Courthouse pressroom soliciting advice and suggestions concerning the kinds of questions one should ask of judges in order to find out a sufficient amount about sentencing in San Diego to be able to write a newspaper article about it. The students, working in teams of two, then proceeded to interview eight San Diego Superior Court judges. Finally, they wrote a joint report summarizing their findings; the report was subsequently published as a one-quarter-of-a-page article in a major San Diego newspaper (Persky, Sprague, & Lowe, 1975).

The reasons for several aspects of the procedure we chose in the interviewing study are of some interest. First, the use of interviewers who were unaware of the overall project and our past work in the criminal justice system assured that our own biases or misgivings about this type of methodology did not influence the conclusions that were reached on the basis of the interview data about the factors that affect sentencing. Second, whereas the interviewees—the judges—could be considered as "content informants," one might think of the newspaper court reporters as "methodological informants." If they were thorough and sincere (the students' impression was that they were), these people provided our interviewers with invaluable information about the details of the procedure by which a very important medium (the newspapers) obtains its information about the operation of the criminal justice system. Third, the publication of the article in a reputable newspaper proved that the students had done a piece of journalistic investigation and writing of at least barely acceptable quality—in terms of the number of interviews conducted, the style of writing, the conclusions reached, and so on.

In short, our study put the interviewing method to the test by simulating the process whereby the media themselves go about collecting information and presenting it to the public. An underlying assumption of this research is that information about the criminal justice system (as well as about many other aspects of social life) trickles down into public consciousness through the methodological and inferential filters imposed by the media. If it were subsequently found—by using other data-collection methods—that the conclusions reached in the published article were misleading or incorrect, our students' publication of the piece essentially would be a clear demonstration of the ease with which the public can be misinformed about key aspects of the actual operation of its most respected institutions, mainly due to the faulty methodology used by the chief information-providers, the media.

Finally, let us turn to the results of the study, or, rather, the conclusions reached in the published article about the factors that affect sentencing. The conclusions were that the sentencing decisions are exceedingly complex, that they are reached after a lengthy consideration and the full application of judicial training and wisdom, and—although there did not seem to be a consensus among the judges—that numerous factors were important in sentencing and all taken into account, including

the nature of the crime, the prior record of the defendant, his or her future behavior as a function of the type and length of sentence, the justification for the crime, the content of the probation officer's report, the content of the letters to the judge by the defendant and other people, sympathy, considerations regarding the defendant's family, chances of rehabilitation, and public cost of imprisonment.

In short, if one were to rely only on the interviewing method of the type employed in our study, the conclusions would be that (a) numerous factors affect, and are integrated into, the sentencing decision, (b) the decisions are highly complex, and (c) every case is different.

Questionnaire (the "Sociological" Approach)

In this study, the San Diego County Superior Court judges again served as subjects, but their sentencing behavior was now investigated by means of a structured questionnaire. The questionnaire is, of course, an extremely frequently used research tool in the social sciences, especially in sociology. Our objective in this study was to simulate, as closely as possible, the research procedure that a sociologist might use to investigate sentencing. Which conclusions would one reach about the factors that influence sentencing if one were to rely solely on the questionnaire? How would these conclusions compare with those reached on the basis of other methods?

Several sociology majors enrolled in our course (one in which the "journalistic" group members were also enrolled) were fully responsible for the development of the questionnaire, for its administration to the judges, and for the writing of the final report. As was the case with the "journalistic" group, the "sociological" research team was not aware of our prior work in the legal area, nor of our hypotheses. The group members were told that their primary objective in the class was to write a comprehensive report about the factors that influence the sentencing of adult felons, and that they should do this on the basis of a research study that would utilize the questionnaire—a method frequently used in the discipline in which they majored.

Members of the group first read about the criminal justice system and sentencing procedures in California, and then searched various methodology texts for information about the details of questionnaire construction. The next step—analogous to the "journalistic" group's use of the court-assigned newspaper reporters as "methodological informants"—involved lengthy consultations with several graduate students in the sociology department at the University of California, San Diego. With the assistance of these graduate students, the questionnaire team constructed an elaborate and detailed 25-item instrument dealing with sentencing. This questionnaire was informally administered to several undergraduates and further refined by removing ambiguities, improving the phrasing of the questions, etc. We felt that the final product was a polished instrument that did a reasonable amount of justice to the questionnaire as a method.

Over a one-month period in 1975, pairs of students arranged to see various judges in the San Diego County Superior Court. Of the 26 judges on the bench (the numbers of their departments are 1–27, but number 13 is missing!), 16 (61%) eventually filled out the questionnaire. Another three judges agreed to fill it out, but it later turned out to be impossible to arrange for an appointment because of the judges' busy schedules. Five judges refused the request to fill out the questionnaire, whereas another two were not even asked, because of illness, leave, etc. The students saw a judge individually in his chambers. The judge filled out the questionnaire himself, in the students' presence,

in sessions ranging from 30 minutes to 2½ hours. We thought it advisable for the students to be there in order to be able to answer any questions and, especially, to ensure a high return rate by preventing the questionnaire from disappearing into files, trash cans, or shredders. As was the case with the "journalistic" group, the students in the questionnaire team invariably reported that the judges had been friendly, informal, and helpful.

After the judges had filled out the questionnaires the students prepared a 30-page report summarizing the findings. The report was in four parts. The first described the sentencing process in California in rough outline. The second dealt with the various details of the questionnaire and the purpose of the various questions in it. In the third part of the report, the students presented and discussed the judges' responses on each of the 25 items; in this section, some statistics concerning the extent of the interjudge agreement on various items was also presented. In the final part of the report, the students described some general conclusions about the type and relative importance of various factors in the process of sentencing of adult felons.

The questionnaire dealt with the judges' treatment of many factors, including the defendant's age, sex, religion, family status, race, political views, military record, education, economic situation, addictions, prior criminal record, severity of the (present) crime (and with regard to the type of crime, the extent of premeditation, the involvement of alcohol or drugs, the amount of violence, whether or not a weapon was used, etc.). In addition, the judges were asked whether or not they took into account the defendant's attitude towards the police and other legal representatives, the defendant's remorse and/or constructive attempts to change his/her lifestyle, and the defendant's behavior during the present and past incarcerations, including his/her general response to previous corrective measures. The judges were specifically asked to pit various factors against each other in terms of importance (the probation officer's recommendation vs. the district attorney's recommendation). The judges were also asked a number of questions about the mechanics of the sentencing process, including about the amount of time they have per file prior to the sentencing hearing (especially in cases where there had been no trial), about the percent of cases in which the judges take part in the various plea- and sentence-bargaining transactions, and about the extent of their involvement in these matters. Finally, judges responded to general questions about their views of the various aspects of the penal code and of the correction system, and about their philosophies of sentencing. Thus, the contents of the questionnaire exceeded somewhat the issues that are of immediate concern in the present section.

The main conclusions in the report can be summarized as follows: (1) almost all judges listed at least four different factors as being highly important in sentencing; (2) the most important factors were: Severity of crime, prior record, family situation, employment status, and "special considerations" (such as drug and alcohol addiction, and mental disorders—presumably because alternative modes of disposition are available in these cases); (3) the significance of the role of the probation officer was consistently played down; this was also true for the roles of the defense and district attorneys, although to a much smaller extent; (4) even when the role of the probation officer was considered to be important, the emphasis was on his/her advisory and informational functions, not on treating the sentencing recommendation as an important *causal* factor; (5) despite the fact that the judges generally minimized the importance of the probation officer, the students' report itself, presumably going beyond the

judges' statements, concluded that the same factors that appeared to play a major part in the judges' sentencing behavior (severity of the crime, and the defendant's prior criminal record, family situation, and employment status) also influenced the contents of the probation report, as revealed by the fact that " . . . these factors are accorded a substantial amount of space in a probation report under the titles: The offense, prior record, and social factors" (Frerichs, McKinney, & Tisner, Note 5, p. 11); thus the students apparently concluded that the same factors *independently influence both* the probation officer and the judge, but did not imply a causal connection between the probation officer's recommendation and the judge's sentence; (6) despite their general predilection to list numerous factors and make the claim that all these factors are important and are seriously considered in the sentencing process, the judges flatly denied the importance of several other factors; notable among these "rejected" factors were the defendant's remorse, attempts to change his/her lifestyle, and the probability of successful rehabilitation.

Our initial plan had been to urge the students to attempt to present a paper dealing with their findings and conclusions at a sociological convention. Analogously to the efforts of the "journalistic" group, we hoped to simulate, as fully as possible, the manner in which the information about the sentencing process would be collected, analyzed, and *disseminated,* if one were to rely on a "sociological" questionnaire as the exclusive research tool. Moreover, we felt that the final report was of sufficiently high quality and contained sufficiently interesting information that there was a good chance that it would be accepted for presentation at a major convention. As it turned out, however, the results from our other studies of sentencing—including the archival analysis project—were beginning to come in and it was immediately obvious that these data would lead to very different conclusions. Because we felt that it would be unethical to pursue our "complete-simulation" idea to the point where highly suspect conclusions about the factors that affect sentencing would be disseminated with our knowledge, the paper by Frerichs, et al. (Note 5) remained unpublished.

Rating-Scale Responses

This study was an attempt to obtain direct ratings of the importance of various factors in the sentencing process. Three different subject populations, which differed sharply in the extent and type of their involvement in sentencing, were used: judges, defense attorneys, and college students.

Eight San Diego County Superior Court *judges* were seen in their chambers by pairs of our research assistants who introduced themselves as UCSD students working on a legal project. Ten randomly chosen judges had initially been approached, but two were unable to participate. The assistants were unaware of the other studies in the sentencing project. If a judge had been a subject in the interview or questionnaire study, and raised the issue, the assistants veridically claimed lack of any knowledge about such studies.

The judges were presented with a booklet, on the first page of which were brief instructions as to how to make the rating-scale responses. In anticipation of a common type of criticism, the judges were also told that the researchers were aware that such responses are quite different from the decisions the judges make in the courtroom, and that the responses may appear unnatural. They were urged to proceed nevertheless, on the grounds that scalar/numerical responses could be statistically analyzed. The remaining pages in the booklet contained rating scales dealing with the eight factors

discussed below.[1] On the top of each page was typed: "Relative to other factors listed in this booklet, how much importance do you attach to the factor listed below in your sentencing decisions in adult felony cases?" Below this question, which was followed by the name of one of the eight factors, was the rating scale. One end of the scale was designated "extremely important" and the other "totally unimportant." The judges were urged by the research assistants to look through the booklet and familiarize themselves with the range of factors before making any responses. The factors were chosen for inclusion in the booklet because they appeared—to us—intrinsically interesting and/or because they had been frequently mentioned by judges in other studies in the project.

On a 100-mm rating scale where one end was designated as "totally unimportant" (0) and the other as "extremely important" (10), the range of the means and standard deviations for the eight factors was 3.00–8.71 and .95–2.57, respectively. The "severity of the crime" (M = 8.71), "prior criminal record" (8.29), and "family situation" (7.12) factors were rated as the most important by the judges. These were also the factors that emerged as the most important in the questionnaire study. However, the "employment status" factor, which was also in the group of the four most important variables in the questionnaire study, was now rated as the *least* important of the factors considered (M = 3.00). It is also of interest to note that the "probation officer's recommendation" factor was rated to be of only average importance (M = 5.71), close to the "drug/alcohol use" (6.43), "probability of rehabilitation" (5.57), and "educational level" (4.94) factors, and significantly less important than the top three factors on the list.

In another part of this study, 33 *defense attorneys* in the San Diego area were visited in their offices by our assistants. Defense attorneys are a valuable subject population in that they do not themselves make sentencing decisions, but are nevertheless intimately involved in the decision-making process and closely observe it on a day-to-day basis. The defense attorneys in our study had a considerable amount of experience in criminal law—namely, an average of 7 years and 4 months of legal practice; moreover, in the previous 18 months, the attorneys had handled an average of 32.95 felony cases (felonies at the time of arraignment).

The attorneys filled out a long questionnaire which covered a wide variety of issues pertaining to their activity in the criminal justice system and their opinions and impressions of its operation. In one section of the booklet, the attorneys rated various factors in terms of their perceptions of the relative weight assigned to these factors by the judges in sentencing. Although the number of factors that the attorneys rated (on scales identical to those used by the judges) was considerably greater than in the case of the judges, each factor rated by the judges was also rated by the attorneys.[2]

The attorneys' ratings were similar to those made by the judges only insofar as the "severity of the crime" and "prior criminal record" factors were again considered

---

[1]In addition, the judges rated the severity of various types of prior criminal record and of various types of felonies. These results are not discussed further in the present article.

[2]The goals of studies in which the quality of the *performance* on the bench of *individual* judges is rated by defense attorneys are, of course, usually quite different from those of our investigation. Such—often very informal—rating studies are typically sponsored and conducted by various Bar Associations (e.g., Konon, 1978).

to be the most important. However, to our considerable surprise, the attorneys felt that the third and fourth most important factors in terms of the weight assigned to these by the judges in sentencing were "race of the defendant" and "defendant's income." In fact, race was considered to be almost as important as the top two factors. Since the judges had not been asked to rate either the race or the income factor, these results cannot, strictly speaking, be regarded as examples of blatant disagreements between the attorneys and the judges. However, it should be noted that in both the interview and questionnaire studies, the judges vigorously denied any effects of the defendant's race and income on sentencing. Furthermore, an examination of the attorneys' ratings of the relative importance of the remaining six factors that had also been rated by the judges revealed very little similarity between the two sets of ratings. For example, the "family situation" and "drug/alcohol use" factors, which were given quite high ratings by the judges, were at the very bottom of the defense attorneys' list.

It would seem, then, that except for the agreement concerning the "severity of the crime" and "prior criminal record" factors, the defense attorneys and the judges have very different opinions about the relative importance of the factors that affect the judges' sentencing decisions.

In the third part of this study, *college students* were used as subjects. These people rated the relative importance of the various factors in the sentencing process on the basis of the impressions formed in the sentencing hearings they attended. Two types of raters within this subject population were distinguished: "Naive" raters ($N = 27$), who did the ratings on the basis of observing no more than a total of four sentencing hearings, involving no less than two and no more than three different judges, and "experienced" raters ($N = 8$), whose ratings were based on observations of no less than 25 sentencing hearings, involving at least six different judges. Both types of college-student subjects rated the relative importance of 17 factors, including the eight rated by the judges, on scales that were identical to those used by the judges and defense attorneys.[8] Unlike the judges and defense attorneys, of course, college students had little or no knowledge of the details of either the law or of the behind-the-scene activities that surround the sentencing process. The students' ratings thus had to be based on (a) the "general-public" conceptions of the sentencing process and (b) exposure to the sentencing hearings, which was either considerable ("experienced" raters) or very limited ("naive" raters).

As it turned out, the number of sentencing hearings that the students attended did not make a great deal of difference in their ratings of the most important factors. Both the "experienced" and the "naive" student raters considered the "severity of the crime," "prior criminal record," and "probation officer's recommendation" factors to be of utmost importance, although the exact ranking differed somewhat for the two groups. Note that the students thus agreed with the judges and defense attorneys with regard to the importance of the severity-of-crime and prior-record factors, but attached far more importance to the probation officer's recommendation than either of the two other groups of subjects.

---

[8]The students also rated the judges' demeanor, courtroom behavior, and personality characteristics. These results are not discussed further in the present article.

Beyond the top three factors, the two groups of student subjects agreed very little either with each other or with the judges and defense attorneys, respectively. For example, whereas the "naive" raters (like the defense attorneys) considered the race of the defendant to be highly important, "experienced" raters thought this to be one of the least important variables. The opposite was true with regard to the rated importance of the "remorse" displayed by the defendant. Furthermore, whereas the "naive" raters agreed with the judges, but disagreed with the defense attorneys and "experienced" raters, in rating the relative importance of the "family situation" factor as quite high, they agreed with the defense attorneys, but disagreed with the judges and "experienced" raters concerning the importance of the "drug/alcohol use" factor.

It is, of course, by no means surprising that various groups of subjects—as different on so many dimensions as the judges, defense attorneys, and "naive" and "experienced" students are—would disagree in their ratings of the relative importance of various factors in sentencing. However, the extent and the type of the disagreement are by no means obvious and would not have been known without actually doing the study. Moreover, it is clear that had only one of the subject groups been used (as is so often the case in legal psychology), one might have been tempted into drawing some potentially quite misleading conclusions. Finally, it is important to note that quite apart from the differences in the results obtained for the three different groups of subjects within the rating-scale approach, the results of *all three* rating-scale studies differ from those obtained by using other research methods to study sentencing.

Experimental Simulations

In this study, three different types of subjects (judges, probation officers, and college students) were presented with brief descriptions of fictitious felony cases. Different levels of various factors that might be important in the sentencing process were defined by the wording of the descriptions. Many aspects of this procedure are, of course, extremely common in psycholegal research.

The *judges* (12 participated as subjects) were presented with fictitious case descriptions that defined a 2 (severity of the crime: forgery vs. armed robbery) × 2 (prior record: 0 vs. 2 prior felony convictions) × 2 (method of conviction: plea of guilty vs. trial) × 2 (family/employment situation: stable family and job vs. no family, no job) × 3 (probation officer's recommendation: prison vs. probation with some time in the custody of the Sheriff vs. straight probation) within-subjects factorial design. For each factorial combination a judge decided on a complete sentence, similar to those passed in real cases [e.g., 36-months probation, with 12 months in the custody of the Sheriff (i.e., in county jail)]. As in real life, the judges thus essentially had three basic sentence options: Prison vs. probation with time in custody vs. straight probation.

Twenty-two *probation officers* with a considerable amount of experience in compiling reports in felony cases were subjects in a similar five-factor within-subjects experimental simulation. Four of the factors were identical to those to which the judges responded (severity of the crime, prior criminal record, method of conviction, and family/job situation), however, an extra level was added to the severity of the crime factor (burglary). The probation officer's recommendation was replaced with remorse. It had two levels: The defendant was either described as showing a great deal of remorse with regard to the crime he had committed, or as showing no remorse. Like the judges, the probation officers were presented with sheets of paper, on each of which was a different factorial combination. The dependent measure was analogous to

that for the judges; the probation officers wrote out a full sentence recommendation, choosing in each case (i.e., each factorial combination) between the three basic sentence options already described for the judges.

Two groups of UCSD *students* enrolled in freshman- and sophomore-level courses were used as subjects in the third experimental simulation. For both groups, the experimental design was a $3 \times 2 \times 2 \times 2$ factorial involving the severity-of-the-crime, prior-criminal-record, family/employment-situation, and remorse factors, with the levels of the four factors identical to those described above for the simulations with probation officers as subjects. (The method-of-conviction and probation-officer's recommendation factors were not included in this study, mainly because they were concerned with aspects of criminal procedure with which most students are not familiar.) One version of this study involved a group of 480 students, each of which was assigned to only one of the 24 factorial combinations described above. Another group of 35 students responded to each of the 24 factorial combinations. Thus, in the case of the student subjects, the simulation was carried out on both a between-subjects and a within-subjects basis. In both cases, the dependent measure was the "duration of the prison sentence in years." This type of measure is very common in psycholegal research; indeed, our student subjects reported no difficulty in using the "years-in-prison" scale. Such a scale, however, could not be given to judges and probation officers; our pilot work indicated that these people simply refused to use this type of sentencing scale on the grounds that it is entirely different from the trichotomous choice (described above) that they have in real life. (This—another Brunswikian—point is instructive and typically overlooked by those researchers in the psycholegal area whose work consists exclusively of simulations with students as subjects.)

The data from the four experimental designs (two involving students and one each with judges and probation officers as subjects) were analyzed by the customary analysis-of-variance procedures. [These results are described in greater detail by Ebbesen & Konečni, Note 3, and Konečni & Ebbesen, in press (b)]. In the case of judges, four of the five factors were statistically significant. The defendants who committed more severe crimes, $F(1,11) = 27.68$, $p < .001$, had more extensive prior criminal records, $F(1,11) = 25.47$, $p < .001$, had been found guilty in a trial, $F(1,11) = 8.80$, $p < .05$, and for whom the probation officer's recommendation was more severe $F(1,11) = 6.69$, $p < .05$, were given harsher sentences by the judges (i.e., the proportion of prison sentences was higher). In contrast, the judges felt that the family/job situation of the defendant was totally irrelevant $F(1,11) = 1.73$, $p =$ ns. Thus, the judges felt that the severity of the crime and the extent of the prior criminal record of the defendant were by far the most important factors, and about equally so.

The simulation with probation officers as subjects revealed that the extent of the prior criminal record was by far the most important factor, $F(1,21) = 81.81$, $p < .001$, followed by the severity-of-the-crime, $F(2,42) = 13.70$, $p < .001$, and family/job-situation and remorse variables $F(1,21) = 8.33$, $p < .01$, and 8.94, $p < .01$, respectively.[4] The method of conviction factor was not significant, $F = 1.49$.

[4]The difference between the results of the simulations with judges and probation officers as subjects, in terms of the relative magnitude of the effects of the prior-record and severity-of-the-crime factors, remains even after the burglary level of the crime factor is removed from the probation officer data set: $F(1, 21) = 25.47$ for the severity-of-the-crime factor (still far smaller than the F-value for the prior record factor).

Thus, the probation officers disagreed with the judges both in terms of the relative importance of the prior-criminal-record and the severity-of-the-crime factors, and with regard to the significance of the method-of-conviction and family/job-situation variables (where complete reversals occurred).

In the between-subjects simulation with college students serving as subjects, prior criminal record, $F(1,456) = 19.71$, $p < .001$, severity of the crime, $F(2,456) = 13.38$, $p < .001$, and remorse, $F(1,456) = 6.46$, $p < .01$, were all significant; in contrast, the family/job factor was far from significant, $F < 1$. On the other hand, in the within-subjects student simulation, all four factors were significant: for severity-of-the-crime, $F(2,68) = 34.24$, $p < .001$, for prior-criminal-record, $F(1,34) = 30.20$, $p < .001$, for remorse, $F(1,34) = 24.96$, $p < .001$, and for family/job, $F(1,34) = 11.08$, $p < .01$.

As is clear from the above results, given that one has chosen the experimental-simulation method of the type used here to study sentencing, quite different conclusions would be reached on the basis of data obtained from different types of subjects. Even when there are certain similarities between the results of the four experimental designs (for example, severity of the crime and prior record were significant in all four), the finer details differ (such as the relative importance of the two factors with regard to each other and other factors). Certain factors seem to be especially sensitive to the type of subjects and experimental designs that are used. For example, the family/job-situation factor was judged as important by the probation officers and students in the within-subjects design, but quite unimportant by the judges and students in the between-subjects design. Such inconsistencies cannot be explained in any reasonable way and underscore the problems with using experimental simulations as a research methodology. The overall differences between the findings obtained in the within- and between-subjects designs with students as subjects are perhaps even more worrisome, in the sense that the methodological difference between the two studies at first appears to be far more trivial than that between the studies using different subject populations.[5]

Observation of Live Sentencing Hearings

The four studies described so far involved quite different research methods, but were all simulations in the sense that none dealt with real-life cases in which the judges pass actual sentences. From our point of view, an essential next research step was to examine the decision strategies which the judges use in real-life cases where they match the sentence options available to them to actual offenders.

As was the case in our research on the courtroom (as opposed to the simulated) bail-setting behavior of the judges, our objective in the real-world studies of sentencing was to analyze the covariation (over cases) between a large number of potential "predictors" and the final sentence decisions. The first step, as in the case of bail setting, was to determine the sources of information available to the judges prior to

[5]The nature of the differences between the results of the within- and between-subjects designs cannot be explained simply by the greater precision of the within-subjects design. Namely, there was a reversal of the order of importance (variance accounted for) of the severity-of-the-crime and prior-criminal-record factors across the two designs. Furthermore, the higher F-values in the within-subjects design were due to a much greater spread of the means for the levels within each factor, rather than merely to smaller variances.

reaching their decisions. Two general sources were isolated. The first consists of the events that occur in the sentencing hearing. The second source of information is a file which the judge reviews prior to the hearing. The present and the next subsection of the article deal with our research on these two sources of information.

The hearing at which the judge imposes the sentence occurs after a defendant has pleaded guilty to a felony charge or been found guilty by a jury. By the time of the hearing, the judge has presumably consulted the probation report, which he or she typically receives only about 24 hours earlier. Five major participants are present at the hearing: The judge, the assistant district attorney, the defense attorney, the offender, and the probation officer. Sentencing hearings seldom last more than five minutes, which is certainly not the impression one would get by focusing only on well-publicized cases (such as the Patty Hearst case).

The purpose of our observational study of sentencing was to break down the various types of information to which the judges are exposed in sentencing hearings into a number of "predictor" variables and unobtrusively code the values of these variables in live hearings. Do any of the factors that can be isolated from the content of the hearing affect the sentencing decision? How well can the sentencing decision be predicted on the basis of the information available exclusively in the hearings (as opposed to various files, written documents, and so on)?

A time-sampling procedure was used to code the verbal exchanges in the sentencing hearings. Our trained assistants observed and unobtrusively recorded every 10 seconds who was talking (one of the five participants mentioned above) and the topic being discussed. A coder had a list of over 70 content categories [see Appendix 1 in Ebbesen & Konečni, in press (c), for a complete listing] printed on a reference sheet in front of him/her. The form on which observations were recorded consisted of a five-person (participants in the hearing) $\times$ n time-interval matrix. The observer recorded an appropriate content code at the end of each ten-second interval in the row representing the person who was speaking. This procedure produced, for each sentencing hearing, a string of codes indicating who talked when, about what, and in response to whom. Prior to the beginning of each sentencing hearing, the coders also rated on ten-point scales the appearance of the defendant (physical attractiveness, dress, etc.). Following the hearing, the coders also rated the grammatical quality of the defendant's speech. Finally, they indicated whether the defendant appeared attentive or indifferent to the proceedings. A total of over 400 sentencing hearings was coded in this manner during 1976 and 1977. Over 30 coders were used at different times. The reliability of the coding was quite high [see Ebbesen & Konečni, in press (c), for a detailed description of the various reliability indices that were used].

The next step was to statistically analyze the relationship between the various factors that had been coded and the sentencing decisions. To our amazement, none of the factors coded in the sentencing hearings were associated in a statistically significant manner with the final sentencing decision. This was true for various demographic variables that could be coded in the live hearing, such as the offender's race, sex, and age, especially when factors such as the type of crime and the extent of the offender's prior criminal record were statistically controlled for (information about the latter two factors had to be obtained from the court files after the sentencing hearings because it was not clearly presented in the hearings themselves). Similarly, the more

social-psychologically relevant attributes of the defendant, such as his or her physical attractiveness, made no difference with regard to the sentence that was imposed.

Finally, none of the formal and content aspects of the sentencing hearings themselves seemed to make any difference. In another article [Ebbesen & Konečni, in press (c)], we reported detailed results concerning (a) various measures of verbal participation by the five major actors in the hearing (the judge, the defendant, etc.), (b) the content of the discussion and the degree of its favorableness to the defendant, and (c) the conditional probabilities that a certain participant in the hearing would speak after another participant had spoken. These data are very interesting and convey the flavor of the sentencing hearings. However, when the verbal-exchange aspects of the hearing are treated as predictors, one finds no statistically significant relationships between these predictors and the sentencing decisions. The single exception was the finding that judges tended to give somewhat lighter sentences (fewer impositions of prison terms) in those cases in which the assistant district attorney *and* the defense attorney raised more positive than negative points; however, even this trivial effect disappeared when we controlled for the type of crime and the extent of the prior record of the defendant (by consulting the court files after the hearing).

The findings of this study suggest that the observational method is a completely inappropriate research tool to study sentencing. One could presumably spend many years sitting around courtrooms and coding the sentencing hearings without being able to isolate any strong predictors of the sentences imposed. This is presumably so because the primary function of sentencing hearings may be a purely ritualistic one of maintaining the public image that the operation of the criminal justice system is a process open to public scrutiny. Little that is said or done in the hearings seems to affect the sentencing decisions, at least within the range of predictors we examined.

Another point is in order here. In the next section of the article, in which we discuss the archival-analysis research approach to sentencing, four factors that *are* significantly associated with the sentencing decisions will be described. The predictors in question are: (a) severity of the crime, (b) prior criminal record of the defendant, (c) whether the defendant was in jail or out on bail while awaiting trial (i.e., prior to the plea of guilty), and (d) the probation officer's recommendation. We mention these significant predictors at this point because one could, quite reasonably, think that the information concerning these factors would also be presented in the sentencing hearings (that is, in addition to being available in the court files), and that, even if the sentencing hearings do not contribute any *unique* predictors of the sentencing decisions, at least they do not mask the predictors available elsewhere, *away* from the public view. As it happens, this is not so. The in-jail vs. out-on-bail variable is never discussed in sentencing hearings. The probation officers seldom verbalize their recommendations, and, in fact, very rarely speak about anything. Only 3.2% of the duration of sentencing hearings we examined was spent on the probation officers' verbal contributions; in only 19.1% of 404 cases did the probation officers make at least one utterance; and the mean length of their utterances in these cases was only 8.2 seconds [see Table (1) in Ebbesen & Konečni, in press (c)]. Even when one of the attorneys in a hearing agrees with the probation officer's recommendation, he/she typically simply states the agreement, without specifying what he/she is agreeing to. Furthermore, the severity of the crime and the extent of the defendant's prior criminal record are discussed by the various participants in a sketchy and/or partisan manner,

so that one cannot get a complete picture of these aspects of the case in the open hearing. When all of these facts are coupled with the speed with which the sentencing hearings are carried out, the mumbled sentences and the incoherent speech of the participants, and the general clamor that characterizes courtrooms (in sharp contrast to the air of dignified silence that the public typically associates with them), it is no wonder that even the predictors of the sentencing decisions that are available elsewhere (i.e., in the court files) do not emerge in these "public hearings," which are, for all practical purposes, merely a smoke screen for the actual decision-making process.

Archival Analysis of Files Pertaining to Sentencing

After the sentence has been passed, a file pertaining to the case is placed at the County Clerk's office. This file contains the probation officer's report and various other documents (e.g., information about the prior criminal record of the defendant, copies of the various documents filed by the district attorney's office, a bail unit report, information about the guilty/not guilty plea, and so on). This is the same file that the judge views one day prior to the hearing. In San Diego County the file is in the public domain, with the exception of the probation report, which cannot be reviewed by the members of the public once 30 days after the sentencing have elapsed.

An instrument quite different from that used for sentencing hearings was developed for the purpose of coding these files. [The entire instrument, consisting of several hundred predictors, is available in Appendix II in Ebbesen & Konečni, in press (c).] Trained coders (close to a hundred were used in 1976 and 1977, the two-year period during which the study was carried out) worked alone in the county facilities transferring information in the files to our coding instrument. Such things as (a) the date of the hearing, (b) the judge's name, (c) the probation officer's name, (d) demographic characteristics of the defendant, (e) the original charges (on the arrest report), (f) the charges that the defendant was convicted of, (g) court-related data concerning prior custody, preplea agreements, bail, and such, (h) aspects of the crime (e.g., nature and number of witnesses and of types of physical evidence), (i) the content of the defendant's statement (e.g., the kinds of factors listed to explain the criminal activity), (j) prior record, (k) employment and social history, (l) medical and psychiatric information, (m) the number of lines the probation officer used to describe positive and negative aspects of the defendant in the evaluation section of the report, (n) the details of the probation officer's sentence recommendation, and (o) details of the final sentence, were coded. The details of the charges and prior record were coded in terms of the California Penal Code. Rating scales were used to code such things as the degree of remorse, the apparent premeditation, the extent of admitted guilt, and the intention to improve expressed by the defendant in his or her statement. Counts of the number of lines dedicated to various topics served as a reliable technique for coding other more variable content areas (see Konečni et al., in press, for a discussion of this procedure). Some content categories required that coders indicate which of a number of predefined topics were raised. As was the case with the sentencing hearings, the reliability of coding was very high [see Ebbesen & Konečni, in press (c), for details]. A total of over 1000 files was coded.

The major statistical analyses used the following four major sentence categories as the "dependent measure": (a) prison as prescribed by law; (b) some period of time

in the county jail, almost always followed by a probationary period; (c) "straight" probation, without any incarceration; and (d) all others (such as the commitment to a mental hospital, a fine without a period of incarceration or probation, and so on).

As was the case in the observational study, after examining the relationships between all of the possible predictors of sentencing that we coded from the files and the actual sentences, we discovered that extremely few of the factors were associated in a statistically significant manner with the final sentencing decision. For example, none of the offender variables that were also coded in the sentencing-hearing study (race, sex, age), nor other offender attributes, which could not be coded in the sentencing-hearing study (e.g., religion, education, marital background, social history, and so on), emerged as statistically significant predictors of the sentencing decisions. Again, this is especially true when the severity of the crime and the extent of the offender's prior criminal record were statistically controlled for. The factors relating to specific features of the criminal activity and the offender's justification for the crime were also not significantly associated with the sentence.

In fact, the statistical analyses revealed that it was reasonable to ignore the relationships between *all but four* of the predictors and the final sentence, in the sense that very little would be lost in terms of the ability to predict the sentence. The four factors that accounted for almost all of the systematic variation in the sentencing decisions were (they were already listed in the section on the observation of sentencing hearings): (a) the *type of crime,* (b) the extent of the offender's *prior criminal record,* (c) the *status* of the offender between his/her arrest and conviction (i.e., was he/she released on his/her own recognizance, released after paying a certain amount of bail, held in jail, or was first in jail and then released on bail; to maintain consistency with our other articles, this variable will be referred to as "status" in the subsequent discussion), and (d) the *probation officer's sentence recommendation.* [See Ebbesen & Konečni, in press (c), for the details of these results and statistical analyses.]

The identification of the four predictors of the sentencing decisions was only the first analytical step. The next step was to examine the relationship between the predictors. One set of analyses revealed that the relationships of the prior-record, severity-of-the-crime, and bail-vs.-jail factors to sentencing are independent of each other. Thus, for example, the status factor (which can be considered *extralegal,* because no sentencing guidelines recommend that it be taken into account in making the sentencing decisions) was not significantly associated with the sentences by virtue of being correlated with the offender's prior record or the severity of the crime. Such findings, of course, do not establish the causal role of the three factors. In fact, it is conceivable that all four predictors are differentially associated with some unmeasured factor which is the *single* real causal variable. Alternatively, the four factors might be correlated with several different causal factors, each to a varying degree. While these explanations could not be conclusively discounted, it is quite difficult to imagine what these other causal factors might be, given the very large number of predictors examined in the study.

The next step was to formulate and evaluate various possible causal models of sentencing, each of which would imply a somewhat different causal chain involving the four predictors (cf. Heise, 1975). Namely, it is possible that only one or two of the four factors are the direct causes of the sentence and that other factors are causes of these causes. Several temporal features of the system make certain chains less likely

than others. For example, it is always the case that prior record, severity of the charge at the time of conviction, and status are determined earlier in the processing of a case than are the probation officer's recommendation and the judge's sentence. Whereas it is possible to imagine a situation where the final sentence causes, for example, the prior record (e.g., through selective reporting or alteration of "rap" sheets that describe the defendant's prior criminal record), the occurrence of such activities was relatively unlikely in the stated circumstances. Accepting for the moment at least the temporal order of events in the processing of a felony case as relevant causal evidence, several logically tenable causal models (chains) can be generated. In one such model, prior record, severity of the crime, and status would be assumed to be the direct causes of the probation officer's recommendation, but these variables would not be assumed to have a *direct* causal link to the judge's sentencing decision. Instead, the judge would be assumed to respond directly only to the probation officer's recommendation (probation officers' recommendations and the judges' sentencing decisions were in agreement in about 85% of the cases studied).

In another possible causal model, the prior record, severity of the crime, and status factors would be assumed to have direct effects on both the probation officer's recommendation and the judge's sentencing decision, but the decisions of the two participants (the probation officer and the judge) would not be considered to be causally related to each other. In other words, the high agreement between the recommendations and the sentencing decisions would be a spurious consequence of the fact that both the judges' sentences and the probation officers' recommendations are caused by the same set of three variables.

A third model actually reverses what might be assumed initially to be the temporal order of events and argues that the probation officer's recommendation is directly caused by the judge's sentence, which is, in turn, caused by the three case factors. One possible interpretation of such a temporal reversal would be based on the assumption that the judge is committed to a specific sentence agreement made between the district and defense attorneys in exchange for a plea of guilty. If the probation officers are aware of such agreements, they may be motivated to match and justify such sentence agreements and therefore may compose their reports and recommendations in *correct anticipation* of the judge's sentencing decision.

The tenability of these three causal models was quantitatively evaluated by log-linear analyses [we used the method recommended by Goodman, 1972, 1973; see Ebbesen & Konečni, in press (c), for details of these analyses]. These analyses revealed that the causal model which assumes that the three case factors (prior criminal record of the defendant, severity of the crime, and status) cause the probation officer's recommendation, which in turn causes the judge's sentencing decision, provides the most satisfactory explanation of the data and is to be preferred over the other two models.

However, several additional models had to be considered before accepting the one just outlined. For example, it seemed possible that any one (or more) of the three prior factors could have a *direct* causal influence on the judge's decision over and above the influence of the probation officer's recommendation. Indeed, the appropriate analyses revealed that the adding to the basic model of a direct causal link between the status variable and the judge's sentence produced a significant increase in the ability of the model to fit the data. This, however, was not the case for the prior-

record and severity-of-the-crime variables.

In summary, the probation officer's recommendation appeared to have a *direct* causal influence on the judge's decision. The relationships between the judicial decision and crime and between that decision and prior record seemed to be entirely due to the fact that the probation officers adjust their recommendations to these factors and the judges then follow the probation officers' recommendations. While a similar pattern held for the offender's status, it also seemed that this factor has some additional, though very small, direct influence on the judge.

The picture of the sentencing process which emerges on the basis of the archival analysis of files pertaining to sentencing thus differs in major ways from the conclusions that one would reach by using the various methods described earlier. These differences will be briefly summarized in the next section. The reasons for having more confidence in the results of the archival analysis than in those obtained by other methods will be outlined in the final section of the article.

A Comparison of Results Obtained by Different Methods of Studying Sentencing Decisions

It is clear that very different conclusions about (a) which factors affect sentencing decisions, (b) what their relative importance is, and (c) the factors' positions in the causal sequence, would be reached by using different methods, different subject populations, the same subject populations exposed to the sentencing process for different lengths of time, the same subject populations in different experimental designs, or various combinations of the above. With regard to the differences between methods, it can be safely stated that no two methods produced identical results in terms of the type, number, and importance of various factors. At one extreme, no important factors whatsoever could be identified by using the method of observation/coding of live sentencing hearings; at the other, on the basis of the journalistic interview, one would have been led to believe that no less than eight factors were highly important. More commonly, factors that emerged as very prominent on the basis of one method appeared to be of negligible importance according to another method. Within the same methodological approach, the responses of different categories of participants in the sentencing process (e.g., judges and defense attorneys) differed a great deal from each other, and both differed considerably from the responses of college students. Even when the same method (e.g., rating scales or experimental simulations) *and* the same subject population (e.g., college students) were used, major differences in the results emerged as a function of factors such as the duration of exposure to the legal system (e.g., "naive" vs. "experienced" observers in the rating-scale study) and the type of experimental design (e.g., between- vs. within-subjects studies with college students in the experimental simulations project).

In the final section of the article, we will outline some of the reasons why we think that the results obtained by the archival analysis can be trusted more than the results from the studies using different methods. It is, therefore, of interest at this point to summarize briefly the results obtained by each of the methods and compare them to those obtained by the archival analysis.

Judging by the *interview* results, the judges' sentencing decisions are exceedingly complex, take a great deal of time to make, and are based on a careful evaluation of numerous factors. In contrast, the analysis of the files pertaining to sentencing revealed that a single factor (the probation officer's recommendation), the information about which is presented in a few lines of text and can presumably be processed in

a few seconds, makes possible a very accurate prediction of the judges' sentencing behavior. Although three other factors were included in the complete causal model, these other factors (with the negligible exception of the offender's out-on-bail vs. in-jail status) appeared not to have a direct causal effect on the judges' sentencing decisions. In other words, the form of the causal model implied that the judges' sentencing behavior is such as it is only because the probation officers' recommendations are a part of the sentencing process. Because these recommendations summarize other variables (prior record, severity of the crime, and status), their removal from the legal system (and, thus, elimination from the causal chain) would presumably cause considerable changes in the sentencing decisions, in that the judges would be forced to combine the information concerning the three mentioned factors themselves and/or would perhaps take into account other factors not presently given a lot of weight by the probation officers.[6]

The *questionnaire* results, like those of the interview, suggested that the severity of the crime and the prior criminal record of the defendant are important, but additional factors (family situation, employment, drug use) were also stressed. The findings of this study completely failed to match those of the archival analysis in that the role of the probation officer was consistently played down by the judges.

In all of the *rating-scale* studies, the severity-of-the-crime and prior-record factors emerged as important, again together with a variety of other factors which were somewhat different for each of the subject populations studied. In addition, the judges in the rating-scale study indicated that the probation officer's recommendation may have some, moderate, importance, but both the "naive" and "experienced" students thought that the probation officer's recommendation was as important as the severity-of-the-crime and prior-record factors.

In the *experimental simulations*, the severity-of-the-crime and prior-record factors again emerged as important in all of the studies. Furthermore, the simulation with the judges as subjects also revealed that they regarded the probation officers' recommendation as important, but far less so than the other two factors. Simulations with other subject populations again indicated that these subjects believed several

---

[6]The contrast between the journalistic-interview account of sentencing and the model which emerged on the basis of the statistical analysis of archival data—especially regarding the importance of the role of the probation officers' recommendations in the latter model—is quite ironic if one were to take seriously another journalistic account of the sentencing process, one that was published in the Los Angeles *Times* (Kistler, 1975, who was *not* a student of ours!). In this article, one finds the following statements:

> The county Probation Department has not been able to maintain even 'a minimum standard of service' to the courts, a Superior Judge told a hearing on the firing of Chief Probation Officer . . . ;
>
> and
>
> The department's deficiencies [the] judge . . . testified, have existed for years and embrace such basic skills as the insufficient ability of many probation officers to read and write at an acceptable court level.

Whatever the quality of their writing and spelling in the rest of the probation reports is, the probation officers would clearly be well advised to clean up the three or four lines at the end of each report where their sentencing recommendation is summarized, lest they cause major perturbations in the pattern of sentencing decisions by confusing or annoying the judges.

other factors to be important in addition to the severity of the crime and prior criminal record.

Several aspects of these findings and comparisons are of interest. First, no method other than the archival analysis revealed that the out-on-bail vs. in-jail status of the defendant was an important factor in sentencing. Second, the results of only a few of the methods even hinted that the probation officer was of some importance, and none led to the inference that the probation officer's recommendation was a primary (let alone only) factor. Third, literally all methods indicated that the severity of the crime and the offender's prior criminal record are highly important. If all of our studies, *except* for the archival analysis, had been carried out, the conclusion about the major importance of these two variables in the sentencing process would presumably have been made with a great deal of confidence, due to the fact that it would have been based on the results which represent a point of convergence of many quite different methods (cf. Webb et al., 1966). Yet, such a conclusion would be entirely wrong, because (a) the role of the probation officer's recommendation as a *primary* cause of sentencing decisions would be ignored, (b) the status variable would be completely ignored, and (c) the exact position of the prior-record and severity-of-the-crime factors within the two-tier causal model would not be apparent. Note that the latter two errors in the description of the judges' sentencing behavior also characterized the otherwise most accurate simulation-based account of the sentencing process, namely, the rating-scale studies with—paradoxically—the "naive" and "experienced" college students. (In these two studies, the severity of the crime, prior record, and probation officer's recommendation emerged as the most—and equally—important factors.)

Finally, even when a *causal model,* as opposed to a mere list of important factors, is stumbled upon by researchers using simulations—essentially by means of speculations that go a long way beyond the actual results—such a model has a low probability of being correct. Namely, the students who wrote the report on the basis of the judges' responses to the "sociological" questionnaire proposed a relatively sophisticated causal model in which the same four factors (severity of the crime, prior record, family situation, and employment) were seen as independently causing both the judges' sentencing decisions and the probation officers' recommendations. However, recall that when a similar model was statistically pitted against the one finally accepted, it turned out to be quite inadequate. Thus, even when the use of a particular simulation methodology did lead to the accumulation of a body of data from which an intuitively appealing causal model of the sentencing process could be developed, the final product was disappointing.

SOME METHODOLOGICAL CONSIDERATIONS

In attempting to cope with external validity problems in our psycholegal research, we have typically used different subject populations, research settings, experimental materials, and research designs. Above all, we have relied on more than one research methodology to study a particular problem, thus generally following the Campbellian tradition (e.g., Campbell & Stanley, 1963; Webb et al., 1966). The use of the multiple-method approach is supposed to greatly increase one's confidence in a

conclusion—provided that several, or all, of the methods lead to it—in comparison to a study in which only one method had been used. Unfortunately, in our work, the results from the multiple methods have seldom led to a single, common conclusion.

If this were to happen in many other areas of research, an impasse would result that could not be resolved by applying *a priori* logical criteria. However, we believe that when one studies an intact functioning social network—such as the criminal justice system—there are often certain logical and practical criteria that can be applied. These lead one to trust the conclusions reached by one method over those reached by another on *a priori* grounds, with the important proviso that the researcher is interested in how the system actually operates, rather than in the phenomenology of the participants.

We believe, first of all, that the results of research efforts that deal with the real-world, consequential legal decisions are far more informative than those that deal with simulated decisions. This implies, among other things, that the decision to collect data from the participants and/or in settings within the criminal justice system in itself by no means resolves all problems regarding generalizability and external validity. For example, the decision to go to the judge's chambers and conduct an interview or a rating-scale study about the factors that affect his/her sentencing decisions may lead to conclusions about the causes of sentencing that are quite incorrect. The judge may deliberately try to mislead the interviewer for self-presentation or political reasons, or else may be quite unaware of the factors that he/she is actually taking into account in sentencing. Thus, neither the decision to deal with the actual participants in the legal system, nor the decision to collect data in legal settings, guarantees that the findings will lead to the discovery of the real causes of a participant's behavior and therefore have external validity.

One could argue that some simulations are better than others and that many problems can be avoided by conducting "good" simulations. However, to the extent that a simulation is trying to discover something about the operation of the real-world legal system, how can one know whether a simulation is "bad" and which of several simulations is the "best," without actually collecting data not only in naturalistic settings and with "real" participants in the legal system, but on "real" legal decisions? Moreover, if one accepts the view that on logical grounds only a real-world study on real-world decisions can validate the results of a simulation, it only makes sense to *begin* a research program by doing real-world studies (especially in a young and largely unmapped discipline such as legal psychology), and that in situations where there are limited funds, time, and personnel—frequently encountered in the social sciences—the choice as to which type of study to do is obvious.

When the above reasoning is applied to our work on bail setting and sentencing, it follows that we would have the most confidence—on *a priori* grounds—in the results of the observational study of bail setting (as opposed to the simulation study, even though real judges were the subjects), and the observational and archival-analysis studies of sentencing (as opposed to the four simulation studies). The use of the observational method to study bail hearings resulted in the formulation of an accurate, powerful, two-tier causal model of bail decisions (the severity of the crime factor is an excellent "predictor" of the district attorney's bail recommendations, which, in turn, accurately "predict" the judge's bail decisions). Similarly, an accurate and powerful

two-tier model of sentencing decisions was formulated on the basis of the results of the archival analysis. In contrast, however, no predictors of the sentencing decisions emerged as important in the observational study of actual sentencing hearings.

Thus, even in the context of a strategy of studying real-world legal decisions, a particular data-collection methodology sometimes "works" and sometimes does not. Perhaps because bail hearings are simple and easily coded (very little information is presented) and, especially, the bail recommendations of the participants (including the assistant district attorney's) are stated explicitly and *orally* (as opposed to being presented to the judge in a written document), the observation/coding of bail hearings was successful. In sentencing, on the other hand, perhaps because the important information is hidden in a file and the hearings themselves represent merely a ritual show, the results of the archival analysis were far more informative than those of the observational method. Such an outcome could not have been predicted: One could have reasonably expected that, for example, some aspect of the participants' verbal interaction, about which codable information was not available in the files, would be an important predictor of sentencing. The lesson to be learned from this is that even the strategy of studying the real-world legal decisions, as opposed to the simulated ones, is not by itself foolproof. The rule "use as many different methods as you can" still applies; one of its consequences is that different types of information available to a decision maker (e.g., the contents of written documents and oral hearings) may be treated as sources of "predictors" of the criterion decisions.

In general, however, when the archival analysis is possible (in the case of bail-setting, there were no written documents to be coded), we are inclined to trust the result of this method more than those of any other, provided that as many as possible of the following criteria are met.[7]

1. The coding categories used are similar to those used by the participants in the real-world system, rather than derived from the currently popular social-psychological theories. This typically means that the coding categories will be concrete and low-level, as opposed to abstract and high-level. An example would be the coding of a category "prior record" in terms of the actual number of prior felony convictions, rather than coding "consistency of prior criminal behavior" (a more abstract concept derived from attribution theory) on a 5-point scale. Such a procedure makes the coding more reliable and also facilitates the communication of the research findings to the participants in the legal system—if one's goal includes producing change in the system.

2. Coding is as exhaustive as possible, covering as much information in the written materials in the file as possible, so that initially a very large number of "predictors" (that is, coding categories) is isolated. This step, of course, minimizes the likelihood that an important predictor will be omitted from the analysis.

3. The statistical analyses examine the effects on the criterion decision of various combinations of "predictors," with a relatively large number of predictors in each "predictive set," so that both main effects and interactions can be captured.

4. Prior to archival analyses, sufficient amount of background research had been done by the investigators concerning the actual, routine, day-to-day operation of the

---

[7]The following section relies heavily on Konečni and Ebbesen [in press (c)].

system so as to leave no doubt that the file that is being coded is, in fact, at the disposal of the decision maker prior to the time when the decision is being made. Note that whereas it is important to demonstrate that the decision maker *could* have seen a particular bit of information in the file, it is not necessary to demonstrate that he/she has actually done so, especially for every case. It may well be that the bits of information that are being coded are correlated with other bits of information at either the same or higher level of abstractness and that the decision maker is actually attending to these other bits of information as he/she examines the file. This, however, in no way precludes treating the *coded* categories as "true predictors." (In fact, one could argue that even if the decision-maker does not see a file, a predictor isolated from the file that accounts for a very large percent of the variance in the decisions could be considered a "true cause.")

Given that such precautions have been taken, the coding of archival materials may have an advantage over the observation/coding of hearings in that (a) more predictors are typically available in the written materials, (b) the nature of the two research situations is such that greater coding reliability can be obtained in the archival case (because of the time and other pressures in observational research), (c) archival research is less obtrusive (although this does not necessarily always have to be the case), and finally, (d) when one examines the system as a whole, it is clear that written materials accompany a defendant through the system; therefore, the predictor of a particular decision that is discovered in the written materials is also more likely to be the predictor of many subsequent decisions by other participants in the system, by virtue of the same piece of paper (such as the "rap sheet," that is, the prior record of the defendant) being a part of the case at almost every node in the system.

An additional advantage of the archival analysis is that certain predictors that can be isolated from the files *temporally precede* the location of other predictors in the causal chain. Occasionally, this means that the best predictor of a particular legal decision may be a factor available in the files (and thus one that can be discovered by both researchers and participants in the system) *before* the offense is even committed! For example, in a study of the processing of the mentally disordered sex offenders (MDSOs) in California (Konečni et al., in press), we found that the convicted offender's prior sex-related criminal record almost inevitably led the court-appointed psychiatrists to diagnose and classify the offender as "sexually deviant" and an MDSO which, in turn, resulted almost automatically in the judge's verdict that the defendant be sent to a mental hospital (rather than be remanded to the trial court for sentencing). In other words, the offender's prior sex-related criminal record is an excellent predictor of both the final and intermediate decisions, and this information is known even before the offense under consideration has been committed. The psychiatric diagnosis and classification, and the judge's verdict, may be correlated with many other predictors, but the simplicity and temporal primacy of the prior sex-related criminal record factor forces other predictors into the role of epiphenomena. For example, differences in the content of probationary reports and psychiatrists' letters can be considered as merely serving to justify an already formed conclusion based on the prior sex-related criminal record, in order to give the appearance of complexity to the processing of MDSOs and smooth out the rough edges of the causal sequence.

What should one do in situations where the research on the real-world legal

decisions cannot be carried out? For example, many aspects of the legal system are confidential. It is impossible for researchers to be present during jury deliberations and it is extremely difficult to obtain access to files containing information that leads to certain decisions (e.g., the prosecutor's files). Many would probably think that simulation research in these cases is fully justified even if all of our criticisms are correct. A more cautious point of view, and one that we favor, is that erroneous information obtained by scientific methods (and therefore having an aura of truth) is more harmful than no information at all, especially when issues as sensitive as legal ones are being dealt with, and people's futures are quite literally at stake.

The point of view just mentioned seems to us to be particularly valid when one is dealing with poorly understood aspects of the legal system and/or is just beginning the research on a particular legal decision. Under some circumstances, simulations may be useful, especially as tools to tease apart further quesions about the real-world process. Even then, however, rather than automatically assume that simulations are useful, one ought to collect sufficient evidence to test whether they have captured the necessary details of the real world to be *real simulations.*

REFERENCE NOTES

1. Miller, G. R., Fontes, N. E., Boster, F., & Sunnafrank, M. Methodological issues in jury research: What can a simulation tell us? Paper presented at the 85th Annual Convention of the American Psychological Association, San Francisco, California, 1977.
2. Bray, R. M. The mock trial: Problems and prospects for jury research. Paper presented at the 84th Annual Convention of the American Psychological Association, Washington, D.C., 1976.
3. Ebbesen, E. B., & Konečni, V. J. Fairness in sentencing: Severity of crime end judicial decision making. Paper presented at the 84th Annual Convention of the American Psychological Association, Washington, D.C., 1976.
4. Vidmar, N. Group-induced shifts in simulated jury decisions. Unpublished manuscript. University of Western Ontario, 1972.
5. Frerichs, J., McKinney, A., & Tisner, P. Judges and their sentencing procedures: A sociological questionnaire study of factors important in the sentencing of defendants in the San Diego Superior Court. Unpublished manuscript, University of California at San Diego, 1975.

REFERENCES

Bermant, G., McGuire, M., McKinley, W., & Salo, C. The logic of simulation in jury research. *Criminal Justice and Behavior,* 1974, **1,** 224–233.

Brunswik, E. *Perception and the Representative Design of Psychological Experiments* (2nd Ed.). Berkeley, California: University of California Press, 1956.

Buckhout, R. Eyewitness testimony. *Scientific American,* 1974, **231** (6), 23–31.

Buckhout, R., Alper, A., Chern, S., Silverberg, G., & Slomovits, M. Determinants of eyewitness performance on a lineup. *Bulletin of the Psychonomic Society,* 1974, **4,** 191–192.

Campbell, D. T., & Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research.* Chicago, Illinois: Rand McNally, 1963.

Coombs, C. H. Thurstone's measurement of social values revisited forty years later. *Journal of Personality and Social Psychology,* 1967, **6,** 85–91.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist,* 1975, **30,** 116–127.

Davis, J. H., Bray, R. M., & Holt, R. W. The empirical study of decision processes in juries: A critical

review. In J. L. Tapp & S. J. Levine (Eds.), *Law, Justice, and the Individual in Society: Psychological and Legal Issues*. New York: Holt, Rinehart and Winston, 1977.

Doob, A. N. Evidence, procedure, and psychological research. In G. Bermant, C. Nemeth, & N. Vidmar (Eds.), *Psychology and the Law*. Lexington, Massachusetts: D. C. Heath, 1976.

Doob, A. N., & Kirshenbaum, H. M. Bias in police lineups—Partial remembering. *Journal of Police Science and Administration*, 1973, **1**, 287–293.

Ebbesen, E. B., & Konečni, V. J. Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, 1975, **32**, 805–821.

Ebbesen, E. B., & Konečni, V. J. On the external validity of decision-making research: What do we know about decisions in the real world. In T. S. Wallsten (Ed.), *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, New Jersey: Lawrence Erlbaum, in press (a).

Ebbesen, E. B., & Konečni, V. J. The setting of bail. In V. J. Konečni & E. B. Ebbesen (Eds.), *A Social-Psychological Analysis of the Criminal Justice System*. San Francisco: W. H. Freeman, in press (b).

Ebbesen, E. B., & Konečni, V. J. The process of sentencing adult felons: A causal analysis of judicial decisions. In B. D. Sales (Ed.), *Perspectives in Law and Psychology. Vol. 2: The Trial Process*. New York: Plenum, in press (c).

Efran, M. G. The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality*, 1974, **8**, 45–54.

Egan, D., Pittner, N., & Goldstein, A. G. Eyewitness identification: Photographs versus live models. *Law and Human Behavior*, 1977, **1**, 199–206.

Farmer, L. C., Williams, G. R., Lee, R. E., Cundick, B. P., Howell, R. J., & Rooker, C. K. Juror perceptions of trial testimony as a function of the method of presentation. In G. Bermant, C. Nemeth, & N. Vidmar (Eds.), *Psychology and the law*. Lexington, Massachusetts: D. C. Heath, 1976.

Gerbasi, K. C., Zuckerman, M., & Reis, H. T. Justice needs a new blindfold: A review of mock jury research. *Psychological Bulletin*, 1977, **84**, 323–345.

Goldfarb, R. L. *Ransom*. New York: Harper & Row, 1965.

Goodman, L. A. A general method for the analysis for surveys. *American Journal of Sociology*, 1972, **77**, 1036–1086.

Goodman, L. A. Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology*, 1973, **78**, 1135–1191.

Heise, D. R. *Causal Analysis*. New York: Wiley, 1975.

Izzett, R., & Leginski, W. Group discussion and the influence of defendant characteristics in a simulated jury setting. *Journal of Social Psychology*, 1974, **93**, 271–279.

Kalven, H., Jr., & Zeisel, H. *The American Jury*. Boston, Massachusetts: Little, Brown, 1966.

Kerr, N. L., Atkin, R., Stasser, G., Meek, D., Holt, K., and Davis, J. H. Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on judgments of mock jurors. *Journal of Personality and Social Psychology*, 1976, **34**, 282–294.

Kistler, R. Probation department's services unsatisfactory, judge says. *The Los Angeles Times*, March 12, 1975.

Konečni, V. J., & Ebbesen, E. B. A critique of theory and method in social-psychological approaches to legal issues. In B. D. Sales (Ed.), *Perspectives in Law and Psychology. Vol. 2: The Trial Process*. New York: Plenum, in press (a).

Konečni, V. J., & Ebbesen, E. B. Sentencing. In V. J. Konečni & E. B. Ebbesen (Eds.), *A Social-Psychological Analysis of the Criminal Justice System*. San Francisco: W. H. Freeman, in press (b).

Konečni, V. J., & Ebbesen, E. B. Social psychology and the law: Some considerations concerning the choice of research problems, settings and methodology. In V. J. Konečni & E. B. Ebbesen (Eds.), *A Social-Psychological Analysis of the Criminal Justice System*. San Francisco: W. H. Freeman, in press (c).

Konečni, V. J., Mulcahy, E. M., & Ebbesen, E. B. Prison or mental hospital: Factors affecting the processing of persons suspected of being "mentally disordered sex offenders." In P. D. Lipsitt & B. D. Sales (Eds.), *New Directions in Psycholegal Research*. New York: Van Nostrand Reinhold, in press.

Konon, M. Ratings could sway judge elections. *The San Diego Evening Tribune*, January 11, 1978.

Landy, D., & Aronson, E. The influence of the character of the criminal and his victim on the decisions of simulated jurors. *Journal of Experimental Social Psychology*, 1969, **5**, 141–152.

Lawson, R. G. Experimental research on the organization of persuasive arguments: An application to courtroom communications. *Law and the Social Order,* 1970, 597–608.

Leon, M., Oden, G. C., & Anderson, N. H. Functional measurement of social values. *Journal of Personality and Social Psychology,* 1973, **27,** 301–310.

Levine, S. J., & Tapp, J. L. The psychology of criminal identification: The gap from Wade to Kirby. *University of Pennsylvania Law Review,* 1973, **121,** 1079–1131.

Loftus, E. F. Leading questions and the eyewitness report. *Cognitive Psychology,* 1975, **7,** 560–572.

Loftus, E. F., Altman, D., & Geballe, R. Effects of questioning upon a witness's later recollection. *Journal of Police Science and Administration,* 1975, **3,** 162–165.

Mitchell, H. E., & Byrne, D. The defendant's dilemma: Effects of juror's attitudes and authoritarianism on judicial decisions. *Journal of Personality and Social Psychology,* 1973, **25,** 123–129.

Myers, D. G., & Kaplan, M. F. Group-induced polarization in simulated juries. *Personality and Social Psychology Bulletin,* 1976, **2,** 63–66.

Nemeth, C., & Sosis, R. H. A simulated jury study: Characteristics of the defendant and the jurors. *The Journal of Social Psychology,* 1973, **90,** 221–229.

Olson, C. L. Some apparent violations of the representativeness heuristic in human judgment. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, **2,** 599–608.

Persky, W., Sprague, M., & Lowe, B. How judges pass sentence. *The San Diego Evening Tribune,* May 22, 1975.

Sigall, H., & Ostrove, N. Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology,* 1975, **31,** 410–414.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Cognitive processes and societal risk taking. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and Social Behavior.* New Jersey: Lawrence Erlbaum, 1976.

Tapp, J. L. Psychology and the law: an overture. *Annual Review of Psychology,* 1976, **27,** 359–404.

Thibaut, J., & Walker, L. *Procedural Justice: A Psychological Analysis.* Hillsdale, New Jersey: Lawrence Erlbaum, 1975.

Vidmar, N. Effects of decision alternatives on the verdicts and social perceptions of simulated jurors. *Journal of Personality and Social Psychology,* 1972, **22,** 211–218.

Walker, L., LaTour, S., Lind, E. A., & Thibaut, J. Reactions of participants and observers to modes of adjudication. *Journal of Applied Social Psychology,* 1974, **4,** 295–310.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. *Unobtrusive Measures: Nonreactive Research in the Social Sciences.* Chicago: Rand McNally, 1966.

Wilson, D. W., & Donnerstein, E. Guilty or not guilty? A look at the simulated jury paradigm. *Journal of Applied Social Psychology,* 1977, **7,** 175–190.