

Spring 1992

Influence of Criminology on Criminal Law: Evaluating Arrests for Misdemeanor Domestic Violence

Lawrence W. Sherman

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/jclc>

 Part of the [Criminal Law Commons](#), [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

Recommended Citation

Lawrence W. Sherman, Influence of Criminology on Criminal Law: Evaluating Arrests for Misdemeanor Domestic Violence, 83 J. Crim. L. & Criminology 1 (1992-1993)

This Symposium is brought to you for free and open access by Northwestern University School of Law Scholarly Commons. It has been accepted for inclusion in Journal of Criminal Law and Criminology by an authorized editor of Northwestern University School of Law Scholarly Commons.

INTRODUCTION

THE INFLUENCE OF CRIMINOLOGY ON CRIMINAL LAW: EVALUATING ARRESTS FOR MISDEMEANOR DOMESTIC VIOLENCE

LAWRENCE W. SHERMAN*

I. INTRODUCTION

On September 19, 1986, the U.S. Food and Drug Administration approved the sale of a new drug, AZT, for the treatment of patients with AIDS.¹ The drug had been in development for several years, and had been tested in a randomized clinical trial involving 282 patients.² Using a standard medical research trial design, 137 patients were chosen by lottery to receive placebo (sugar) pills, and 145 were chosen to receive pills containing the real drug. In a followup period lasting up to six months, nineteen of the placebo group and one of the AZT group died. The experiment was halted ahead of schedule, although not nearly as soon as demanded by doctors and AIDS patients who wanted the drug to be approved before the research was completed. The FDA held fast to the policy it had adopted after the thalidomide disaster of the early 1960s, refusing to allow any drug to enter the market until it has been tested in a randomized clinical trial.³ The drug immediately became the

* Symposium Editor. Professor of Criminology, University of Maryland and President, Crime Control Institute, Washington, D.C. Ph.D., Yale University, 1976. Diploma in Criminology, Cambridge University, 1973. M.A., University of Chicago, 1970. B.A. Denison University, 1970. This research was supported in part by grant number 86IJCXKO43 from the National Institute of Justice to the Crime Control Institute. Points of view or opinions stated herein are those of the author and do not necessarily represent the official views of the U.S. Department of Justice.

¹ Philip J. Hilts, *Results of Preliminary Tests Prompt AIDS Drug's Release*, WASH. POST, Sept. 20, 1986, at A1.

² Margaret A. Fischl, et al., *The Efficacy of Azidothymidine (AZT) in the Treatment of Patients with AIDS and AIDS-Related Complex*, 317 N. ENG. J. MED. 185 (1987).

³ STUART J. POCOCK, *CLINICAL TRIALS: A PRACTICAL APPROACH* 26 (1983).

standard treatment for all AIDS patients.

On May 27, 1984, the U.S. National Institute of Justice announced the results of a randomized clinical trial of the use of arrest for misdemeanor domestic violence.⁴ The 314-case experiment, conducted with the Minneapolis Police Department, used a lottery method that assigned about one third of the probable cause suspects to be arrested, one third to be advised, and the rest to be sent away from the home on threat of arrest.⁵ Over a followup period lasting at least six months, about ten percent of the arrested suspects and about twenty percent of the suspects not arrested were officially detected to have committed one or more repeated domestic assaults.⁶ Citing these results, the Attorney General of the United States four months later issued a report recommending that arrest be made the standard treatment in cases of misdemeanor domestic assault.⁷ Within two years, "preferred" arrest became the most common urban police policy for those cases.⁸ By 1989, mandatory or preferred arrest policies were reported by eighty-four percent of urban police agencies.⁹ By late 1991, fifteen states and the District of Columbia had passed mandatory arrest statutes for cases in which there was probable cause to believe that misdemeanor domestic violence had occurred.¹⁰

The parallels between AZT for AIDS and arrest for domestic violence seem obvious. Both ailments are serious, afflicting millions of people and killing thousands each year. Victims of both ailments had strong political constituencies pressing to put the treatment into widespread use as soon as possible. Both treatments were carefully evaluated by scientists in collaboration with clinical practitioners, using the most advanced research design available for inferring cause and effect (the randomized, controlled experiment).¹¹ Both evaluations, by the basic sciences of biochemistry and criminology, respectively, had an apparently strong influence on the respective professional practices. Unfortunately, both treatments were shown

⁴ Lawrence W. Sherman & Ellen G. Cohn, *The Impact of Research on Legal Policy: The Minneapolis Domestic Violence Experiment*, 23 LAW & SOC. REV. 117, 121 (1989).

⁵ Lawrence W. Sherman & Richard A. Berk, *The Specific Deterrent Effects of Arrest For Domestic Assault*, 49 AM. SOC. REV. 261 (1984).

⁶ Richard A. Berk & Lawrence W. Sherman, *Police Responses to Family Violence Incidents: An Analysis of an Experimental Design With Incomplete Randomization*, 83 J. AM. STAT. ASSN. 70 (1988).

⁷ 1984 U.S. ATT'Y GEN. TASK FORCE ON FAMILY VIOLENCE REP. 17 (1984).

⁸ Sherman & Cohn, *supra* note 4, at 125.

⁹ LAWRENCE W. SHERMAN, *POLICING DOMESTIC VIOLENCE* ch. 5 (1992).

¹⁰ Joan Zorza, *The Criminal Law of Misdemeanor Domestic Violence, 1970-1990*, 83 J. CRIM. L. & CRIMINOLOGY 46 (1992).

¹¹ POCOCK, *supra* note 3.

only to reduce the suffering associated with the target problems, but not to cure those problems.

A final parallel is less obvious. Both experiments raised basic questions about the relationship between science and professional practice, generating intense controversy. Both of them led to clear recommendations to change the standard procedures governing scientific influences on policy-making. Both of them led to charges that scientists were morally insensitive to the interests and suffering of victims. The irony of this final parallel is that the two respective controversies went in opposite directions. Critics of the FDA charged that it used too *much* science, caution and delay before adoption of the new treatment.¹² Critics of the NIJ experiment charged that it used too *little* science, caution and delay before recommending adoption of the new treatment.¹³ The FDA responded by considering faster procedures for approving new drugs, possibly including approval without waiting for the results of controlled experiments. The NIJ responded by funding replications of the Minneapolis experiment in six new cities (Atlanta, Charlotte, Colorado Springs, Metro-Dade [Miami], Omaha, and Milwaukee), which are the subject of this symposium issue.

Explaining this irony is easy, but important. Over the preceding half-century, field experiments in biochemistry had become the basis for regulating the use of drugs as "dangerous commodities." The common wisdom in medicine was that it was unsafe to approve new drugs without large-scale controlled tests.¹⁴ The common wisdom in criminal law was just the opposite: penalties were a philosophical matter of just deserts, not an empirical question of effectiveness.¹⁵ Despite the pleas of some legal scholars to take questions of sanctioning efficacy more seriously,¹⁶ legislatures generally went about the process of making criminal law without much concern about its "safety." And while the FDA had never before encountered such a politically volatile disease, field experiments in criminology had never before achieved such influence over the

¹² Herbert Burkholz & Beverly Zakarian, *A Shot in the Arm for the F.D.A.*, N.Y. TIMES, July 21, 1991, § 6 (Magazine), at 6.

¹³ Richard Lempert, *From the Editor*, 18 LAW & SOC'Y REV. 505, 509-10 (1984); Richard Lempert, *Spouse Abuse: Ann Arbor Rushed Into Arrest Ordinance Without Studying Side Effects*, ANN ARBOR NEWS (June 21, 1987); Richard Lempert, *Humility is a Virtue*, 23 LAW & SOC'Y REV. 146 (1989) [hereinafter Lempert, *Humility is a Virtue*]. See also James W. Meeker & Arnold Binder, *Reforms as Experiments: The Impact of the 'Minneapolis Experiment' on Police Policy*, 17 J. POL. SCI. & ADMIN. 147 (1990); Arnold Binder & James W. Meeker, *Experiments as Reforms*, 16 J. CRIM. JUST. 347 (1988).

¹⁴ POCOCK, *supra* note 3; Burkholz & Zakarian, *supra* note 12.

¹⁵ See, e.g., ANDREW VON HIRSCH, *DOING JUSTICE* 45-6 (1976).

¹⁶ Norval Morris, *Impediments to Penal Reform*, 33 U. CHI. L. REV. 627-28 (1966).

course of the criminal law.¹⁷ As one observer put it, “the Minneapolis experiment probably reached the high water mark of research impact: in view of the publicity that the research received and the climate in which it was released, one can probably expect that social research will seldom have as much impact as this experiment did.”¹⁸

Just as some AIDS patients were shocked to discover that they could not have legal access to any new drug they chose even though they were dying, some criminologists were shocked to discover that criminology could have a clear impact on the shape of the criminal law. While that possibility is arguably one of the basic purposes of this journal, the lack of precedent for the dramatic impact of the Minneapolis experiment left criminology understandably unprepared to deal with the many moral and technical questions that impact raised. Is criminology a mature enough science—especially compared to biochemistry—to be guiding public policy at all? How much research is enough to support a policy recommendation? Should policy research results be publicized before they have been replicated? Should criminologists report only the average effect of a criminal law policy on a given sample, or must they also examine any systematic differences in the effects of the punishment on different kinds of people? Is crime control the only metric by which criminal law should be evaluated, or should other quality of life criteria—family unity, offender employability, children’s trauma at seeing parents arrested—also be considered? Should the burden of proof lie more heavily on criminologists recommending changes in current practice than on those supporting the status quo, even when no research is available to justify current practices?

In choosing the Minneapolis experiment and its replications as the subject of this symposium volume, the editors have two goals. One goal is to report the most comprehensive information available about the effects of arrest for this most pervasive problem of violence. With some 2,000 murders and over four million police encounters each year in the United States alone, the subject of domestic violence needs little justification for a journal of criminal law and its basic science.¹⁹ The other goal is equally important: to explore the concrete problems of using criminology to influence the criminal law, with the domestic violence arrest experiments as a case study. These problems become especially complex when a crucial experiment is replicated, and generates conflicting results in differ-

¹⁷ *But see* JOAN PETERSILIA, *THE INFLUENCE OF CRIMINAL JUSTICE RESEARCH* 94 (1987).

¹⁸ Sherman & Cohn, *supra* note 4, at 126.

¹⁹ SHERMAN, *supra* note 9.

ent cities. Interpreting those diverse findings poses a substantial challenge for the science of criminology; determining how the criminal law should respond to them is an even greater challenge.

This introduction attempts to guide the reader towards accomplishing both goals. Starting with a review of the rationale for conducting controlled experiments in criminal sanctions, it shows why police policy on domestic violence was ripe for such an experiment in 1980. The Minneapolis experiment is then described in some detail, including its policy recommendations *against* mandatory arrest laws and its relative influence in the passage of such laws. The five replication experiments are also described in some detail, so as to make possible an accounting for their diverse results. The foreword concludes with an assessment of the teachings of the past decade about both of this Volume's concerns: policing domestic violence more effectively, and using experimental criminology more wisely.

II. CONTROLLED EXPERIMENTS IN CRIMINAL SANCTIONS

The importance of controlled experiments in criminal law derives largely from our ignorance of the true nature of criminal deterrence. As Professor Norval Morris has observed, every criminal law system in the world (except Greenland's) has deterrence as its "primary and essential postulate."²⁰ As Sir Arthur Goodhart once observed, if punishment "cannot deter, then we might as well scrap the whole of our criminal law."²¹ Yet for most of human history, the evidence of the deterrent effects of criminal law has been little more than what Morris calls "a surfeit of unsubstantiated speculation."²²

In the past quarter century, substantial strides have been made toward filling the knowledge gap about the deterrent effects of criminal sanctions. A series of theoretical treatises²³ was followed by a prestigious National Academy of Sciences panel report on the methodological limitations of existing deterrence studies,²⁴ a series of survey studies published in this Journal²⁵ and elsewhere,²⁶ cross-sectional analyses of the relationship of criminal sanctions to crime

²⁰ MORRIS, *supra* note 16, at 631.

²¹ *Id.*

²² *Id.*

²³ JOHANNES ANDENAES, PUNISHMENT AND DETERRENCE (1974); FRANKLYN E. ZIMRING & GORDON J. HAWKINS, DETERRENCE (1973); JACK P. GIBBS, CRIME, PUNISHMENT AND DETERRENCE (1975).

²⁴ NATIONAL RESEARCH COUNCIL, DETERRENCE AND INCAPACITATION: ESTIMATING THE EFFECTS OF CRIMINAL SANCTIONS ON CRIME RATES (Alfred Blumstein et al. eds. 1978).

²⁵ Harold G. Grasmick & Donald E. Green, *Legal Punishment, Social Disapproval and Internalization as Inhibitors of Illegal Behavior*, 71 J. CRIM. L. & CRIMINOLOGY 325 (1980); Raymond Paternoster et al., *Estimating Perceptual Stability and Deterrent Effects: The Role of*

rates,²⁷ and quasi-experimental (before and after) evaluations of the effects of sudden changes in sanctions like capital punishment,²⁸ mandatory prison sentences,²⁹ and police crackdowns.³⁰ None of these research methods, however, has been able to resolve the lingering problems of distinguishing mere correlations from true causation. As a result, our knowledge of the deterrent or other consequences of criminal sanctions—including a possible increase in crime³¹—remains sketchy and uncertain.

Controlled experiments are fundamentally different from all other kinds of research. They are uniquely capable of eliminating alternative causes for observed effects, or plausible rival hypotheses also consistent with the evidence.³² In all other research designs the scientist must specify which rival theories must be tested and eliminated. In controlled experiments, even theories that the scientist never considered can usually be eliminated automatically. By making two groups comparable with respect to virtually all characteristics (within the limits of sampling error) *except* the factor under study (like AZT or arrest), a controlled experiment leaves very little doubt about inferring causation from any observed correlations—at least within the particular sample under study.³³ As we shall see, how-

Perceived Legal Punishment in the Inhibition of Criminal Involvement, 74 J. CRIM. L. & CRIMINOLOGY 270 (1983).

²⁶ Gary F. Jensen, *Crime Doesn't Pay: Correlates of a Shared Misunderstanding*, 17 SOC. PROB. 189 (1969); Matthew Silberman, *Toward a Theory of Criminal Deterrence*, 41 AM. SOC. REV. 442 (1976); Daniel S. Nagin & Raymond Paternoster, *The Preventive Effects of the Perceived Risk of Arrest: Testing an Expanded Conception of Deterrence*, 29 CRIMINOLOGY 561 (1991).

²⁷ Robert J. Sampson & Jacqueline Cohen, *Deterrent Effects of Police on Crime: A Replication and Theoretical Extension*, 22 LAW & SOC'Y REV. 163 (1988). In this author's opinion, this is by far the best of these analyses.

²⁸ David P. Phillips, *The Deterrent Effect of Capital Punishment: New Evidence On an Old Controversy* 86 AM. J. SOC. 139 (1980); Ruth D. Peterson & William C. Bailey, *Felony Murder and Capital Punishment: An Examination of the Deterrence Question*, 29 CRIMINOLOGY 367 (1991).

²⁹ Colin Loftin & David McDowall, *One With A Gun Gets You Two: Mandatory Sentencing and Firearms Violence in Detroit*, 455 ANNALS AM. ACAD. POL. & SOC. SCI. 150 (1981); Colin Loftin & David McDowall, *The Deterrent Effects of the Florida Felony Firearm Law*, 75 J. CRIM. L. & CRIMINOLOGY 250 (1984).

³⁰ H. LAURENCE ROSS, *DETERRING THE DRINKING DRIVER* (1981); Lawrence W. Sherman, *Police Crackdowns: Initial and Residual Deterrence*, in 12 CRIME AND JUSTICE: A REVIEW OF RESEARCH 1 (Michael H. Tonry & Norval Morris, eds., 1990).

³¹ A theory about which much exposition and some evidence are also accumulating. See RICHARD LEMERT, *SOCIAL PATHOLOGY* (1951); HOWARD S. BECKER, *OUTSIDERS* (1963); David Farrington, *The Effects of Public Labelling*, 17 BR. J. CRIMINOLOGY 112 (1977); Raymond Paternoster & LeeAnn Iovanni, *The Labelling Perspective and Delinquency: An Elaboration of The Theory and An Assessment of the Evidence*, 6 JUST. Q. 359 (1989).

³² THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION* 7-8 (1979).

³³ RONALD A. FISHER, *THE DESIGN OF EXPERIMENTS* (1971).

ever, generalizing from one sample to other populations is quite another matter.

The power to infer cause and effect has made controlled experimentation especially prominent in medical research, which has been blessed with far greater financial resources than criminological research. It has used those resources and the experimental method to accomplish the testing and approval of the Salk vaccine for polio, the use of penicillin to control infections, and the use of streptomycin to treat tuberculosis.³⁴ Of equal importance is the role that controlled experiments have played in putting a stop to harmful medical practices, like blindness-causing oxygen treatments for premature infants,³⁵ bleeding as a standard medical treatment,³⁶ removal of intestinal parts to cure epilepsy, removing teeth to cure a pitcher's sore arm, and extended bed rest after heart attacks.³⁷ It is largely because of the demonstrated dangers of treatments introduced without controlled tests that such tests became mandatory for new drugs in 1969, and why they continue to be used to expose established treatments that in fact make patients sicker.³⁸

The power to determine what works and what doesn't has not been lost on scholars of the criminal law. As early as 1959, Professors Zeisel, Kalven and Buchholz advanced the case for conducting controlled experiments in law.³⁹ Their argument overruled the standard objection to such research: the claim that the Constitution prohibits random assignment of punishment options as discriminatory. They pointed out that present decisionmaking in criminal justice is already so shot through with arbitrary and discriminatory practices that, if anything, controlled experiments tend to make decisionmaking *less* discriminatory. By creating equal probability for each subject to receive each of the alternative treatments, random assignment formulas in controlled experiments remove the influences of race, sex, class and demeanor. The only disparity left is between the experimental and control groups, a disparity which constitutes no discrimination against any class protected under the Fourteenth Amendment. While the disparity is arguably unjust for

³⁴ POCOCK *supra* note 3, ch. 1.

³⁵ William A. Silverman, *The Lessons of Retrolental Fibroplasia*, 236 (6) SCI. AM. 100 (1977).

³⁶ POCOCK, *supra* note 3.

³⁷ Howard H. Hiatt, *Will Your Next Hospital Stay Be Necessary?*, WALL ST. J., Nov. 18, 1986, § 1, at 32.

³⁸ Roy W. Beck et al., *A Randomized, Controlled Trial of Corticosteroids in the Treatment of Acute Optic Neuritis*, 326 NEW ENG. J. MED. 581 (1992); Sandra Blakeslee, *Treatment for Eye Disease May Worsen Condition, Study Says*, N.Y. TIMES, Feb. 27 1992, at A14.

³⁹ HANS ZEISEL ET AL., DELAY IN THE COURT 241 (1959).

those particular subjects, that cost is balanced against the benefits of the knowledge that a controlled experiment can produce.

This argument later led Chief Justice Burger's advisory committee on legal experimentation to endorse controlled experiments in criminal law.⁴⁰ The committee's logic rested heavily on the benefits that were presumed to result from completion of a well-designed, statistically powerful experiment. Their report created a special ethical burden to guard against methodologically weak criminology or research designs that were doomed from the start (from such problems as inadequate sample size) to be unable to accomplish their objectives; these would impose the cost of disparity with no countervailing benefit. But as a report of the Social Science Research Council observed, it is extremely difficult to conduct controlled experiments properly.⁴¹ The organizational complexity of controlled experiments far exceeds that of standard social science research methods, and fits in very uneasily with the social science research culture of universities.

It is not surprising, then, that controlled experiments in criminal law got off to a slow start. Lack of funding, ethical objections, and organizational demands combined to discourage would-be experimenters. These obstacles were not impossible to overcome, however, either in the U.S. or in England.⁴² In state correctional agencies like the California Youth Authority, and in private "think tanks" like the Vera Institute of Justice in New York, controlled experiments began to thrive in the early 1960s.⁴³ In perhaps the most famous of these, the Vera Institute developed and tested an alternative to money bail called "release on recognizance" (ROR). In a random assignment of persons with "community ties" to either money bail or ROR, virtually all those assigned to ROR appeared at court as scheduled.⁴⁴ The influence of this experiment was wide-

⁴⁰ FEDERAL JUDICIAL CENTER, ADVISORY COMMITTEE ON EXPERIMENTATION IN THE LAW, EXPERIMENTATION IN THE LAW (1981).

⁴¹ HENRY W. RIECKEN & ROBERT F. BORUCH, SOCIAL EXPERIMENTATION 153 (1974).

⁴² See David Farrington, *Randomized Experiments on Crime and Justice*, 4 CRIME & JUSTICE 257 (1983).

⁴³ See, e.g., LAMAR T. EMPEY & STEPHEN D. LUBECK, THE SILVERLAKE EXPERIMENT: TESTING DELINQUENCY THEORY AND COMMUNITY INTERACTION (1971); LAMAR T. EMPEY & MAYNARD L. ERICKSON, THE PROVO EXPERIMENT (1972); Marguerite Warren, *The Case for Differential Treatment of Delinquents*, 38 ANNALS AM. ACAD. POL. & SOC. SCI. 47 (1969); Ted B. Palmer, *California's Community Treatment Program for Delinquent Adolescents*, 8 J. RESCH. CRIME & DELINQ. 74 (1971); PAUL LERMAN, COMMUNITY TREATMENT AND SOCIAL CONTROL (1975).

⁴⁴ Charles Ares et al., *The Manhattan Bail Project: An Interim Report on the Use of Pre-Trial Parole*, 38 N.Y.U. L. REV. 67 (1963); Bernard Botein, *The Manhattan Bail Project: Its Impact in Criminology and the Criminal Law Process*, 68 TEX. L. REV. 319 (1965).

spread, resulting in the adoption of ROR by legislatures across the U.S. and in several other countries.

Even bolder experiments followed, although with less fame and influence. In the early 1970s, the California Department of Corrections conducted an experiment of a reduction of the time served by inmates in prison.⁴⁵ The experiment randomly assigned 1,135 prisoners to either "full" or "reduced" prison terms, which represented a six month reduction. The average full term was 37.9 months; the average reduced term was 31.3 months. The researchers structured the data analysis to show that there was no significant difference between the two groups in the rate of "unfavorable" parole outcomes. But Duke University Professor Philip Cook's reanalysis of the data found a significantly higher prevalence of recidivism measured by arrest after release of the reduced prison term group, compared to the full prison term group.⁴⁶

Given our general ignorance about the effects of various types and doses of criminal sanctions, such experiments clearly meet the threshold justifications for randomized experiments recommended by the Federal Judicial Center.⁴⁷ Those requirements provide that

1. The present practice must either need substantial improvement or be of doubtful effectiveness.
2. There must be significant uncertainty about the value of the proposed innovation.
3. There must be no other practical means to resolve uncertainties about the value of the proposed innovation.
4. The experiment must be seriously intended to inform a future choice between retaining the status quo or implementing the innovation.

A further ethical requirement suggested by Professor Morris was not included in the Federal Judicial Center's list. Writing in the 1960s, Morris suggested a principle of "less severity" than the status quo in any innovations to be tested by randomized experiments.⁴⁸ At that time, most of the policy options under debate, such as deinstitutionalization of juvenile delinquents, were pointed in the direction of less severity. By the 1980s, however, the tide was running the other way, with most interest groups demanding greater severity in criminal justice responses. The choice in that context

⁴⁵ JOHN E. BERECOCHEA ET AL., RES. DIVISION, CAL. DEPT. OF CORRECTIONS, RES. REP. NO. 62: TIME SERVED IN PRISON AND PAROLE OUTCOME: AN EXPERIMENTAL STUDY, REPORT NO. 2. (1981).

⁴⁶ Address by Professor Philip Cook, Carnegie Mellon University, School of Urban & Public Affairs, April 1988.

⁴⁷ FEDERAL JUDICIAL CENTER, *supra* note 40, at 11.

⁴⁸ Morris, *supra* note 16, at 648.

was to accede to greater severity without any evaluation, or to conduct a randomized experiment in which punishments of greater severity were randomly assigned. Nowhere, perhaps, was that choice more clearly framed than in the question of police responses to misdemeanor domestic violence.

III. POLICE RESPONSES TO DOMESTIC VIOLENCE

The historical custom of police in the U.S. was to make arrests only rarely in cases of misdemeanor domestic violence. In the later 1960s, this custom was reinforced by federal sponsorship of training in police mediation of domestic "disturbances," including those in which minor assaults had occurred. By the late 1970s, women's advocates used litigation and legislation to press for a policy innovation: much greater use of arrest. From the 1980s to the present, the innovation many have recommended is mandatory arrest (required by state law) whenever police have probable cause to believe that a domestic assault has occurred. This recommendation clearly constitutes a substantial increase in the severity of criminal sanctions for this particular offense. What is less understood is that it constitutes a departure from, rather than an equalization with, the level of enforcement severity for most other misdemeanors and many felonies.

A. UNDER-ENFORCEMENT: DOMESTIC AND OTHER VIOLENCE

As recently as 1967, the leading police professional organization, the International Association of Chiefs of Police, declared in its training manual that "in dealing with family disputes, the power of arrest should be exercised as a last resort."⁴⁹ This position was endorsed by the American Bar Association, whose 1973 *Standards for the Urban Police Function* said that police should "engage in the resolution of conflict such as that which occurs between husband and wife . . . in the highly populated sections of the large city, without reliance upon criminal assault or disorderly conduct statutes."⁵⁰ In 1977, police in three metropolitan areas were observed to take slightly longer to respond to domestic disturbance calls (4.65 minutes) than non-domestic disturbances (3.86 minutes). Police in these areas openly told observers it was the officers' policy (not the department's) "to proceed slowly in the hope that the problem would be resolved or that a disputant would have left before they

⁴⁹ INTERNATIONAL ASSOCIATION OF CHIEFS OF POLICE, TRAINING KEY 16: HANDLING DISTURBANCE CALLS (Gaithersburg, Md.: IACP 1967).

⁵⁰ ABA, PROJECT ON STANDARDS FOR CRIMINAL JUSTICE, STANDARDS FOR THE URBAN POLICE FUNCTION 12 (1973).

arrived.”⁵¹ The expression of such policies led many to conclude that male police officers practice discriminatory enforcement in such cases because they side with male offenders.

The evidence is far from clear, however, that police practiced *more* under-enforcement in domestic situations than in other cases in general, or in cases of interpersonal violence in particular. By the 1970s, the best evidence from observations of police work suggested that there was no less enforcement in domestic violence cases than in other cases of personal violence, although there was less enforcement in cases involving a male and a female than in cases involving two males. The evidence remains inconclusive largely because of imprecise data on the levels of injury involved in the different categories of cases.

The pattern of under-enforcement itself is clear. In 1966, Professors Albert J. Reiss, Jr. and Donald J. Black conducted the first multi-city study of police arrest decisions using systematic personal observations. This study of thousands of police-citizen encounters in Boston, Chicago and Washington found that, in cases where both victim and offender were present when police arrived, only forty-five percent of all felonies involving family members resulted in arrest; the proportion was fifty-five percent if the victim asked police to make an arrest. The arrest rate was about the same with respect to family misdemeanors overall (47%), although police were more likely to comply with victim requests for misdemeanor arrest (80%).⁵² In 1977, a similar study was conducted by Indiana University Professor Elinor Ostrom and her colleagues in twenty-four police departments in three metropolitan areas: Rochester (NY), Tampa-St. Petersburg (FL) and St. Louis (MO). One analysis of these data found that arrests were made in only twenty-two percent of all family assault cases (including those where one party had left the scene).⁵³ Another study of over 3,000 family violence records in an Ohio county in 1978 found that arrests were made in twelve percent of cases involving current or former spouses or lovers.⁵⁴

Many reasons may account for this low level of enforcement.⁵⁵ One is the common law in-presence requirement for misdemeanor arrests generally, which technically barred officers in many states

⁵¹ Nan Oppenlander, *Coping or Copping Out: Police Service Delivery in Domestic Disputes*, 20 CRIMINOLOGY 453 (1982).

⁵² DONALD J. BLACK, *THE MANNERS AND CUSTOMS OF THE POLICE* 94 (1980).

⁵³ Oppenlander, *supra* note 51 at 455.

⁵⁴ Edna Erez, *Intimacy, Violence and the Police*, 39 HUM. REL. 265 (1986).

⁵⁵ SHERMAN, *supra* note 9.

from making warrantless arrests unless they had witnessed the offense. Other reasons include the muddy practical distinctions between felony and misdemeanor assaults, possibly erroneous police perceptions of domestic violence situations as extremely dangerous to police,⁵⁶ frequent victim preferences against arrest,⁵⁷ and possibly even support by some police for the practice of spouse-beating. What is poorly understood by most policy activists in this area, however, is that police under-enforce the laws generally, for a very wide range of offenses.

The myth of full enforcement has been demolished by several careful field studies of police arrest behavior. These studies show clearly that full enforcement is not standard police practice. They also show mixed evidence on whether police take family matters less seriously than crimes among acquaintances or strangers. The 1966 study for the President's Commission on Law Enforcement and Administration of Justice⁵⁸ found that, in 176 encounters in which both a suspect and a complainant were present, and in which there was legally sufficient evidence for making an arrest, arrests were made in only fifty-eight percent of the reported felonies and forty-four percent of the reported misdemeanors⁵⁹. Similar evidence comes from Ostrom's 1977 study of policing in sixty neighborhoods in three major metropolitan areas. Of the 742 cases (of all kinds of offenses) in which police had a legal basis for an arrest of a single suspect present at the scene, arrests were made in forty-two percent of the felonies and fourteen percent of the misdemeanors⁶⁰. There was no victim present in over half the cases. The victim expressed a clear preference for arrest in only ten percent of these cases, yet police still failed to make arrests in over half (53%) of those. *Across all types*

⁵⁶ The literature now suggests that domestic violence is not very dangerous in terms of the danger of police officers dying, but domestic situations do appear to constitute a frequent source of injuries to police. Mona Margarita, *Killing the Police: Myths and Motives*, 452 ANNALS AM. ACAD. POL. & SOC. SCI. 63 (1980); David N. Konstantin, *Homicides of American Law Enforcement Officers, 1978-80*, 1 JUST. Q. 29, 42-43 (1984); JOEL GARNER & ELIZABETH CLEMMER, U.S. DEP'T OF JUSTICE, DANGER TO POLICE IN DOMESTIC DISTURBANCES—A NEW LOOK (1986); Craig Uchida et al., *Danger to Police During Domestic Encounters: Assaults on Baltimore County Police, 1984-86*, 2 CRIM. JUST. POL'Y REV. 357, 365; Rosemary Stanford & Bonney Lee Mowry, *Domestic Disturbance Danger Rate*, 17 J. POLICE SCI. & ADMIN. 244 (1990).

⁵⁷ SHERMAN, *supra* note 9.

⁵⁸ See Donald J. Black & Albert J. Reiss, *Patterns of Behavior in Police and Citizen Transactions*, in, STUDIES IN CRIME AND LAW ENFORCEMENT IN MAJOR METROPOLITAN AREAS (Field Surveys III, Vol. 2, 1967). See also ALBERT J. REISS, THE POLICE AND THE PUBLIC (1971); Black, *supra* note 52.

⁵⁹ BLACK, *supra* note 52, at 90.

⁶⁰ Douglas A. Smith & Christy A. Visher, *Street-Level Justice: Situational Determinants of Police Arrest Decisions*, 29 SOC. PROB. 167, 170 (1981).

of observed encounters in the summer of 1977, police in three metropolitan areas chose not to make an arrest in eighty-three percent of the cases where there was legal basis to do so. Even where the suspect and victim were strangers, police failed to make arrests in sixty-six percent of the cases (compared to eighty-three percent of cases where the parties were acquainted).

Observation studies have also witnessed police officers ignoring burglary, larceny, malicious destruction of property, drunk driving, hit and run accidents, and a broad range of other offenses.⁶¹ If there is a police agency that practices full enforcement anywhere in this country, researchers have yet to find it.

B. DOING SOMETHING: MEDIATION VS. ARREST

In the late 1960s, clinical psychologists like Professor Morton Bard recommended that police try to do more than walk away from domestic calls.⁶² They developed and trained officers in techniques of conflict mediation, a concept later found offensive by those who see assault as crime, not "conflict." These techniques taught police to be more on-the-spot marriage counselors than assessors of possible law breaking. The techniques quite sensibly included separation of the man and woman from each other and, if possible, other members of the household. Each party would then be able to give the officer her or his version of what happened without fear of being contradicted by the other party, leading to more shouting or worse. After hearing the two versions, police were supposed to consult with each other to discuss alternative actions. A preferred method was to get the two parties to calm down, sit down, and rationally discuss what would happen next. If that was not possible, officers would often advise one of the parties to leave for a cooling off period. Another option Bard stressed was referral to counseling or other social service agencies. Consistent with past practice, arrest was reserved for cases of serious injury or assaults on police.

This training resulted in some cities in a decrease in arrests. Some training explicitly made this a goal. This 1977 observation of a police encounter shows how the mediation training was carried out in many cases:

The officer received a call for family trouble. A woman met us at the door. She was crying and very upset. She had some bruises on her face and her lip was swollen and bleeding. She said that her husband

⁶¹ BLACK, *supra* note 52, at 185; MICHAEL K. BROWN, *WORKING THE STREET* 182-220 (1981).

⁶² MORTON BARD, *NAT'L INST. L. ENFORCEMENT & CRIM. JUST., TRAINING POLICE AS SPECIALISTS IN FAMILY CRISIS INTERVENTION* (1970).

had hit her, that she was not going to take it any more, and that she wanted him arrested. The officer had her sign a complaint form. At this point the man upstairs began yelling and cursing at the woman. It turned out that the man and the woman are not married but have been living together for some time and have a small baby. The woman thought that they were married, but the man is in the process of getting a divorce from his wife. The officer went upstairs [with the woman] to try to talk to the man who was very angry and yelled at the woman and the officer. The officer threatened him with arrest. Finally, the officer shut the door to keep out the woman. She got very upset and felt that the officer was taking the side of the man. The back-up officer arrived and found the woman in the corner of the small room, still crying. She told the second officer that she thought she was legally married to the man. The officer told her that she could not be legally married to him, since he was still legally married to his wife. When the man came downstairs with the first officer, he said he was going to his mother's house. The officer asked if everything would be okay. The man said yes. The woman remained sitting in the corner crying. The officers left.⁶³

In this example, mediation was substituted for a legally valid arrest. It also consisted primarily of a one-on-one discussion with the suspect. As Professor Oppenlander's analysis of the 1977 data points out, police discussions with one party are more frequent than those that engage both parties simultaneously. Yet it is hard to call such methods "mediation" if they do not involve consultation with both parties about the solutions being reached, regardless of whether there are face-to-face discussions.

In the majority of cases in which no crime has been committed, mediation techniques might make a great deal of sense in the short run. They are clearly focused on the immediate "crisis," and not on any long-term pattern or future risks of violence in the relationship. The implicit objective is to minimize the risk of harm while police are present at the scene, rather than to prevent a recurrence of violence. Since police mishandling of interpersonal relationships in the encounter can arguably escalate the risk of violence rather than defuse it, their mediation skills are important targets for training.⁶⁴ Whether accomplishing the objective of minimizing violence at the police encounter will reduce domestic violence overall, however, is debatable. Unfortunately, no rigorous evaluation of mediation training was ever done that adequately addressed the key question of violence reduction.

The mediation approach also hypothesized that arrest should be avoided, even when a misdemeanor had been committed, be-

⁶³ Oppenlander, *supra* note 51, at 460-61.

⁶⁴ BARD, *supra* note 62.

cause of the danger that arrest might backfire. This reflected a widespread belief among police at that time. As one observer summarized the views of police he interviewed:

An arrest in a family squabble does not resolve the underlying problems, and it may simply aggravate matters, especially if the husband is arrested. He will eventually get out of jail, and he may return home, angrier than before, to pick up where he left off.⁶⁵

In the 1970s, however, women's advocates clearly rejected that hypothesis. As Joan Zorza's article in the present volume amply documents, the women's movement adopted a full enforcement position. This position implicitly hypothesizes that more arrests (and prosecution, and sentencing) would help to reduce male violence in the home against women. The first line of attack on police under-enforcement was through litigation. But litigation alone could not address the underlying statutory problem of the in-presence requirement for warrantless misdemeanor arrests. Accordingly, legislation was sought and obtained in several states in the late 1970s granting a specific exception to the in-presence requirement for domestic violence misdemeanors. One of the states in which this occurred was Minnesota, in 1978. Predictably, however, Minnesota police made little use of the expanded arrest powers that the legislature had granted them.⁶⁶

In this context, a controlled experiment was clearly justified under Federal Judicial Center guidelines.⁶⁷ Practitioners had great uncertainty about a proposed innovation they were being asked to adopt, an innovation in practice if not in black letter law. Two competing hypotheses about the effects of arrest, at least in specific cases, could be set against each other on the criterion measure of repeat domestic violence. While a controlled experiment in arrest would not be able to address the *general* deterrent effects of arrest, it could examine the *specific* deterrent effects in which individual victims would have the greatest personal stake. Fortunately, Minneapolis appointed a new police chief in 1980, Anthony V. Bouza, who was willing to undertake the risk of conducting the first random assignment experiment in field arrests.

⁶⁵ BROWN, *supra* note 61, at 204.

⁶⁶ Joan Potter, *The Police and the Battered Wife: The Search for Understanding*, 1 POLICE MAG. 40.

⁶⁷ FEDERAL JUDICIAL CENTER, *supra* note 40.

IV. THE MINNEAPOLIS EXPERIMENT

A. DESIGN AND IMPLEMENTATION

In order to find which police approach was most effective in deterring future domestic violence by the same offender against the same victim, the Police Foundation and the Minneapolis Police Department agreed to conduct a controlled experiment with the support of the National Institute of Justice.⁶⁸ The design of the experiment called for a lottery selection of three treatments for all suspects legally eligible for arrest under the 1978 Minnesota warrantless arrest statute. The lottery selection, as noted above, minimized the pre-existing differences among the three groups of suspects, and helped eliminate any other possible cause of observed differences in repeat violence rates besides the three treatments themselves. The treatments included the following:

1. arrest (with at least one night in jail)
2. sending the suspect away from the scene of the assault for eight hours (or arresting the suspects if they refused)
3. giving the couple some form of advice, which could include mediation at an officer's discretion.⁶⁹

The criterion for comparing the relative success of the three treatments was the frequency and seriousness of any future domestic violence over the next six months.

The experiment involved only simple (misdemeanor) domestic assaults, where both the suspect and the victim were present when the police arrived. Thus, the experiment included only those cases in which police were empowered, but not required, to make arrests. The police officer needed to have probable cause to believe that a cohabitant or spouse had assaulted the victim within the past four hours. Police did not need to witness the assault. Cases of life-threatening or severe injury, usually labeled as felonies (aggravated assault), were excluded from the design. So were cases in which the victim demanded an arrest, cases in which a court order of protection was in effect, cases in which the suspect assaulted a police officer, or any case in which a police officer believed an arrest was necessary due to an imminent threat of harm to the victim.

⁶⁸ Sherman & Berk, *supra* note 5.

⁶⁹ After meeting with the experimental officers for two days, the principal investigators decided that no greater homogeneity of a "mediation" treatment was feasible in the circumstances. This weak test of "mediation" was arguably typical for police departments at that time. See OPPENLANDER, *supra* note 51. On the other hand, it created a bias against the success of a non-arrest compliance treatment. See Albert J. Reiss, *Consequences of Compliance and Deterrence Models of Law Enforcement For the Exercise of Police Discretion*, 47 L. & CONT. PROB. 106 (1984).

The design called for each participating officer—all of whom were volunteers—to carry a pad of report forms, color coded for the three different police responses. Each time the officers encountered a situation that fit the experiment's criteria, they were to take whatever action was indicated by the top report form on the pad. The forms were numbered and arranged for each officer in an order determined by the lottery. The consistency of the lottery assignment was to be monitored by research staff observers riding on patrol for a sample of evenings. After a police action was taken at the scene of a domestic violence incident, the officer was to fill out a brief report and give it to the research staff for follow-up. As a further check on the lottery process, the staff logged in the reports in the order in which they were received and made sure that the sequence corresponded to the original assignment of responses.

The experiment employed two key measures of repeat violence: official police records and victim interviews. A predominantly minority, female research staff was employed to contact the victims for a detailed, face-to-face interview, to be followed by telephone follow-up-interviews every two weeks for twenty-four weeks. The interviews were designed primarily to measure the frequency and seriousness of victimizations caused by a suspect after police intervention. The research staff also collected criminal justice reports that mentioned suspect's names during the six-month follow-up period.

As is common in field experiments, the actual research process in Minneapolis differed somewhat from the original plan. None of these differences, however, seriously threatened the experiment's validity. There is little doubt that many of the officers occasionally failed to follow fully the experimental design. Some of the failures were due to forgetfulness, such as when officers left report pads at home or at the police station. Other failures stemmed from confusion over whether the experiment applied in certain situations. Whether any officer intentionally subverted the design is unclear. The plan to monitor the lottery process with ride-along observers broke down because of the unexpectedly low frequency of cases meeting the experimental criteria, a problem that also affected the replications. Thus, the possibility existed that police officers finding the upcoming experimental treatment unpalatable may have occasionally decided to ignore the experiment. They may have chosen, in effect, to exclude certain cases in violation of the experimental design. Such action would have biased the selection of the experiment's sample of cases, but not the results of the experiment among the cases included.

Moreover, the plans assumed that there would be legitimate reasons why the three treatments were not always delivered as designed. Ninety-nine percent of the suspects targeted for arrest actually were arrested; seventy-eight percent of those scheduled to receive advice did; and seventy-three percent of those who were to be sent out of the residence for eight hours actually were sent (most of those who were not were arrested, as planned). One explanation for this pattern, consistent with experimental guidelines, is that mediating and sending off were more difficult ways for police to control a situation.

This pattern could have biased estimates of the relative effectiveness of arrest by removing uncooperative and difficult offenders from mediation and separation treatments. Any deterrent effect of arrest could be underestimated and, in the extreme, arrest could be shown to increase the chance of repeat violence. In effect, the arrest group could have received too many "bad risks" relative to the other treatments. Fortunately, Professor Berk's statistical analysis of this process shows that the delivered treatments conformed very closely to the experimental design, with no problems of bias.⁷⁰

More substantial problems arose with the interviews of victims. The majority of the victims could not be followed up for the full six months, and less than two-thirds even granted an initial interview. Many of the victims simply could not be found, either for the initial interview or for the follow-ups. They had left town, moved somewhere else, or refused to answer the phone or doorbell. The research staff made up to twenty attempts to contact these victims and often employed investigative techniques (asking friends and neighbors) to find them. Sometimes these methods worked, only to have the victim give an outright refusal, or break one or more appointments to meet the interviewer at a "safe" location for the interview. Fortunately, the experimental treatment assigned to the offender did not affect the victim's decision to grant initial interviews. Statistical tests showed the victims' willingness to give interviews did not depend upon what police did, the race of the victim, or the race of the offender.

Despite these limitations, the experiment succeeded in producing an experimental sample of 314 cases with complete official outcome measures and an apparently unbiased sample of responses from the victims in those cases.

⁷⁰ Sherman & Berk, *supra* note 5.

B. RESULTS

Consistent with the kinds of cases coming to police attention in most big cities, the sample contained a disproportionate number of unmarried couples with lower than average educational levels, minority couples, and couples who were very likely to have had prior violent incidents with police intervention. The sixty percent unemployment rate among the experiments's suspects was strikingly high. The suspects' fifty-nine percent prior arrest rate was also strikingly high, suggesting (with the eighty percent prior domestic assault rate) that the suspects generally were experienced law-breakers who were accustomed to police interventions. But with the exception of the heavy representation of Native-Americans due to Minneapolis' proximity to many Indian reservations, many of the sample's characteristics were probably close to those of domestic violence cases coming to police attention in other large U.S. cities.

The results, based on both official records and victim interviews, showed that arrest produced the lowest prevalence of repeat violence of any of the three treatments. The results taken from the police records on subsequent violence showed ten percent of the *arrested* suspects with at least one repeat incident, nineteen percent of the *advised* suspects, and twenty-four percent of the suspects *sent* away for eight hours. The arrest treatment was clearly an improvement over sending the suspect away, which produced two and a half times as many repeat incidents as arrest. The advise treatment was statistically not distinguishable from the other two police actions.

The victim interviews showed a somewhat different picture. According to the victims' reports of repeat violence, nineteen percent of the suspects in the *arrest* group, thirty-three percent of those in the *send* group, and thirty-seven percent of those in the *advise* group committed at least one repeat attack on the victims (defined as including assaults, threats or property damage). This ranking varies from the official measure results by reversing the send and advise groups. In this measure, sending the suspect away produced results that were not statistically distinguishable from the results of the other two actions. It is not clear why the order of the three levels of repeat violence was different for these two ways of measuring the violence. But it is clear that arrest worked best by either measure, at least in comparison to the two most common police alternatives as they were currently practiced by patrol officers. This does not mean that arrest was be superior to more powerful alternatives; it only means that it worked better than the alternatives then in use.

C. LIMITATIONS

The experiment generally won high praise for the quality of the research design. It has been described as "a landmark study,"⁷¹ and as "arguably the best field experiment on a criminal justice policy problem to date."⁷² Nonetheless, the study had a number of limitations, many of which were pointed out by the authors, and some of which were later identified by other commentators.⁷³

Perhaps the most important limitation was that the study was unable to say *why* arrest had the lowest rate of repeat violence. While the authors presented it as a specific deterrent effect, it may also have been a displacement effect, meaning that the suspects moved on to abuse new victims.⁷⁴ The low completion rate of victim interviews also made it difficult to estimate the rate of breakup in relationships across treatment groups, which may also have reduced the risk of repeat violence. It was, however, large enough to refute the hypothesis that arrested suspects' victims were intimidated into remaining silent about new violence, since the interview completion rates did not vary by treatment group.

The study's summary report for practitioners identified several problems involved in generalizing from the study's results to other cities.⁷⁵ One was the sample size, which was too small to yield information regarding the possibility of different effects among different kinds of people. Another was the automatic night in jail attendant to domestic violence arrest in Minneapolis, which is not true of all cities; without it, a deterrent effect might not be achieved. A third was the unusual combination of sixteen percent Native-American and thirty-six percent African-American suspects in the sample, which might not allow generalization to cities with, say, large proportions of Hispanics. A fourth was the fact that interviewers attempted to contact all the victims, which attention might have artificially enhanced the effects of the police actions.

Other critics have attacked the methods of analysis used in the original report,⁷⁶ although a subsequent analysis addressed the purely statistical issues and found similar results no matter how the

⁷¹ Delbert S. Elliott, *Criminal Justice Procedures in Family Violence Crimes*, in FAMILY VIOLENCE: 11 CRIME & JUSTICE 458 (Lloyd Ohlin & Michael Tonry eds., 1989).

⁷² Professor Richard Lempert, *quoted in* Sherman & Cohn, *supra* note 4, at 118.

⁷³ For a full review of the limitations, see SHERMAN, *supra* note 9, ch. 4.

⁷⁴ Albert J. Reiss, *Some Failures in Designing Data Collection That Distort Results*, in COLLECTING EVALUATION DATA: PROBLEMS AND SOLUTIONS 161 (Leigh Burstein et al., eds., 1985).

⁷⁵ Lawrence Sherman & Richard Berk, *The Minneapolis Domestic Violence Experiment*, 1 POLICE FOUND. REP. 7 (1984).

⁷⁶ Binder & Meeker, *supra* note 13.

data were analyzed.⁷⁷ The major debate about the limitations of the experiment concerned its appropriateness for serving as the basis of any national policy recommendations. One commentary, for example, called the experiment a mere "pilot study," and described using its results for national policy-making as "foolishness bordering on irresponsibility."⁷⁸ Another suggested that the limitations implied a need for "some caution in rushing to a policy recommendation on the strength of these findings alone."⁷⁹ These comments raise the distinct questions of what the authors actually recommended on the basis of the results, and what others recommended on the basis of the findings that the authors actively helped to publicize.

D. RECOMMENDATIONS

The authors actually recommended three policies.⁸⁰ One was that the estimated twenty-two states then barring warrantless arrest for misdemeanor domestic violence not committed in the officer's presence change their laws to allow such arrests.⁸¹ This recommendation seemed to be amply supported by the research, and was the least controversial among both researchers leaning against drawing strong inferences from the study and women's advocates leaning in favor of drawing strong inferences. Professor Lempert, the first to call for caution in making innovations based on the results, concurred that "if I were a police chief I would change a 'do not arrest' policy because of the study's results, although I would not mandate arrest."⁸²

The Minneapolis experiment's authors concurred in cautioning against *mandatory* arrest, although they did write that, "on the basis of this study alone, police should probably employ arrest in most cases of minor domestic violence."⁸³ They went on to cite the Minneapolis police department's policy (which I drafted) as the kind that "did not make arrest 100 percent mandatory. The policy did, however, require officers to file a written report explaining why they failed to make an arrest when it was legally possible to do so."⁸⁴ The authors' opposition to mandatory arrest statutes stemmed primarily from the study's small sample size. This meant that, "be-

⁷⁷ Berk & Sherman, *supra* note 6.

⁷⁸ Binder & Meeker, *supra* note 13, at 354.

⁷⁹ Elliott, *supra* note 71, at 454.

⁸⁰ Sherman & Berk, *supra* note 75.

⁸¹ Lisa G. Lerman & Franci Livingston, *State Legislation on Domestic Violence*, 4 RESPONSE 1 (1983).

⁸² Lempert, *Humility is a Virtue*, *supra* note 13, at 158.

⁸³ Sherman & Berk, *supra* note 75.

⁸⁴ *Id.*

cause of the relatively small numbers of suspects in each subcategory (age, race, employment status, criminal history, etc.), it is possible that this experiment failed to discover that for some kinds of people, arrest may only make matters worse. Until subsequent research addresses the issue more thoroughly, it would be premature for state legislatures to pass laws requiring arrests in *all* misdemeanor domestic assaults."⁸⁵

In making this recommendation, the authors ran afoul of both academic critics and women's advocates. The academic critics argued, in the spirit of the Federal Judicial Center guidelines, that the experiment had too many limitations to justify a national policy innovation in the direction of greater severity.⁸⁶ Women's advocates, in contrast, generally argued that the research showed the value of arrest, and that police would not make more arrests without a state statute requiring it.⁸⁷ Having offended critics on both ends of the spectrum, the authors may claim to have taken the middle road. And judging by the standards for the approval of new drugs by the FDA, the case can be made that the Minneapolis experiment would have been sufficient evidence for the FDA to approve a "preferred arrest" policy of the kind the authors recommended.⁸⁸

The FDA analogy breaks down, however, with the authors' third recommendation: that the Minneapolis experiment be replicated in other cities. It appears to be unusual for randomized clinical trials to continue after the FDA has authorized a new drug, in part because it would mean withholding a treatment of proven effectiveness from a patient in need.⁸⁹ This was the position taken by some opponents of the Minneapolis replication program, who argued for leaving well enough alone.⁹⁰ The authors again took a middle ground, suggesting that the research results should be used provisionally. The authors stressed that more research was needed to insure that the findings would apply to other cities and that arrest would not create more violence among any particular subgroups. As the results reported in this Symposium issue demonstrate, both concerns were valid.

⁸⁵ *Id.*

⁸⁶ Binder & Meeker, *supra* note 13; Lempert, *supra* note 13.

⁸⁷ See Zorza, *supra* note 10.

⁸⁸ Sherman & Cohn, *supra* note 4, at 136.

⁸⁹ *Id.*

⁹⁰ Personal communication from Deputy Inspector Dean J. Collins, Milwaukee Police Department, concerning lobbying against Milwaukee Common Council approval of the Milwaukee domestic violence experiment, August, 1986. The Common Council subsequently gave the experiment unanimous approval.

E. THE EXPERIMENT'S INFLUENCE

It is not clear exactly how much influence the experiment had in changing police policies and state laws, nor is it clear that those changes created substantial changes in police practice. What is clear is that the policies and laws governing police responses to domestic violence underwent a radical change in the years following the extensive publicity reporting the experiment's results.

In 1984, on the eve of the announcement of the final Minneapolis results, a national telephone survey of a sample of police departments serving cities of over 100,000 found that only ten percent of them encouraged their police officers to make arrests in misdemeanor domestic violence cases. By 1986, a repeat survey found that forty-six percent of a (slightly different) sample of those cities encouraged arrest in those circumstances. There is some evidence that cities were more likely to have made that change if the respondents could properly identify the results of the Minneapolis experiment, but some might read those results as showing only a modest direct influence of the experiment on policy.⁹¹

What may have had greater influence on the policy change was the highly publicized \$2.5 million jury verdict against the Torrington, Connecticut Police Department for repeatedly failing to make arrests of an abusive husband who ultimately caused serious injury to his wife, Tracey Thurman.⁹² The threat of civil liability is often mentioned when police discuss this issue, even among police officers who have never heard of the Minneapolis experiment. The indirect influence of the experiment, however, might have come from Professor James Fyfe, who cited its findings as an expert witness testifying on behalf of plaintiff.⁹³

Whatever the relative contributions of the experiment, the Duluth mandatory arrest demonstration project,⁹⁴ the Attorney General's Task Force on Family Violence,⁹⁵ documentary and dramatized television programs, and other attention paid to the issue in the mid-1980s, police policies continued to change. By 1989, a sample of big city police agencies found that eighty-four percent had adopted at least a preferred arrest policy, and seventy-six per-

⁹¹ Sherman & Cohn, *supra* note 4, at 125.

⁹² A.P., *Officers Must Pay \$2.3 Million to Wife Maimed by Husband*, N.Y. TIMES, June 26, 1985, B6.

⁹³ SHERMAN, *supra* note 9, at ch. 5.

⁹⁴ Jan Hoffman, *When Men Hit Women*, N.Y. TIMES, Feb. 16, 1992, at § 6 (mag.) 23. This article about mandatory arrest nationally attributes great influence to the (unevaluated) Duluth project, and makes no mention of the Minneapolis experiment.

⁹⁵ *Supra* note 7.

cent had adopted mandatory arrest policies.⁹⁶ A sample of rural and urban police agencies in 1988 found that the Minneapolis experiment was ranked last among eight possible influences on police policies for domestic violence, but the authors of that survey argued that the experiment had indirect influence through lobbying on state legislatures.⁹⁷ Whatever the influence on the state legislatures, all but one of the fifteen states with mandatory arrest statutes adopted them in the period after the Minneapolis experiment was released.⁹⁸

In recommending against such mandatory arrest laws and policies, the authors of the Minneapolis experiment warned that such laws would merely "invite circumvention" by police officers resistant to policy constraints.⁹⁹ According to the available research on the implementation of such policies, that is generally what happened. Professor Ferraro's study of the Phoenix Police Department's field practices after the adoption of a mandatory arrest policy found that police made arrests in only forty-three percent of the cases where there was probable cause and the offender was present.¹⁰⁰ Professor Balos, in her review of police compliance with the Minnesota mandatory arrest statute for violations of court orders of protection (about which the Minneapolis experiment was silent), found that police in Hennepin County, which includes Minneapolis, made arrests in only twenty-two percent of the cases where arrest was required by state law.¹⁰¹ Only Milwaukee, of all cities studied to date, appears to have achieved a high level of compliance with a mandatory arrest policy.¹⁰² Yet despite the resistance to full implementation of these policies, the national arrest rate for all simple assaults (most of which are probably domestic) rose by seventy percent from 1985 to 1989.¹⁰³ In states where such laws were passed, moreover, it became legally impossible to replicate the Minneapolis experiment with a control group of non-arrested suspects, regardless of the *de facto* circumvention.

Fortunately, misdemeanor domestic violence arrests remained

⁹⁶ J. David Hirschel & Ira W. Hutchison III, *Police-Preferred Arrest Policies*, in WOMAN BATTERING: POLICY RESPONSES 59 (Michael Steinman, ed., 1991).

⁹⁷ Meeker & Binder, *supra* note 13, at 151.

⁹⁸ NATIONAL CENTER ON WOMEN AND FAMILY LAW, MANDATORY ARREST SUMMARY CHART (1991).

⁹⁹ Sherman & Berk, *supra* note 5, at 270.

¹⁰⁰ Kathleen J. Ferraro, *Policing Woman Battering*, 36 SOC. PROB. 61, 63-64 (1989).

¹⁰¹ Beverly Balos & Katie Trotzky, *Enforcement of the Domestic Abuse Act in Minnesota: A Preliminary Study*, 6 LAW & INEQUALITY 83, 103 (1988).

¹⁰² SHERMAN, *supra* note 9, chapter 5.

¹⁰³ Computations from FBI data reported in Sherman, *supra*, note 9, Figure 5.1.

legally discretionary in enough states to allow the National Institute of Justice to sponsor a series of replications of the experiment—perhaps the clearest and most direct influence of the Minneapolis experiment on public policy.

V. THE REPLICATIONS

The five completed replications of the Minneapolis experiment have produced two key findings. One finding is that, in cases of arrest for probable cause, misdemeanor domestic assault has different effects in different cities. Specifically, the results in three cities (including Minneapolis) show evidence of a deterrent effect, while those in three others show evidence of increased violence overall. The second finding is that such arrests have different effects on different kinds of people, with a consistent variation depending on the employment status of the suspect. Employed suspects tend to be deterred by arrest, while unemployed suspects tend to become more frequently violent following arrest. These findings are drawn from the six experiments funded to replicate the Minneapolis experiment, all but one of which (Atlanta) has produced research reports as of this writing.¹⁰⁴

Understanding these results requires some appreciation of the differences among the experiments in their research designs and their implementation, as well as some of the possible reasons for the findings. It also requires special attention to the details of the Metro-Dade (Miami) experiment, which alone among the completed experiments is not reported by its authors in this volume.¹⁰⁵ Finally, the question of what influence the replications have had on the criminal law to date reveals important differences between them and the Minneapolis experiment.

A. RESEARCH DESIGNS

None of the replications was an exact reproduction of the Minneapolis research design. Rather than slavishly following the model of the physical sciences, the new experiments all improved upon the Minneapolis design in basic ways, and most of them tested different

¹⁰⁴ See articles in this Symposium issue, as well as Franklyn Dunford et al., *The Role of Arrest in Domestic Assault: The Omaha Police Experiment*, 28 CRIMINOLOGY 183 (1990); J. DAVID HIRSCHL, ET AL., CHARLOTTE SPOUSE ASSAULT REPLICATION PROJECT: FINAL REPORT, (NAT'L INST. OF JUST., 1990); Anthony M. Pate et al., *Metro-Dade Spouse Abuse Replication Project Draft Final Report* (NAT'L INST. OF JUST., 1991); Lawrence W. Sherman et al., *From Initial Deterrence to Long-Term Escalation: Short-Custody Arrest for Poverty Ghetto Domestic Violence*, 29 CRIMINOLOGY 821 (1991).

¹⁰⁵ This unfortunate lacuna is not due to the lack of an invitation.

combinations of treatments against arrest. These changes in research design made the results arguably more powerful and interesting. The cost of those benefits, however, is some complication in our ability to make direct comparisons and draw conclusions across the studies.

The most basic improvement was the separation of police screening of cases for eligibility in the experiment from the application of the randomized treatment. That is, all five replications assigned the treatments after the police officer declared the cases eligible, and the officers had no idea of what the treatment would be at the time they made the eligibility decision. This was a potential problem inherent in the use of the color-coded report sheets that officers were issued in Minneapolis, due to the possible bias it could introduce into the officers' decision to include a case in the experiment. A "nice" suspect, for example, might have been excluded if the officer knew that arrest was the next treatment in his randomized sequence. By requiring officers to call headquarters to find out the treatment, the design greatly reduced the potential for such bias in sample selection.

Another improvement benefiting all but Omaha was a substantial increase in funding to allow much larger sample sizes. The larger sample sizes had several benefits. One was an increased number of treatments that could be tested in some experiments (Miami and Colorado Springs). Another was greater statistical power for examining different effects of arrest on different types of suspects. A third was greater power overall for detecting effects of arrest that are less substantial than the large differences found in Minneapolis.

The treatment innovations responded well to Professor Reiss's comment about comparing arrest to rather weak alternatives.¹⁰⁶ The Colorado Springs experiment included immediate counseling at police headquarters, as well as police issuance at the scene of an emergency protection order legally ejecting the offender from the premises. The Metro-Dade experiment compared arrest with and without followup police counseling some days later to no arrest with and without counseling. The Charlotte experiment compared arrest to a ticket-style citation at the scene, which did not require the suspect to leave. Milwaukee pursued the question of jail time associated with arrest, randomly assigning three-hour and twelve-hour arrests to see what difference the variations in jail time across police agencies might make in the generalizability of the findings. Only

¹⁰⁶ See Reiss, *supra* note 69.

Omaha repeated the three Minneapolis alternatives, providing the closest replication of any of the five.

Eligibility criteria varied somewhat, sometimes unavoidably. In Metro-Dade, for example, warrantless arrest was illegal for unmarried couples, even if they were cohabiting, until near the end of the experiment. Thus, their sample has the highest proportion of married couples (79%) of any of the experiments, although it was closely followed by the sixty-nine percent married couples in the Colorado Springs sample—the only two experiments with a majority of married couples.¹⁰⁷ The inclusion of non-spouse-like cases, such as homosexual lovers and other cohabitants, an issue raised by Ms. Zorza in this symposium, represented a trivial difference across the cities. While Charlotte, Colorado Springs and Metro-Dade excluded such cases, they constituted only eight percent of cases in Milwaukee, ten percent of cases in Omaha, and fifteen percent of the cases in Minneapolis.¹⁰⁸ Alone among the six experiments, Charlotte excluded male victims entirely.

Perhaps the most striking difference was in the legal threshold of eligibility. While misdemeanor assault (or battery) was the only offense eligible in four of the experiments, Omaha and Colorado Springs accepted other “domestic” offenses. Such cases constituted only twenty-three percent of the Omaha cases, but they constituted over half (59%) of the Colorado Springs cases.¹⁰⁹ It is not clear, then, that the Colorado Springs results are directly comparable to the other experiments, since most of its cases involved criminal “harassment” without evidence of physical contact. One experimental case I observed in Colorado Springs led to the arrest of a man who had been assaulted by a woman, on the rationale that he had come to her apartment that evening against her wishes. Under such circumstances in the Milwaukee experiment, and probably elsewhere, the woman would have been the experiment’s suspect.

These variations, however, are also largely true of the police environment to which the results of these experiments must be generalized if they are to inform policymaking. What all six experiments had in common was that they all randomly assigned arrest and some sort of non-arrest. This alone made them among the most comparable studies of the effects of police practices ever undertaken.

¹⁰⁷ Pate, et al, *supra* note 104; SALLY FREELS & ROBERT F. BORUCH, CROSS-SITE DATA REPORT, SPOUSE ASSAULT REPLICATION PROJECT 12 (NAT’L INST. OF JUST., Jan. 26, 1990).

¹⁰⁸ *Id.*

¹⁰⁹ *Id.*

B. IMPLEMENTATION

The comparability of results was further limited, however, by the diversity in actual implementation of the experiments. Table 1 summarizes these differences, along with the differences in the design elements. The Table reveals how much information we still lack on some key aspects of comparability of the experiments, such as the length of time arrestees are held in custody. Most of the important characteristics for comparison of the designs have been reported, however. They show great variance on some issues, and much less on others. What is most striking is that there is no consistent difference in any of these characteristics that matches the division of the six experiments into the two different substantive findings of deterrence or escalation effects.

The greatest variation across the experiments was in the demographic characteristics of the samples, which is exactly what the Minneapolis authors had recommended. Such variations helped to reveal whether arrest would have the same effects in different kinds of cities.

What was less variable, although far from consistent, was the nature of the arrest treatment. While the average number of hours spent in custody after arrest varied in percentage terms, it was generally reported to be in the range of nine to fifteen hours. Judging by the relatively small difference between the effects of short (three hour) and long (twelve hour) arrest in Milwaukee, these variations in custody time made little difference—assuming that they were measured accurately, which may not have been the case.¹¹⁰ Greater variability is found in the percentage of suspects who were handcuffed during arrest. While generally high where it was reported, Table 1 shows that Colorado Springs was again extremely different from the other experiments in having a rate of handcuffing four-fifths lower than those in the other cities. The greatest variability is found in the percentage of cases with prosecution leading to convictions and some sort of sentence which ranged from one percent in Milwaukee to sixty-four percent in Omaha. While the Minneapolis experiment, which had four percent convictions, could eliminate prosecution as a contributing influence on the effects of arrest, that was clearly not the case in Charlotte and Omaha.

More reassuring from a research design standpoint is that all six experiments had relatively high levels of compliance with the randomized designs. All of them improved on the Minneapolis compliance rate of eighty-two percent, ranging up to the Milwaukee rate of

¹¹⁰ Sherman, et al., *supra* note 104, at 845.

TABLE 1
DESIGN VARIATIONS IN SIX ARREST EXPERIMENTS^a

Design Elements	Deterrent Effect Cities			Escalation Effect Cities		
	Minneapolis	Miami	Colorado Springs	Milwaukee	Omaha	Charlotte
Custody Time (hours)	24+	14.6	?	2 and 12	16	9
Suspects Handcuffed	?	93%	22%	96%	?	80%
Suspects Convicted	4%	?	?	1%	64%	28%
Treated as Randomized	82%	90%	87%	98%	92%	84%
Married couples	35%	79%	69%	30%	42%	50%
Black suspects	36%	42%	31%	75%	43%	70%
White & Hispanic						
Victims	57%	60%	71%	28%	59%	30%
Suspects Unemployed	60%	29%	30%	47%	31%	22%
High-Crime Area						
Sample	Yes	No	No	Yes	No	No
Sample Sizes	314	907	1658	1200	330	650
Victims (6 mos+)						
Interviewed	23%	42%	58%	79%	73%	50%
Victim Medical Treatment	23%	?	?	13%	19%	4%

Key: information not reported = ?

^aLAWRENCE W. SHERMAN, POLICING DOMESTIC VIOLENCE (1992).

ninety-eight percent. While none of the experiments is perfect, in the historical experience of both medical and criminological experiments the compliance levels are substantial.

The most important methodological question, however, is whether the misassignment rate exceeded the base rate of recidivism, since that could severely distort an analysis of the results according to how each case was *supposed* to have been treated, rather than to what actually happened.¹¹¹ That approach to analysis is the standard for medical experiments, and was the method NIJ instructed the replication experiments to employ in the analyses summarized in Tables 1 and 2. Based on the current reports alone, it is impossible to tell how much statistical power each experiment may have lost from a high ratio of recidivism to the misapplication of randomly assigned treatments.

What is easier and equally important to assess is the effect of differences in response rates to the six-month followup victim interviews. These varied from twenty-three percent in Minneapolis to the seventy percent range in Omaha and Milwaukee. Where the response rates were under sixty percent (Charlotte, Metro-Dade, and

¹¹¹ Gerald S. Weinstein & Bruce Levin, *Effect of Crossover on the Statistical Power of Randomized Studies*, 48 ANN. THORACIC SURG. 137 (1989).

TABLE 2
SUMMARY OF RESULTS OF SIX IMMEDIATE ARREST
EXPERIMENTS, SAME VICTIM^a

Finding	City					
	Minneapolis	Omaha	Charlotte	Milwaukee	Col.Sprs.	Miami
6 month deterrence official measures	yes	no	no	no	no	1 of 2
6 month deterrence victim interviews	yes	border	no	no	yes	yes
6-12 month escalation official measures	no	yes	yes	yes	no	no
6-12 month escalation victim interviews	—	no	no	no	no	no
30-60 day deterrence official measures, any or same victim	yes	no	no	yes	no	1 of 2
30-60 day deterrence, victim interviews	yes	border	no	yes	—	yes
Escalation effect for unemployed	—	yes	—	yes	border	—
Deterrence for employed	—	yes	—	yes	yes	—

Key: relationship not reported = —

^aLAWRENCE W. SHERMAN, *POLICING DOMESTIC VIOLENCE* (1992).

Colorado Springs), there is a substantial likelihood of the interview sample having substantively different characteristics from the full sample. At the very least, this may account for any differences between the victim interviews and official records, which were found in all three replication cities with low response rates.

C. RESULTS

The replication results, just like the Minneapolis results, were examined in several different ways. The logic of employing multiple analyses is to increase our confidence in results that are consistent across different approaches and criteria. The difficulty this approach can create, however, is in the interpretation of conflicting results even within experiments. As Table 2 summarizes, that is what occurred in almost all of the replications.

One of the key questions about a treatment is how long the effects last. As Table 2 shows, there was evidence of short-term deterrence followed by a longer-term decay in Milwaukee, and evidence that the effects of arrest worsened over time in both Omaha and Charlotte. A major contribution of Dr. Dunford's article in this symposium, in fact, is his demonstration that the conventional six-month followup period, driven largely by funding agencies' needs to

contain costs, can substantially distort the measurement of the treatment effects. It is not clear why the effects of arrest worsen over time, but it is clear that the finding is not due to the idiosyncracies of any one experiment.

Another difference in the results concerns the suspects' repeat violence against the same victim or any other victim. The hypothesis that arrest displaces repeat violence onto new victims¹¹² has its adherents, and so the Milwaukee experiment tested it. The results of both approaches were essentially the same, but there was admittedly little deterrent effect to explain away with displacement. As Table 2 shows, most of the other research reports stuck to the focus of the Minneapolis design, which examined repeat violence only against the same victim.

The most basic difference in the analysis of these results is the method of measuring recidivism: official records versus victim interviews. Low response rates in the latter, however, can make these two measures inappropriate for comparison. Even where response rates are relatively high, they can still capture different populations. For example, the sample represented in victim interviews might generally consist of couples who are more socially bonded and therefore easier to locate than those on the full sample captured by official measures. This disparity produces some striking differences in results between the official and interview measures, but in highly consistent ways.

For example, *all of the evidence of escalation effects are found in the official data, and not in the victim interviews*. The evidence in Omaha, Milwaukee, and Charlotte that the arrested suspects became significantly more violent than those not arrested is only found in the full samples, and not in the reduced portions of the sample granting interviews. While this may be due to inherent differences between interviews and official records, it is equally plausible that escalation effects are more pronounced among the less socially bonded couples who are harder to interview. It is also important that the reader note this before reading the individual experimental reports, since the authors do not put their own findings in the context of these other results. Thus the Charlotte experiment reported in this issue, for example, makes little of the significant escalation effect it found in one measure because it was not confirmed by the other. The replication of this difference in two other cities, however, gives the finding much more prominence as a clue to the overall puzzle of these diverse results.

¹¹² See REISS, *supra* note 74.

Another striking fact about the two different measures is that *interviews generally show better results from arrest than does official data*. In both Colorado Springs and Miami, for example, at least one official measure shows no deterrent effect, while the victim interviews in both cities show a clear deterrent effect. This is again consistent with the hypothesis that the interviewed victims were different in important respects from the full sample. It is somewhat complicated, however, by the appearance of a deterrent effect as measured by one of the official measures in Miami.

These differences in the "main effects," or overall results for each randomly assigned treatment group, may mask underlying differences in the effects of arrest within each group. Ideally, such differences would be hypothesized in advance and built into the random assignment design by randomization within each subgroup, such as employed versus unemployed suspects. The political difficulties of obtaining approval for such a design in 1986, however, were insuperable. Thus, the experiments had an advance hypothesis that less socially bonded persons would react to arrest differently,¹¹³ but the full statistical power of the research design was not built around that hypothesis. This creates, in effect, some of the same problems of correlation versus causation which randomized experiments were designed to resolve. Some statisticians, however, endorse such sub-analyses as merely constituting experiments within experiments.¹¹⁴

The results of these tests with regard to interaction effects are, so far, consistently in support of the social bonding hypothesis. *Arrest has consistently more crime reduction effect on employed suspects than it does on unemployed suspects*. That is the good news. The bad news is that *the weaker a suspect's social bonds, the more likely it appears to be that arrest will backfire by causing increased violence*. The initial data analysis demonstrating this effect is found in the article by Sherman and his colleagues in this issue. That analysis was subsequently confirmed with more powerful analytic techniques, which showed in particular that the greatest escalation effect was among suspects who were both unmarried and unemployed.¹¹⁵ An identical analysis was performed on the Omaha data, which showed even stronger differences (although not statistically significant ones, due to the smaller sample

¹¹³ Lawrence W. Sherman, *Experiments in Police Discretion: Scientific Boon or Dangerous Knowledge?* 47 LAW & CONT. PROB. 61, 78 (1984).

¹¹⁴ Hans Zeisel, *Disagreement Over the Evaluation of a Controlled Experiment*, 88 AM. J. SOC. 378 (1982); Hans Zeisel, *Hans Zeisel Concludes the Debate*, 88 AM. J. SOC. 394 (1982).

¹¹⁵ Lawrence W. Sherman & Douglas A. Smith, *Crime, Punishment and Stake in Conformity: Legal and Extralegal Control of Domestic Violence*, AM. SOC. REV. (forthcoming 1992).

size). The interaction of unemployment and arrest was later replicated by Professor Berk and his colleagues, as reported in this symposium, although without taking marriage or other social bonding indicators into account. Finally, this pattern is consistent with the differences in Miami and Colorado Springs between the official and victim interview recidivism data.

Whether the pattern of interaction between social bonds and the effect of arrest holds up in Charlotte and Metro-Dade still remains to be seen. The fact that the interaction has been found in three very different cities, however, is a strikingly consistent finding no matter what the other experiments show. It demonstrates that in at least some cities, arrest does have different effects on different kinds of people. Under those conditions, it seems unlikely that the FDA would approve the marketing of mandatory arrest for all suspects.

Metro-Dade Details. While the statistical details of the other five experiments are published and accessible either in this issue or elsewhere, that is not true of the Metro-Dade experiment. Therefore, a full understanding of the replications is served by a brief summary of the Miami-Dade details.¹¹⁶

The Miami Metro-Dade experiment randomly assigned four different treatments. Arrest and no arrest were each divided into two groups: those with and without followup counseling by a special police unit. These followup visits occurred within about a week after the call to police during which arrest was randomly assigned. There were no significant differences associated with the counseling treatments. The effects of arrest in Miami are clearest when we collapse the four groups into two, comparing all arrest cases to all non-arrest cases.

The Miami victim interviews clearly show deterrent effects of arrest, although we must recall that the victim interview response rate of forty-two percent of the full sample suggested very different characteristics of that sub-sample compared to the full sample of 907 cases on which official records were gathered. At the time of the initial interview, eighteen percent of the non-arrest victims reported at least one incident in which the suspect hit, slapped, hurt, or tried to hurt the victim. This compares to only ten percent of the arrest group victims. The frequency differences are similar: a rate of 345 per 1,000 suspects in the non-arrest group, compared to 182 in the arrest group. The differences in both measures are statistically

¹¹⁶ This section is taken from SHERMAN, *supra* note 9, ch. 6, and based entirely upon Pate, et al., *supra* note 104.

significant, reducing the risk of repeat violence by about half. At the six month interviews, the prevalence effect was about the same. Twenty-seven percent of the non-arrest victims were hurt at least once compared to fifteen percent of the arrest group victims. The frequency difference was also of the same magnitude: 527 repeat incidents per 1,000 no-arrest suspects compared to 281 repeat incidents per 1,000 arrested suspects. These results were confirmed by significant differences in the results of three different tests of time to "failure," or the date of the first repeat incident of violence.

Repeat arrests for domestic violence also showed deterrent effects in Miami. While the number of repeat arrests in all groups was low, the prevalence rates were still significantly different among the suspects arrested and those not arrested. Out of the 465 arrested suspects, five (1.1%) were re-arrested for at least one new offense against the same victims within six months. Among the 442 suspects randomly assigned to non-arrest, the comparable number was seventeen (3.8%) subsequent arrests, or almost four times the prevalence rate. The difference in frequency was of about the same magnitude as the difference in prevalence (although possibly due to chance): twenty-two subsequent arrests per 1,000 arrested suspects, compared to seventy-seven per 1,000 among the suspects not arrested. Once again, all three tests of the difference in average time to first repeat offense showed significant benefits of arrest.

The one measure that showed no statistically significant difference was the offense reports naming the same suspect as the offender against the same victim. At six months, the prevalence rates of this measure were almost identical for the arrest and non-arrest suspects. This was true whether the repeat offense was aggravated or non-aggravated battery, or if it was any other offense. The frequency of any offense, however, was still somewhat higher among those suspects who had not been arrested (342 per 1,000) than among those who had been (290 per 1,000). The time to first repeat offense also showed no differences.

Why the two official measures should disagree is unclear, although we can speculate that there was greater measurement error—or variability in police recordkeeping—in the offense reports than in the arrest reports. Had they been combined, as they were in Minneapolis, they would have probably shown no differences by treatment, given the far greater volume of offense reports than of arrests. Citing the Miami experiment as a strong confirmation of the Minneapolis results must therefore be done with some caution. Nonetheless, there is still support for positing a deterrent effect in at least one official measure, and very strong, consistent support in all

five ways of analyzing the victim data—consistent with the social bonding hypothesis.

D. EXPLAINING DIVERSE RESULTS

Unfortunately, the social bonding hypothesis does not yet fully explain the differences in findings across cities. In order to give that hypothesis a full and fair test, the raw data from all of the six experiments must be merged and analyzed together. Different combinations of risk factors, for example, must be calculated, such as unmarried and unemployed (a category very rare in Colorado Springs) versus married and unemployed. Since not all of the raw data are yet publicly available, that analysis has not been possible. When it is, however, it may allow an explanation of the diverse results that goes beyond a simple bivariate inspection of the demographic differences portrayed in Table 1.

Taken by itself, Table 1 shows no clear demographic correlates between the deterrence and the escalation cities. Even where some differences appear, they apply to only two out of three cases. For example, the deterrent effect cities had a much lower average proportion of black suspects (38%) than the escalation effect cities (63%). The almost identical proportions of black suspects in escalation effect Omaha and deterrent-effect Miami, however, seriously weakens race as an explanation for the diverse results. The raw data might, however, reveal other characteristics correlated with race in the different cities, which may in turn support a social bonds theory or some other interpretation of differential reactions to domestic violence arrests.

One such characteristic, for example, may be the social structure of the community setting in which the suspects reside. Unemployment or social bonds may simply be a correlate of a more powerful neighborhood effect on suspects' reactions to arrest. As Professors Sampson and Wilson have observed, much social science has a tendency towards the "individualistic fallacy" of assuming that all causation is located in characteristics of individuals rather than of their social settings.¹¹⁷ Given the strong ecological correlates of crime rates, it is equally plausible that there are structural characteristics of neighborhoods—such as the proportion of unemployed males or of unmarried couples or of persons with criminal histories—that could determine how neighborhood residents react to arrest for domestic violence.

¹¹⁷ Robert J. Sampson & William J. Wilson, *Race, Crime and Urban Inequality*, in *CRIME AND INEQUALITY* (John L. Hagan, ed., forthcoming 1992).

This hypothesis would predict very different effects of unemployment on arrest reactions in Colorado Springs, for example, and in Milwaukee. The neighborhood proportions of unemployed persons, like the sample proportions reported in Table 1, are likely to be far higher in Milwaukee than in Colorado Springs. Thus we would expect more of an escalation or labeling effect in Milwaukee than in Colorado Springs. Similarly, we would expect more of a deterrent effect of arrest among employed persons in Colorado Springs than in Milwaukee, assuming the same ecological differences. As this issue reports, that is exactly what has been found to date. What remains to be done is to collect census tract and block level data for all the cases in the six experiments and test the ecological hypothesis. While such an analysis may be years in coming, it would be well worth the investment in unraveling the puzzle of the diverse results.

VI. THE SYMPOSIUM ISSUE

The main articles in this symposium make several contributions to the evolution of our knowledge about policing domestic violence. The gracious agreement of the authors to prepare papers for the issue has led, for the first time, to an assembly in one place of the original reports of research on most of the experiments. This will greatly simplify the task of policymakers and advisors seeking to digest the findings of the entire research program, and it will increase the odds that criminology may have an appropriate influence on the criminal law. The many requests for information on the experiments from state legislative counsel, even while the issue was in preparation, suggests the great need for such a compendium.

This symposium also allows comparison of the ways in which different researchers interpret the same data. The authors of neither the Omaha nor the Charlotte report would say that their experiments found that arrest caused an overall escalation in domestic violence, although that is how this introduction has characterized them. The reader is urged to attend to the reasons for these differences in interpretation, such as the consistent showing of escalation in official data, but not in victim interviews, in three experiments. Only by interpreting the experiments in relation to each other, which this symposium invites us all to do, can we fully understand the results.

The article by Ms. Zorza lays an excellent criminal law foundation for the significance of these criminological experiments. It comprehensively reviews the history of litigation and legislation that

helped to frame the key questions to which the entire research program responded. It clearly demonstrates the significant stimulus that criminal law reform movements can have in shaping criminology, which in turn may shape criminal law. And as a representative statement of the reaction of many domestic violence victim's advocates to the results of the replication studies, Ms. Zorza's commentary on the experiments illustrate the great divide in language, epistemology, treatment of scientific evidence, values and assumptions separating criminologists and activists on this issue.

While the flaws of the Minneapolis experiment were substantial, they go unmentioned in the advocates' critique. The flaws of the replications—at least those of the first three reported, which do not support a deterrent effect—are highlighted instead, because of what Zorza describes as the researchers' appearance of seeming "intent on returning to the old do-nothing or even blame-the-victim practices."¹¹⁸ Such *ad hominem* comments are unfortunately frequent in the attacks on the replication results, but they go with the territory. An assistant attorney general of the United States once told this writer it was a good thing that the Minneapolis experiment had found a deterrent effect, for otherwise that official would have been compelled to attack the experiment's methodology. It is unfortunate that so many advocates cannot accept the concern of criminologists with the plight of domestic violence victims, apparently because we do not always reach politically correct conclusions.

It is even more unfortunate that the battered women's movement in Milwaukee and elsewhere has shown little concern for the evidence that arrest positively harms black women in at least one poverty ghetto, where the majority of the suspects are unemployed and unmarried. Zorza's characterization of those results typically obscures the issue:

[E]ven if arrest may not deter unemployed abusers in ghetto neighborhoods, arrest still deters the vast majority of abusers . . . [w]e do not consider eliminating arrest for other crimes (e.g., robbery), however, because it may not deter a particular individual or class of individuals.¹¹⁹

The Milwaukee finding is not the failed deterrence of arrest, but the substantial *increases* arrest produces in the total volume of violence against victims of the ghetto poor unemployed. We have no evidence that arrest for robbery increases the total number of robbery offenses robbers commit, nor is arrest without prosecution the typical response to robbery—as it is in the realm of domestic assault. If

¹¹⁸ Zorza, *supra* note 10, at 72.

¹¹⁹ *Id.* at 66.

we had evidence that the typical criminal justice response to robbery backfired, we might respond to it with longer prison sentences upon conviction in order to counteract higher recidivism rates with greater incapacitation effects. Whether such a response makes sense when applied to a crime as pervasive as domestic violence—either for the families involved or for society—is another question altogether. What seems clear is that prosecutors are generally unwilling to do much with domestic violence cases, especially in cities with high volumes of such arrests (like Milwaukee). As long as that is true, we must soberly assess the wisdom of an “arrest-and-nothing-else” policy, since that is all we seem likely to get.

The comment that “most abusers” are deterred by arrest also misses a key point. While most abusers may be white, married and employed, the abusers *coming to police attention* may not be. Most of the crime, most of the police, and most police responses to reports of domestic violence in this country are found in cities of over 100,000.¹²⁰ Most of those cities, in turn, have substantial minority populations, in which victims disproportionately call on the police for assistance.¹²¹ Disregarding these facts in order to pursue a policy beneficial to women who do not live in poverty stricken ghettos—primarily white women—displays an unfortunate racial and economic insensitivity to the overall effects of mandatory arrest. Even if most abusers coming to police attention are not ghetto dwellers, we cannot write off as unimportant the victims of those who are.

Professors Hirschel and Hutchison have provided an exemplar of a comprehensive report on a randomized experiment in criminal law. Their careful, cautious and thorough analysis provides ample detail to answer a wide range of questions about how the experiment was done and why it might have found what it found. One may take issue with their conclusions, but not with the admirable way in which they present their data.

Their conclusion that Charlotte showed no difference in the effects of the arrest, citation, and separate-or-advise treatments seems difficult to support.¹²² The fact is that offenders who were arrested and issued citations (tickets to appear in court) showed significantly higher rates of violence than did other offenders, at least as measured by the official data. The authors attempt to explain the discrepancy between the data obtained in official reports and that

¹²⁰ Computations from FBI, *CRIME IN THE UNITED STATES* (1991).

¹²¹ BLACK, *supra* note 52, at 134.

¹²² See J. David Hirschel & Ira W. Hutchison III, *Female Spouse Abuse and the Police Response: The Charlotte, North Carolina Experiment*, 83 J. CRIM. L. & CRIMINOLOGY 73 (1992).

obtained in victim interviews by citing the under-reporting of repeat violence in the former, and the much higher rate of reported violence in the latter. We must recall, however, a more serious and profound difference between the official and victim data: the one-hundred percent coverage of the official data compared to the fifty percent coverage of the sample with the victim interview data. The two measures, in effect, compare a fruit basket (the full sample) to a group of oranges (the victim interviews). The differences between them may be just as plausibly related to the samples as to the method of measurement. The official data showing that arrest increased violence cannot be discounted, since it was the only measure for the entire fruit basket that constituted the randomized experiment. The causal inference value of the random assignment for such a small and possibly biased victim interview sample of what was randomized is highly questionable at best, just as it was in Minneapolis.

One point in the presentation of the Charlotte findings should also be clarified.¹²³ The analyses taking into account race, prior record and other variables might be read as showing that there is no interaction effect between those variables and arrest. That is not what the article says. The authors actually state that there was no two-way correlation between repeat domestic violence and the respective predictor variables of race, age, marital status and employment status. That does not mean that there is no three-way association between employment, arrest and repeat violence, which is what is reported in the Milwaukee, Omaha and Colorado Springs studies. A three-way interaction may be present even when there is no two-way correlation, just as it was when no two-way correlation was found in Milwaukee.¹²⁴ Thus, the Charlotte experiment does not falsify the prediction of such an interaction, at least as reported in this symposium. The three-way analysis still remains to be reported.

As the first to begin and report a Minneapolis replication experiment, Dr. Dunford has already detailed many of his results elsewhere.¹²⁵ In this symposium, however, he contributes new and vital information. In the original report of the Omaha replication experiment, for example, he reported only the first six months' followup data, none of which showed any differences in repeat violence rates between arrest and non-arrest treatments. In this analysis, however,

¹²³ *Id.*

¹²⁴ SHERMAN, *supra* note 9, app. 2, pt. 4.

¹²⁵ Dunford, et al., *supra* note 104; Franklyn W. Dunford, *System-Initiated Warrants for Suspects of Misdemeanor Domestic Assault: A Pilot Study*, 7 JUST. Q. 631 (1990).

Dr. Dunford includes data covering a full year's followup (and more), and while he concludes that these extended data do not alter the finding of no deterrent effect for arrest,¹²⁶ they clearly support an alternative interpretation: that arrest causes more repeat violence than non-arrest.

Dr. Dunford's Table 1 indicates that the frequency rate of repeat violence (measured by new arrests) in a one year followup was 248 incidents per 1,000 suspects among those who had been arrested, but only 154 incidents per 1,000 among those who had not been arrested. This difference is substantial: arrest caused a relative increase of sixty-one percent more repeat incidents than no arrest. The odds that this finding is *not* due to chance are ninety-three percent. This just barely misses the conventional standard of ninety-five percent, and clearly satisfies a widely used standard of ninety percent. A not quite as large difference found in the other official measure (39% more violence with arrest than non-arrest) has an eighty-four percent likelihood of not being a chance result. These substantial differences are thus very good bets to be more than mere flukes.

Dunford's main conclusion is that longer time periods are needed to assess the full effects of arrest and other interventions.¹²⁷ While the victim interview data show no differences at twelve months, they showed non-significant differences in favor of deterrence at six months. Thus both the interview and the official data show the same trend in effects from six months to twelve months: a decreasing benefit and an increased cost associated with arrest. This prompts us to ask how much more repeat violence would be associated with arrest in an eighteen month followup. Viewed in the context of Milwaukee's and Charlotte's results, the Omaha findings are consistent with the pattern of official measures showing that arrest increases recidivism.

Equal time requires a comment on the article about the Milwaukee experiment, for which I bear the principal responsibility. An objective commentator should raise the question of a "shotgun" approach to data analysis in which a lack of significant differences in the main experiment leads the analysts to search anywhere for differences within subgroups, just to find some significant differences to report. An objective commentator would also have to note, however, that a "rifle shot" hypothesis that zeroes in on a few theoretic-

¹²⁶ Franklyn W. Dunford, *The Measurement of Recidivism in Cases of Spouse Assault*, 83 J. CRIM. L. & CRIMINOLOGY 120 (1992).

¹²⁷ *Id.*

cally driven variables produces a far more compelling set of findings, less likely to be spurious or chance correlations. The commentator might concede that that was what happened in the Milwaukee analysis, given the publication of the "rifle shot" hypothesis in another law review symposium some eight years earlier.¹²⁸ A more skeptical writer than the present one might still harbor lingering doubts, however, and tend to place greater reliance on the outcome of attempts to replicate the employment interactions in other experiments. That skeptic might then seize upon Professor Berk's analysis as evidence that the increase in violence among the unemployed suspects caused by arrest has been exaggerated by the Milwaukee writers. In doing so, the skeptic would be basing a conclusion on less than the full story.

In the article by Professor Berk and his colleagues, we are once again treated to that which he has so often provided: the application to a criminological problem of a cutting-edge statistical technique. The lawyers who read this volume should take comfort in the fact that Bayesian analysis is not only unfamiliar to them, but to most criminologists as well. The technique consists of analyzing data in one study—Colorado Springs—based on findings from two prior experiments in which interactions between arrest and unemployment had been found (Milwaukee and Omaha).¹²⁹ The conclusion that an interaction effect is found in all three experiments is consistent with the social bonds hypothesis discussed above. The conclusion that an increase in domestic violence among the unemployed is not replicated warrants further comment.¹³⁰

Professor Berk's analysis of the effects of arrest in this symposium are limited to the *prevalence* of recidivism, or the percentage of suspects with one or more detected acts of repeat domestic violence. Of greater significance to both victims and police, however, may be the *frequency* of repeat violence, or the total number of attacks. As the report on the Milwaukee experiment in this symposium demonstrates, prevalence and frequency can lead to somewhat different interpretations. Indeed, the conclusion that arrest increases repeat violence among unemployed suspects is based primarily on the results for frequency, not for prevalence. Yet Professor Berk does not report the frequency results for Colorado Springs, and does not incorporate the frequency results for Omaha and Milwaukee.

As it happens, we do have the frequency results for both Mil-

¹²⁸ Sherman, *supra* note 113.

¹²⁹ Sherman & Smith, *supra* note 115.

¹³⁰ Richard A. Berk et al., *A Bayesian Analysis of the Colorado Springs Spouse Abuse Experiment*, 83 J. CRIM. L. & CRIMINOLOGY 170 (1992).

waukee (reported in this issue) and Omaha (reported elsewhere).¹³¹ In both experiments, the frequency data show that arrest produces a pronounced increase in repeat violence among the unemployed. While Berk correctly reports a small prevalence difference in Omaha among unemployed suspects (57% among those arrested and 53% among those not), the frequency difference is in the same direction but much larger. The Omaha frequency data show that arrest reduces the rate of future violence among the employed from 280 to 176 incidents per 1,000 suspects per year (a 37.1% reduction). Among the unemployed, arrest increases the rate of future violence by 52.2% from 412 to 627 incidents per 1,000 suspects per year. Given the small number of cases in the Omaha sample (sixty-four unemployed and 175 employed persons), these differences do not achieve statistical significance. Yet the differences between these point estimates of the rate of future violence are larger than in Milwaukee (which had a forty-three percent increase in frequency caused by arrest among the unemployed), and in the same direction. It seems hard to justify a conclusion that arrest does not backfire for unemployed suspects, at least in those two cities. Frequency data for Colorado Springs may well show similar results.

Even if they do not, however, we must again note the differences between Colorado Springs and the other experiments: most of the suspects there were married and most of the offenses involved not physical violence but verbal harassment. Furthermore, Colorado Springs apparently has no poverty-stricken ghettos with high proportions of persons with low marriage rates, long-term unemployment, low prevalence of high school education, high prevalence of prior arrests, and other similar characteristics. Thus, the "good risks" seem better and the "bad risks" seem not so bad in "hi-tech" Colorado Springs as in "rust belt" Milwaukee and Omaha. Given the major differences between the samples in the three cities, it is arguably impressive that the employment interaction results are as similar as they are.

I will not be commenting on the other commentaries in this issue. Suffice it to say that the Journal editors have chosen a distinguished group of commentators representing a broad range of viewpoints, from police to women's advocates to legal scholars.

VII. WHAT HAVE WE LEARNED?

After more than a decade of evaluating arrest for misdemeanor domestic violence, we still have much to learn. The jigsaw puzzle of

¹³¹ SHERMAN, *supra* note 9, ch. 7.

diverse results in different cities has not been put together, and too many pieces are still missing. Many alternate approaches still remain untested or unreplicated, such as the impressive Omaha result that issuing warrants for *absent* offenders reduced repeat violence by fifty percent.¹³² Nonetheless, it is time we took stock of what we have learned, both about the substance of the problem and about the process of doing policy-relevant criminology.

A. DOMESTIC VIOLENCE ARRESTS

One response to the replication results is that it is too early to reach any policy recommendations.¹³³ This view implies that the burden of proof must be on any argument to undo mandatory arrest laws, or stop them from being passed. Such a view, however, runs contrary to the principles laid down by the Federal Judicial Center's Advisory Committee on Experimentation in the Law, as cited above. The question as of 1984 was whether an innovation of greater severity should be adopted—preferred or mandatory arrest. The initial experiment supported the innovation, but with reservations. On balance, the subsequent experiments have not.

Even if we disregard the evidence of increased domestic violence caused by arrest in some cities and with some kinds of offenders, the weight of the evidence fails to justify an innovation of greater severity—at least on specific deterrent grounds alone. Yet it is those grounds, alone, which have morally justified the entire program of research. If the FDA had to make a decision about allowing mandatory arrest to go on the market based on these experiments, it seems doubtful that it would. On the other hand, if it were asked to allow some doctors to use arrest on a selective basis when it was most likely to be effective, they might well do that.

Arrest is not a drug, of course, and it is constrained by principles of justice. The unfairness of arrest guidelines based on employment status would be unthinkable, regardless of its effectiveness. What may ultimately be acceptable, under the existing principles of community policing, is different police policies or practices for different neighborhoods. Police discretion already varies widely by neighborhood,¹³⁴ and community policing is trying to make it vary even more explicitly in response to community pref-

¹³² Dunford, *supra* note 126.

¹³³ See, for example, the comments of Professor James Q. Wilson in Daniel Goleman, *Do Arrests Increase the Rates of Repeated Domestic Violence?*, N.Y. TIMES, Nov. 27, 1991, at C8.

¹³⁴ Douglas A. Smith, *The Neighborhood Context of Police Behavior*, in COMMUNITIES AND CRIME, 8 CRIME AND JUSTICE: A REVIEW OF RESEARCH (Albert J. Reiss, Jr. & Michael Tonry eds., 1986).

erences.¹³⁵ A local option approach, informed by research on the specific deterrent effects of arrest in different communities, might be the best way to develop a workable policy from the findings.

This possibility can be fully assessed, however, only after further analyses explore the neighborhood basis for the interaction effects observed to date. Whether that analysis can predict the likely effects of arrest based on census tract characteristics remains to be seen. It may well be more effective at that task, however, than we have been in predicting city-level effects so far.

Whatever approach may be taken on structuring discretion to use arrest, the key question is whether any discretion should remain in the hands of the police. This question has both a philosophical and a practical dimension. On philosophical grounds, it is clear that large segments of the legal and advocacy communities want no discretion invested in the police; they can cite legions of horror stories in support of their positions. On the practical side, no one has ever figured out how to eliminate police discretion. As Ms. Zorza quite correctly points out, we have learned that mandatory arrest laws are widely circumvented. That is all the more reason, it would seem, to develop an alternative approach.

The available research cannot say what that approach should be. All it can say is what the results of the six experiments show. Therein lies the lesson for the influence of criminology on the criminal law.

B. EXPERIMENTAL CRIMINOLOGY AND CRIMINAL LAW

The domestic violence experiments show that criminology can provide factual information about the criminal law and its consequences. That is about all it can do. It cannot, for example, control the ways in which participants in the political process describe (or distort) research results in advancing a point of view. It cannot ensure that its recommendations will be heeded, or that its conclusions will be believed. It cannot speak to value judgments about "just deserts," even when they are conveniently raised as a fallback position when evidence of deterrence is weakened. It cannot guarantee that its findings will resonate with the prevailing ethos of the age, as the Minneapolis findings did but the replication findings did not.

The Minneapolis findings stirred enormous interest by a wide range of writers and editorialists, who hailed the results as a breakthrough.¹³⁶ The replication results received grudging acceptance in

¹³⁵ MALCOLM K. SPARROW ET AL., *BEYOND 911: A NEW ERA FOR POLICING* (1990).

¹³⁶ See e.g., Ellen Goodman, *Using 'Muscle' Against Wife-Beaters*, WASH. POST, April 19,

some of those quarters, and complete silence in most others. They were even attacked editorially by the Milwaukee newspapers.¹³⁷ It is clear that our zeitgeist in the 1990s still favors "getting tough," and that greater severity is more politically correct than lesser severity among a broad coalition of both liberal and conservative groups. This carries a sobering lesson: provisional policy recommendations made on initial research results may be widely accepted in support of that broad coalition, but subsequent findings that run against it may have far less influence. Undoing the effects of initial results may be much harder than some criminologists imagined,¹³⁸ largely because there is less rational interest in minimizing violence than one might have assumed. It appears that preferences for punishment have more ideological than pragmatic foundations, and that criminology can only speak to the pragmatic.

This is a sad commentary for a system of criminal law founded on the presumption of deterrence. It suggests that as criminology unravels the deterrence hypothesis in its full complexity, the criminal law is unlikely to respond to that information in ways that will maximize crime control. Rather, the principle of appropriate vengeance, already so strong in the sentencing guidelines movement, may become even stronger, making deterrence irrelevant. If this keeps up, we will have no need for a JOURNAL OF CRIMINAL LAW AND CRIMINOLOGY; a Journal of Just Deserts will do just fine.

But times change, and knowledge takes a long time to accumulate. By the time we have fully assembled the puzzle of diverse effects of domestic violence arrests, perhaps the political culture may become more open to adopting columnist Ellen Goodman's point of view:

What is progress after all in the course of sexual politics? Is it marked by an increase in the number of men in jail? Or by a decrease in the number of assaults? I don't want to choose between law enforcement and "crime prevention," but I would chart the long run of progress by the change in men's behavior.¹³⁹

1983, at A21; editorial, *Crime in the Kitchen*, N.Y. TIMES, April 15, 1983, at A30; editorial, *Arresting Wife Beaters*, WASH. POST, April 16, 1983, at A22.

¹³⁷ Editorial, *Don't Let Abusers Off the Hook*, MILWAUKEE J., Nov. 21, 1991.

¹³⁸ Sherman & Cohn, *supra* note 4.

¹³⁹ Ellen Goodman, *I'd Have Let Him Walk*, WASH. POST, Dec. 14, 1991, at A27.