Nebraska Law Review

Volume 81 | Issue 2 Article 5

2002

Evaluating Empirical Research Methods: Using Empirical Research in Law and Policy

Jennifer K. Robbennolt
University of Illinois at Urbana-Champaign, jrobbenn@illinois.edu

Evaluating Empirical Research Methods: Using Empirical Research in Law and Policy

*Jennifer K. Robbennolt**

TABLE OF CONTENTS

I. INTRODUCTION

David Baldus and his colleagues have conducted an extensive and useful study of capital punishment decisionmaking in the state of Nebraska.[1] Their research is a fine example of how empirical research can be relevant to questions of law and policy. In determining whether to have a system of capital punishment and, if so, how to structure such a system, it is extremely useful to have data about various aspects of how the existing system operates. The same can be said for innumerable legal and policy issues ranging from various aspects

* Associate Professor of Law & Senior Fellow at the Center for the Study of Dispute Resolution, University of Missouri School of Law. B.A. 1991, Willamette University; J.D. 1996, University of Nebraska-Lincoln; Ph.D. (Psychology) 1998, University of Nebraska-Lincoln. I appreciate the comments of David Baldus, Chris Guthrie, Monica Johnson, Grant Robbennolt, Bob Schopp, and Rod Uphoff, the research assistance of Jason Moore, Mitch Berry, and Josie Pottebaum, and the financial support of the University of Missouri Law School Foundation.

1. David C. Baldus et al., *Arbitrariness and Discrimination in the Administration of the Death Penalty: A Legal and Empirical Analysis of the Nebraska Experience (1973-1999)*, 81 NEB. L. REV. 486 (2002) [hereinafter *Nebraska Study*].

of jury decisionmaking in both criminal and civil cases,[2] to eyewitness identifications,[3] to inheritance policy,[4] to affirmative action,[5] and to contract law,[6] to mention just a few. Legislators routinely consider issues that are informed by empirical research and judges are called upon to act as gatekeepers who regulate the admissibility of scientific evidence in court.[7]

In any discussion of the relevance and usefulness of empirical research in the law, there is, however, persistent tension between the methods of social science and the theory, goals, and settings of law and policy. As David Faigman notes,

> A more vexing aspect of data collection for the law comes from the basic incongruity between what scientists study and what the law is interested in knowing. . . . The law invariably relies on applied science, which typically involves extrapolating from controlled laboratory tests that are highly artificial or generalizing from noncontrolled field tests that contain many confounding variables.[8]

Using the study of capital punishment in Nebraska as a starting point, the purpose of this paper is to explore more broadly the tensions and tradeoffs at the intersection of social science methodology and the law from the perspective of legal and policy decisionmakers who are called upon to utilize empirical research. Those who use social scientific research to inform the law must do more than distinguish "good" research from "bad," they must also face the inevitable question of how to appropriately use well-done, but inherently imperfect research, for legal and policy purposes.

---

2. *See generally* Edith Greene et al., *Jurors and Juries: A Review of the Field*, *in* TAKING PSYCHOLOGY AND LAW INTO THE TWENTY-FIRST CENTURY, (J. Ogloff ed., 2002); Neil Vidmar, *The Performance of the American Civil Jury: An Empirical Perspective*, 40 ARIZ. L. REV. 849 (1998).
3. *See generally* Gary L. Wells et al., *Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads*, 22 LAW & HUM. BEHAV. 603 (1998); Gary L. Wells, et al. *From the Lab to the Police Station: A Successful Application of Eyewitness Research*, 55 AM. PSYCHOL. 581 (2000).
4. *See, e.g.*, Mary Louise Fellows et al., *Committed Partners and Inheritance: An Empirical Study*, 16 LAW & INEQ. 1 (1998); Monica K. Johnson & Jennifer K. Robbennolt, *Using Social Science to Inform the Law of Intestacy: The Case of Unmarried Committed Partners*, 22 LAW & HUM. BEHAV. 479 (1998); Rita J. Simon et al., *Public Versus Statutory Choice of Heirs: A Study of Public Attitudes About Property Distribution at Death*, 58 SOC. FORCES 1263 (1980).
5. *See, e.g.*, David L. Chambers et al., *Michigan's Minority Graduates in Practice: The River Runs Through Law School*, 25 LAW & SOC. INQUIRY 395 (2000).
6. *See, e.g.*, Russell Korobkin, *Inertia and Preference in Contract Negotiation: The Psychological Power of Default Rules and Form Terms*, 51 VAND. L. REV. 1583 (1998); Dennis P. Stolle & Andrew J. Slain, *Standard Form Contracts and Contract Schemas: A Preliminary Investigation of the Effects of Exculpatory Clauses on Consumers' Propensity to Sue*, 15 BEHAV. SCI. & L. 83 (1997).
7. *See* Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
8. DAVID L. FAIGMAN, LEGAL ALCHEMY: THE USE AND MISUSE OF SCIENCE IN THE LAW 53 (1999).

Toward this end, Part II uses empirical research on capital punishment decisionmaking to illustrate the tension between methodological concerns and legal theory or settings. The first example is drawn from the recent study of capital punishment in Nebraska conducted by Baldus and his colleagues and demonstrates a tension between methodological concerns and legal theory. The second example draws on experimental research that examines the effects of death qualifying jurors and the comprehensibility of capital punishment penalty phase jury instructions. This example highlights the tradeoff between realism and experimental control, a tension inherent in social science research that has implications for how empirical research is used in the law. Part III identifies several obstacles to careful consideration of these tensions and tradeoffs, and suggests that cognitive bias and a lack of familiarity with the scientific method can interfere with evaluations of the methodology of empirical research. Part IV suggests ways of overcoming these obstacles in order to better achieve careful consideration of the tensions and tradeoffs that exist at the intersection of law and methodology.

II. SOCIAL SCIENCE METHODS AND LEGAL QUESTIONS

There are any number of methodological approaches to addressing research questions that have relevance for law and policy, each with its own advantages and limitations. In considering the relevance of empirical research for answering legal and policy questions, decisionmakers ought to be concerned about several different aspects of the research. For example, those evaluating empirical research should be concerned that the measures used to quantify legal concepts are sufficient to capture the construct at issue (construct validity),[9] the degree to which the research design allows inferences to be drawn about causal relationships between variables (internal validity),[10] and the degree to which the research findings can be generalized to persons, times, and settings beyond those in which the research was conducted (external validity).[11] The following sections illustrate some of the tensions between these methodological concerns and legal concepts or settings.

9. THOMAS D. COOK & DONALD T. CAMPBELL, QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS FOR FIELD SETTINGS 39, 59-64 (1979); *see also* Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 80-81 (2002) ("This means that our comparisons and, ultimately, our answers to research questions are only as valid as the measures we have developed. If those measures do not adequately mirror the concepts contained in our theories, the conclusions we draw will be faulty.").

10. COOK & CAMPBELL, *supra* note 9, at 38.

11. *Id.* at 39.

A. Archival Research - The Nebraska Death Penalty Study

The study of capital punishment in Nebraska recently conducted by Baldus and his colleagues is an example of archival research in which the researchers measure variables present in actual cases and attempt to identify relationships between those variables. Archival studies, such as this one, have the important benefit of high external validity. That is, this study was able to examine decisions from real cases in which judges and prosecutors were faced with rich case facts and subtleties of evidence, and were required to make decisions about sentences for real defendants. Archival research in which case characteristics and decisions are carefully coded is extremely useful for identifying patterns and associations between aspects of the cases and the decisions that are made.[12]

Specifically, the study was largely concerned with the relationship between culpability and sentencing outcomes. Thus, the study examined the disposition of death-eligible cases over a twenty-six year period, measuring culpability primarily by looking at the record in each case and coding "the strength of evidence of each of the statutory aggravating and mitigating circumstances" in order to examine the ways in which dispositions varied with culpability.[13]

In order to quantify different levels of the legally relevant concept of case "culpability," some degree of abstraction is necessary; that is, the concept of culpability must be operationalized so that it may be measured. In measuring culpability, the study relied heavily on counting the number of statutory aggravating and mitigating factors in each case.[14] There are several good justifications for this approach. While a wide variety of case characteristics could be and, in fact, were coded, the small sample of capital cases in Nebraska precludes the use of a large number of variables in the statistical analyses.[15] Accordingly, counting statutory aggravators and mitigators is a relatively straightforward way in which to measure defendant culpability. Such an approach is quantifiable, it is based on factors that the legislature has deemed to be notably relevant case characteristics, and it likely captures many of the case factors that are important to decisionmak-

12. For a discussion of archival research methods see generally Robert J. MacCoun, *Inside the Black Box: What Empirical Research Tells Us About Decisionmaking by Civil Juries*, *in* VERDICT: ASSESSING THE CIVIL JURY SYSTEM 137 (Robert E. Litan ed., 1993); Neil Vidmar, *Making Inferences About Jury Behavior from Jury Verdict Statistics: Cautions About the Lorelei's Lied*, 18 LAW & HUM. BEHAV. 599 (1994).

13. *Nebraska Study*, *supra* note 1, at 532, app. A.

14. *Id.* at 534.

15. *Id.* at 535-36. There were only 185 prosecutions in death-eligible cases (covering 175 death-eligible defendants) over the period of the study. *Id.* at 541, 543.

ing. Indeed, these measures appear to have had a high degree of predictive power.[16]

Nonetheless, these measures do not completely correspond to the legal theory of capital punishment, which emphasizes individualized assessment of each case. The Supreme Court has emphasized that the decisionmaker is to make an "individualized decision" about whether a particular defendant should be sentenced to death.[17] In *Woodson v. North Carolina*, the Court found that:

> A process that accords no significance to relevant facets of the character and record of the individual offender or the circumstances of the particular offense excludes from consideration in fixing the ultimate punishment of death the possibility of compassionate or mitigating factors stemming from the diverse frailties of humankind. It treats all persons convicted of a designated offense not as uniquely individual human beings, but as members of a faceless, undifferentiated mass to be subjected to the blind infliction of the penalty of death.[18]

The Court concluded that, "in capital cases the fundamental respect for humanity underlying the Eighth Amendment requires consideration of the character and record of the individual offender and the circumstances of the particular offense as a constitutionally indispensable part of the process of inflicting the penalty of death."[19] Consistent with this principle, the Court has held that the sentencer must be able to consider, "*as a mitigating factor*, any aspect of a defendant's character or record and any of the circumstances of the offense that the defendant proffers as a basis for a sentence less than death."[20]

---

16. *Id.* at 550 ("The most significant factor explaining the pattern of capital charging and sentencing outcomes in Nebraska is the number of statutory aggravating circumstances in the cases.").
17. Lockett v. Ohio, 438 U.S. 586, 605 (1978) ("Given that the imposition of death by public authority is so profoundly different from all other penalties, we cannot avoid the conclusion that an individualized decision is essential in capital cases."); *see also* Sumner v. Shuman, 483 U.S. 66, 78 (1987) ("[W]e conclude that a departure from the individualized capital-sentencing doctrine is not justified and cannot be reconciled with the demands of the Eighth and Fourteenth Amendments."). Courts are expected to guide the discretion of the jury in conducting this individualized assessment. *See* Gregg v. Georgia, 428 U.S., 153, 189 (1976) ("Furman mandates that where discretion is afforded a sentencing body on a matter so grave as the determination of whether a human life should be taken or spared, that discretion must be suitably directed and limited so as to minimize the risk of wholly arbitrary and capricious action."); Furman v. Georgia, 408 U.S. 238 (1972).
18. Woodson v. North Carolina, 428 U.S. 280, 304 (1976) ("[A] . . . shortcoming of the North Carolina statute is its failure to allow the particularized consideration of relevant aspects of the character and record of each convicted defendant before the imposition upon him of a sentence of death."); *see also* Roberts v. Louisiana, 428 U.S. 325 (1976).
19. *Woodson*, 428 U.S. at 304.
20. *Lockett*, 438 U.S. at 604. Moreover, "a statute that prevents the sentencer in all capital cases from giving independent mitigating weight to aspects of the defen-

Nebraska law regarding capital sentencing is consistent with these dictates. In particular, Nebraska Revised Statute section 29-2522 requires that sentencing judges consider "[w]hether sufficient aggravating circumstances exist to justify imposition of a sentence of death" and "[w]hether sufficient mitigating circumstances exist which approach or exceed the weight given to the aggravating circumstances."[21] In interpreting these statutory requirements, the Nebraska Supreme Court has made it clear that the capital sentencing decision is to be based on an individualized evaluation of all of the circumstances in the case, rather than on a process that simply counts statutory aggravators and mitigators. In *State v. Joubert*, the court stated:

> Once the existence of one or more aggravating circumstances has been found to exist, § 29-2522 requires *not a mere counting of aggravating and mitigating circumstances* but, rather, a reasoned judgment as to what factual situations require the imposition of death and which of those can be satisfied by life imprisonment in light of the *totality of the circumstances* present"[22]

---

dant's character and record and to circumstances of the offense proffered in mitigation creates the risk that the death penalty will be imposed in spite of factors which may call for a less severe penalty." *Id.* at 605; *see also* Eddings v. Oklahoma, 455 U.S. 104, 112, 114 (1982) ("By holding that the sentencer in capital cases must be permitted to consider any relevant mitigating factor, the rule in *Lockett* recognizes that a consistency produced by ignoring individual differences is a false consistency."); Penry v. Lynaugh, 492 U.S. 302, 327-328 (1989) ("Indeed, it is precisely because the punishment should be directly related to the personal culpability of the defendant that the jury must be allowed to consider and give effect to mitigating evidence relevant to a defendant's character or record or the circumstances of the offense.").

21. Nebraska Revised Statute section 29-2522 provides that:

> After hearing all of the evidence and arguments in the sentencing proceeding, the judge or judges shall fix the sentence at either death or life imprisonment, but such determination shall be based upon the following considerations:
> (1) Whether sufficient aggravating circumstances exist to justify imposition of a sentence of death;
> (2) Whether sufficient mitigating circumstances exist which approach or exceed the weight given to the aggravating circumstances; or
> (3) Whether the sentence of death is excessive or disproportionate to the penalty imposed in similar cases, considering both the crime and the defendant.

NEB. REV. STAT. § 29-2522 (Reissue 1995). Statutory aggravating and mitigating circumstances are listed in Nebraska Revised Statute section 29-2523. Nebraska's system of capital punishment differs from that of most states that have such a system in that the judge or panel of judges is to determine the sentence, rather than a jury. However, the United States Supreme Court has recently held that capital defendants are entitled to have a jury determine any facts, such as aggravating circumstances, that are "necessary for imposition of the death penalty." Ring v. Arizona, 122 S. Ct. 2428 (2002).

22. State v. Joubert, 224 Neb. 411, 425-26, 399 N.W.2d 237, 248 (1986) (emphasis added); *see also* State v. Reeves, 239 Neb. 419, 476 N.W.2d 829 (1991); State v. Simants, 197 Neb. 549, 551, 250 N.W.2d 881, 892-93 (1977) ("In the balancing of the aggravating and mitigating circumstances, we emphasize that a death pen-

Toward this end, sentencing courts are "required to consider any relevant evidence in mitigation," including mitigating circumstances beyond those listed in the statute.[23] Importantly, for present purposes, such evidence would not be captured by a counting of statutory mitigators.

Thus, it is clear that the legal conception of capital sentencing decisionmaking contemplates individualized assessment of all of the facts and circumstances in the case that are relevant to the offender's culpability in order to determine whether a sentence of capital punishment is appropriate. However, it would be methodologically difficult, if not impossible, to measure "the totality of the circumstances" and "any relevant evidence." This is the other side of the richness attendant to archival research; each case has different facts and may present qualitative differences or difficult to quantify subtleties that influence prosecutorial or sentencing decisions. There may, then, be additional factors that were not controlled for, or even coded, and these additional factors may confound the results obtained.[24] Accordingly, there

---

alty will not be imposed simply because the aggravating circumstances may outnumber the mitigating circumstances. Rather the test is whether the aggravating circumstances in comparison outweigh the mitigating circumstances."); State v. Stewart, 197 Neb. 497, 250 N.W.2d 849 (1977). Judges interviewed by Baldus and his colleagues expressed a view of the jurisprudence of capital punishment consistent with this approach. The judges

> completely rejected the suggestion that a 'rule of one' or any other quantitative standards were perceived to exist or were applied by either the Nebraska Supreme Court or the trial courts. In stating this judgment the former judges relied heavily on the language of the Nebraska Supreme Court that outcomes could not be based on mere counts, as well as their belief that the sentencing courts meticulously adhered to this rule.

Nebraska Study, *supra* note 1, at 512.

23. State v. Victor, 235 Neb. 770, 457 N.W.2d 431 (1990); *see also* State v. Holtan, 205 Neb. 314, 287 N.W.2d 671 (1980) (noting that the court is not limited to consideration of the mitigating factors listed in the statute, but shall consider "any matter relevant to the imposition of the sentence and receive any such evidence which the court deems to have probative value as to the character of the defendant").

24. For example, the researchers were unable to control for the presence or absence of victim impact evidence (or the nature of such evidence). *See Nebraska Study*, *supra* note 1, at 619. It may be difficult to code all possibly relevant variables, in part, because "[t]o identify before the fact those characteristics of criminal homicides and their perpetrators which call for the death penalty, and to express these characteristics in language which can be fairly understood and applied by the sentencing authority [or a researcher doing coding], appear to be tasks which are beyond present human ability." McGautha v. California, 402 U.S. 183, 204 (1972). Thus, no system of measurement is likely to capture every relevant piece of information. But, as Epstein and King note, "measurement allows us to put many apparently disparate events or subjects on the same dimension, making it far easier to comprehend at least one aspect of the phenomenon under study. . . . Even more to the point, understanding the real world always requires a certain level of abstraction, and so measurement of some kind plays a central role in empirical research. The key is that we abstract the right dimensions for our purposes, and that we measure enough dimensions of each subject to capture all the

is not a precise one-to-one correspondence between the measures used and the requirements of the law.

Given the difficulties posed by potentially uncontrolled variables and the lack of precise correspondence with the jurisprudence of capital sentencing inherent in this type of archival research, we should not uncritically accept the results of the study at face value as exactly measuring everything that a decisionmaker takes into account in making a decision. It is clear that they do not. There are a host of variables, both legal and extra-legal that may influence decisions regarding offender culpability or sentencing (e.g., details of the crime, aspects of the defendant's character, juror attitudes and affective responses, attorney argument styles, and reactions of the local community).[25]

Conversely, however, given the advantages of archival research, neither should we uncritically reject the results for this lack of precise correspondence.[26] It may not be the case that this lack of correspondence systematically influences the pattern of results.[27] For this reason, the results of the study of capital punishment in Nebraska provide an extremely useful look at both prosecutorial and judicial decisionmaking in capital cases, demonstrate a number of interesting relationships between defendant culpability and decisionmaking, and suggest several ways in which arbitrariness exists in the system.[28]

---

parts that are essential to our research question." Epstein & King, *supra* note 9, at 81.

25. *See* Richard Weiner, *Death Penalty Research in Nebraska: How Do Judges and Juries Reach Penalty Decisions?*, 81 NEB. L. REV. 757, 760 (2002). One study of prosecutors' closing arguments in the penalty phase of capital trials demonstrated that prosecutors tend to focus more on the moral and emotional aspects of the case in their arguments than on aggravating and mitigating circumstances. D. D. Logan, Why You Should Not Kill This Man, presented at the British Psychological Society's International Conference on Psychology and Law, Swansea, Wales, *described in* Valerie P. Hans, *Death By Jury* in CHALLENGING CAPITAL PUNISHMENT: LEGAL AND SOCIAL SCIENCE APPROACHES (Kenneth C. Haas & James A. Inciardi eds., 1988) ("Although it was also typical for prosecutors to state that the aggravating circumstances outweighed the mitigating circumstances, the overall pattern of their arguments in the penalty phase suggested that they perceived the emotional dimensions of the decision to be paramount.").

26. *See* Shari Seidman Diamond, *Instructing on Death: Psychologists, Juries, and Judges*, 48 AM. PSYCHOL. 423, 430-31 (1993) (noting that the possibility of unmeasured variables "offers a convenient rationale for discounting the results of such an archival analysis to an audience that may not be inclined to acknowledge the existence of undependable decisionmaking in capital cases").

27. *See infra* section IV.A.

28. *See Nebraska Study*, *supra* note 1.

B. Experimental Research - Death Qualification and Penalty Phase Instructions

A second type of tension that has implications for how empirical research is used in law involves a tradeoff between the ability to control and manipulate variables that are of interest and the degree to which the research setting reflects the relevant legal conditions (that is, the verisimilitude of the study). An example of this tension comes from a growing body of experimental research examining the factors that influence capital sentencing. In a typical experiment, a large number of participants evaluate the same simulated case.[29] All characteristics of the case (the details of the crime, the backgrounds and characteristics of the defendant and victim, etc.) are held constant; the only attribute varied is the characteristic or procedure of interest. Thus, observed differences in responses, such as sentencing decisions, can be attributed to the variable of interest, unconfounded by other influences. Simulation methodology allows isolation of specific variables, permits observation of deliberation processes, and allows experimental manipulation of legal rules and procedures as well as case and party characteristics.[30]

Experimental methodology has been usefully employed to examine a number of aspects of capital case decisionmaking—only two of which will I note here. First, a number of experiments have been conducted to explore the effects of death qualifying jurors in capital cases.[31]

29. In this context, I will focus primarily on jury simulation studies. Field experiments, in which trial features of interest are systematically manipulated within actual trials, may also be conducted. *See, e.g.*, Valerie P. Hans et al., *The Arizona Jury Reform Permitting Civil Jury Trial Discussions: The Views of Trial Participants, Judges, and Jurors*, 32 U. MICH. J. L. REF. 349 (1999); Larry Heuer & Steven D. Penrod, *Increasing Jurors' Participation in Trials: A Field Experiment with Jury Notetaking and Question Asking*, 12 LAW & HUM. BEHAV. 231 (1988); Larry Heuer & Steven D. Penrod, *Instructing Jurors: A Field Experiment with Written and Preliminary Instructions*, 13 LAW & HUM. BEHAV. 409 (1989). Field experiments may be feasible in some circumstances with the cooperation of the court (e.g., examinations of juror notetaking), but may not be possible in others (e.g., plaintiff or defendant characteristics).

30. For discussion of simulation research, see generally Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?* 23 LAW & HUM. BEHAV. 75 (1999); Robert M. Bray & Norbert L. Kerr, *Use of the Simulation Method in the Study of Jury Behavior: Some Methodological Considerations*, 3 LAW & HUM. BEHAV. 107, 117 (1979); Shari Seidman Diamond, *Illuminations and Shadows from Jury Simulations*, 21 LAW & HUM. BEHAV. 561 (1997); MacCoun, *supra* note 12; Wayne Weiten & Shari Seidman Diamond, *A Critical Review of the Jury Simulation Paradigm: The Case of Defendant Characteristics*, 3 LAW & HUM. BEHAV. 71, 75-83 (1979); *see also* Gregory Mitchell, *Taking Behavioralism Too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law*, 43 WM. & MARY L. REV. 1907 (2002).

31. Brooke M. Butler & Gary Moran, *The Role of Death Qualification in Venirepersons' Evaluations of Aggravating and Mitigating Circumstances in Capital Trials*, 26 LAW & HUM BEHAV. 175 (2002); C. L. Cowan et al., *The Effects of Death*

Briefly, in capital cases jurors are screened for their attitudes toward capital punishment and may be constitutionally disqualified from jury service if their attitudes would "prevent or substantially impair the performance of [their] duties as a juror in accordance with [their] instructions and oath";[32] the resulting jury is said to be "death qualified." Experimental studies in this area typically compare jurors or juries who have been death qualified with those who have not in their verdicts and other judgments in response to a simulated case. These studies have generally demonstrated that juries that have been death qualified tend to be more conviction prone than are juries that have not been so qualified.[33]

Second, there have been a number of studies investigating the comprehensibility of penalty phase jury instructions in capital cases.[34] In these studies, lay decisionmakers are asked to read or lis-

---

Qualification on Jurors' Predisposition to Convict and on the Quality of Deliberation, 8 LAW & HUM. BEHAV. 53 (1984); Pheobe C. Ellsworth et al., *The Death-Qualified Jury and the Defense of Insanity*, 8 LAW & HUM. BEHAV. 81 (1984); Craig Haney, *On the Selection of Capital Juries: The Biasing Effects of the Death-Qualification Process*, 8 LAW & HUM. BEHAV. 121 (1984).

32. Wainwright v. Witt, 105 S. Ct. 844, 852 (1985); *see also* Witherspoon v. Illinois, 391 U.S. 510 (1968); Hovey v. Superior Court, 616 P.2d 1301 (Cal. 1980).

33. See M. T. Nietzel et al., *Juries: The Current State of the Empirical Literature*, *in* PSYCHOLOGY AND LAW: THE STATE OF THE DISCIPLINE 23-52 (R. Roesch et al. eds., 1999), for a meta-analytic review (finding a small but reliable correlation between death penalty attitudes and verdict). *See, e.g.*, Butler & Moran, *supra* note 31; Cowan et al., *supra* note 31. A number of studies have explored the reasons why death-qualified juries might be more conviction prone, finding evidence that death-qualified jurors differ from "excludables" demographically as well as in their attitudes toward crime control and due process, and in their attitudes toward the prosecution and defense, *see, e.g.*, Robert Fitzgerald & Phoebe Ellsworth, *Due Process vs. Crime Control: Death Qualification and Jury Attitudes*, 8 LAW & HUM. BEHAV. 31 (1984), that they differ in how they interpret evidence, *see, e.g.*, William C. Thompson et al., *Death Penalty Attitudes and Conviction Proneness: The Translation of Attitudes into Verdicts*, 8 LAW & HUM. BEHAV. 95 (1984), that they differ in how they evaluate aggravating and mitigating circumstances, *see, e.g.*, Butler & Moran, *supra* note 31, and that exposure to voir dire that includes a death qualification process leads to more conviction proneness, *see, e.g.*, , Haney, *supra* note 31; Craig Haney, *Examining Death Qualification: Further Analysis of the Process Effect*, 8 LAW & HUM. BEHAV. 133 (1984); *see also* Craig Haney, Aida Hurtado, & Luis Vega, *"Modern" Death Qualification: New Data on its Biasing Effects*, 18 LAW & HUM. BEHAV. 619 (1995).

34. *See, e.g.*, Shari Seidman Diamond, *supra* note 26; Shari Seidman Diamond & Judith N. Levi, *Improving Decisions on Death by Revising and Testing Jury Instructions*, 79 JUDICATURE 224 (1996); Craig Haney & Mona Lynch, *Comprehending Life and Death Matters: A Preliminary Study of California's Capital Penalty Instructions*, 18 LAW & HUM. BEHAV. 411 (1994) [hereinafter Haney & Lynch, *Comprehending Life and Death*]; Craig Haney & Mona Lynch, *Clarifying Life and Death Matters: An Analysis of Instructional Comprehension and Penalty Phase Closing Arguments*, 21 LAW & HUM. BEHAV. 575 (1997) [hereinafter Haney & Lynch, *Clarifying Life and Death*]; James Luginbuhl, *Comprehension of Judges' Instructions in the Penalty Phase of a Capital Trial*, 16 LAW & HUM.

ten to a set of jury instructions and to respond to questions designed to assess their comprehension of the instructions. In general, this research has documented that capital case penalty phase instructions are not well comprehended by jurors.[35] In particular, lay decisionmakers have difficulty understanding the concepts of aggravation and mitigation, particularly mitigation, and even misconstrue individual factors as falling in the opposite category.[36] Moreover, a number of studies have demonstrated that participants' level of understanding of the instructions influences their use of the evidence and their sentencing decisions.[37]

Concerns about experimental methods, such as the methods used in the studies just described, typically center around the external validity of the simulations, that is, the degree to which the effects observed generalize to the actual trial setting. Because field experiments are often not feasible, experimental simulations are conducted in the laboratory with varying degrees of verisimilitude (i.e., ecological validity). The participants range from college students in some studies to jury-eligible community members in others, comprehension of instructions may or may not be assessed in the context of a simulated penalty phase trial, case materials may be provided in written form or via audio or videotape, and so on. In order to gain control of specific variables of interest, some degree of realism is sacrificed.

BEHAV. 203 (1992); Mona Lynch & Craig Haney, *Discrimination and Instructional Comprehension: Guided Discretion, Racial Bias, and the Death Penalty*, 24 LAW & HUM. BEHAV. 337 (2000); Richard L. Wiener et al., *The Role of Declarative and Procedural Knowledge in Capital Murder Sentencing*, 28 J. APPLIED SOC. PSYCHOL. 124 (1998) [hereinafter Wiener et al., *Capital Murder Sentencing*]; Richard L. Wiener et al., *Comprehensibility of Approved Jury Instructions in Capital Murder Cases*, 80 J. APPLIED PSYCHOL. 455 (1995) [hereinafter Wiener et al., *Comprehensibility*]. *See generally*, Peter Meijes Tiersma, *Dictionaries and Death: Do Capital Jurors Understand Mitigation?*, 1995 UTAH L. REV. 1. For psychological research on jury instructions see generally AMIRAM ELWORK ET AL., MAKING JURY INSTRUCTIONS UNDERSTANDABLE (1982) (jury instructions on negligence); Amiram Elwork et al., *Juridic Decisions: In Ignorance of the Law or in Light of It*, 1 LAW & HUM. BEHAV. 163 (1977) (jury instructions on negligence); Vicky L. Smith, *When Prior Knowledge and Law Collide: Helping Jurors Use the Law*, 17 LAW & HUM. BEHAV. 507 (1993); Vicky L. Smith, *Prototypes in the Courtroom: Lay Representations of Legal Concepts*, 61 J. PERSONALITY & SOC. PSYCHOL. 857 (1991). For reviews see Peter English & Bruce Sales, *A Ceiling or Consistency Effect for the Comprehension of Jury Instructions*, 3 PSYCHOL. PUB. POL'Y & L. 381 (1997); Joel Lieberman & Bruce Sales, *What Social Psychology Teaches Us About the Jury Instruction Process*, 3 PSYCHOL. PUB. POL'Y & L. 589 (1997).

35. *See* Haney & Lynch, *Comprehending Life and Death*, *supra* note 34; Haney & Lynch, *Clarifying Life and Death*, *supra* note 34; Lynch & Haney, *supra* note 34; Weiner et al., *Comprehensibility*, *supra* note 34; Weiner et al., *Capital Murder Sentencing*, *supra* note 34.

36. *Id.*

37. *See* Lynch & Haney, *supra* note 34; Weiner et al., *Comprehensibility*, *supra* note 34; Weiner et al., *Capital Murder Sentencing*, *supra* note 34.

Therefore, it is inevitable that an experimental simulation will not precisely match all the conditions of the relevant legal situation.

While decisionmakers are (or ought to be) concerned primarily with the external validity or generalizability of the research findings, they often focus on the verisimilitude of a study as a proxy for external validity or as a convenient way in which to dismiss the research.[38] Indeed, a representativeness heuristic may operate such that decisionmakers discount the usefulness of research that does not precisely match the legal context.[39] Thus, courts have rejected empirical studies that do not mirror the trial setting. In *Lockhart v. McCree*, the Court rejected several studies of death qualification because the participants "were not actual jurors sworn under oath to apply the law to the facts of an actual case involving the fate of an actual capital defendant."[40] Courts have also rejected empirical studies of the comprehensibility of capital phase jury instructions on similar grounds.[41]

However, it is not necessary that a study mirror the conditions of an actual trial for the study to have high external validity and for it to be useful in a legal or policy context. An experimental simulation might be a relatively bare-boned approximation of a particular legal setting, but if it evokes behavior similar to, or consistently predictive of, that exhibited in the real world, the research has a high degree of generalizability. For example, Lynch and Haney investigated the capital case penalty phase instruction comprehension of jury-eligible community members and found a "pattern of errors [that] was strikingly

38. For discussions of the ecological validity of simulation studies see Bray & Kerr, *supra* note 30; Diamond, *supra* note 30; Weiten & Diamond, *supra* note 30.

39. *See* Margaret Bull Kovera et al., *Reasoning About Scientific Evidence: Effects of Juror Gender and Evidence Quality on Juror Decisions in a Hostile Work Environment Case*, 84 J. APPLIED PSYCHOL. 362 (1999). Using the representativeness heuristic, decisionmakers make categorizations based on the degree to which the object of the evaluation is representative of the category to the neglect of other relevant considerations. *See generally* Daniel Kahneman & Amos Tversky, *Belief in the Law of Small Numbers*, 76 PSYCHOL. BULL. 110 (1971).

40. Lockhart v. McCree, 476 U.S. 162, 171 (1986). *See* Phoebe C. Ellsworth, *Unpleasant Facts: The Supreme Court's Response to Empirical Research on Capital Punishment*, *in* CHALLENGING CAPITAL PUNISHMENT: LEGAL AND SOCIAL SCIENCE APPROACHES (Kenneth C. Haas & James A Inciardi, eds., 1988), for a detailed discussion of the criticisms of the empirical research in the *Lockhart* opinion.

41. For example, the court in *State v. Deck*, 994 S.W.2d 527, 542 (Mo. 1999), stated:

> Dr. Weiner's [sic] study, however, must be discounted because the people interviewed for the study did not act as jurors. They were given hypothetical facts that were different than the facts in this case, and they did not hear the testimony of witnesses, observe physical evidence or deliberate with eleven other jurors. More importantly, in the context of the instructions as a whole, the term 'mitigating' is always contrasted with the term 'aggravating' so that no reasonable person could fail to understand that 'mitigating' is the opposite of 'aggravating.'

similar to, and in most respects worse than, the ones . . . obtained in earlier studies using college student subjects."[42]

Thus, just as with archival research, there are disadvantages of simulation research that suggest that one should not uncritically accept the results of such studies at face value as representing exactly how jurors or judges make decisions in real contexts. Those evaluating the research ought to be aware of the potential effects of a lack of one-to-one correspondence between the simulation method used and the conditions of the actual legal decisions. Conversely, however, neither should one uncritically reject the results because the simulation did not exactly match the conditions in the real world. Experimental research provides useful information about how people make decisions, understand instructions and evidence, and respond to differences in a variety of contextual factors.

III. OBSTACLES TO EVALUATING EMPIRICAL METHODOLOGY

Even from these brief examples, it should be fairly clear that any individual empirical study will inevitably have some drawbacks as well as some advantages. Experimental research emphasizes control at the expense of realism, while archival research has the advantage of realism at the expense of tight control.[43] Of concern, however, is

42. Lynch & Haney, *supra* note 34, at 346; *see also* Bornstein, *supra* note 30 (reviewing jury simulation research comparing types of participants and methods of trial presentation and finding few differences in study results).

43. For purposes of simplicity, I have focused here on the broad distinction between archival and experimental research, but the same is true for other empirical methods. For instance, valuable information can be obtained from interviewing participants in capital cases such as jurors and judges. Indeed, interviewing jurors is the approach that has been taken by the Capital Jury Project. *See, e.g.*, William J. Bowers, *The Capital Jury Project: Rationale, Design, and Preview of Early Findings*, 70 IND. L. J. 1043 (1995); Theodore Eisenberg & Martin T. Wells, *Deadly Confusion: Juror Instructions in Capital Cases*, 79 CORNELL L. REV. 1 (1993); James Luginbuhl & Julie Howe, *Discretion in Capital Sentencing Instructions: Guided or Misguided?* 70 IND. L. J. 1161 (1995). The advantage of such an approach is that the information is obtained directly from the people who heard the evidence and made the decisions and who can speak to what they reacted to or thought was important. *See* Diamond, *supra* note 26, at 431. However, participants' reports are subject to a host of potential perceptual and memory errors; research in psychology has demonstrated that people are not always able to report the bases for their decisions and that their perceptions and memories can change once they have made a decision. *See generally*, RICHARD NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1980); J.W. Brehm, *Postdecision Changes in the Desirability of Alternatives*, J. ABNORMAL & SOC. PSYCHOL. 384 (1956); E. Greene, *Whodunit? Memory for Evidence in Text*, 94 AM. J. PSYCHOL. 479 (1981); Richard E. Nisbett & Timothy D. Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 PSYCHOL. REV. 231 (1977).

the inclination of many consumers of social science data to disregard these nuances of methodology, look at studies in isolation, and to either accept their results at face value as providing the answers to difficult legal or policy problems or, alternatively, to reject the results out of hand because the study did not account for every conceivable variable, did not perfectly mirror legal conditions, or did not measure legal variables in ways that perfectly capture the law. It would be more useful, however, for such consumers to engage in a more systematic consideration of these tensions and tradeoffs, rather than to merely conduct such a casual consideration of the relationship between research methods and legal or policy concerns.[44]

While it would be ideal for consumers of social science data to engage in careful review of policy relevant to empirical research, there are obstacles to accomplishing this thorough consideration of studies, methods, and legal relevance. First, psychological research has demonstrated that whether one uncritically rejects an empirical study out of hand or uncritically accepts it at face value is not a random process, but can be determined by cognitive and motivational biases. Second, many observers lack the background in research methodology that would facilitate more systematic thinking about these tensions.

A. Biased Assimilation

First, the phenomenon of biased assimilation has "troubling implications for efforts to ground contemporary policy debates in empirical analysis."[45] Biased assimilation is the tendency for evaluations of the methodology and persuasiveness of empirical research to be influenced by the extent to which the results of the research are consistent with the attitudes or expectations of the person doing the evaluation.[46] Thus, "judgments about the validity, reliability, relevance, and sometimes even the meaning of proffered evidence are biased by the apparent consistency of that evidence with the perceiver's theories and expectations."[47]

In the paradigmatic study of biased assimilation, Lord, Ross, and Lepper examined how people evaluated empirical research related to capital punishment and its deterrence of crime.[48] Specifically, they examined the "consequences of introducing the opposing factions to

44. *See infra* section IV.A.

45. Robert J. MacCoun, *Biases in the Interpretation and Use of Research Results*, 49 ANN. REV. PSYCHOL. 259, 267 (1998).

46. *See* Charles G. Lord, Lee Ross, & Mark R. Lepper, *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*, 37 J. PERSONALITY & SOC. PSYCHOL. 2098 (1979).

47. *Id.* at 2099.

48. *Id.*

relevant and objective data."[49] Participants who either favored capital punishment and believed it to have a deterrent effect or who opposed capital punishment and questioned its deterrent effect were recruited to participate in the study. Participants were asked to evaluate two empirical studies of the deterrent effect of capital punishment, one study described as finding an effect and the other described as failing to find such evidence. Thus, each participant read about one study that corroborated their initial beliefs about capital punishment and deterrence and one that refuted these beliefs. For each study, participants were given a description of the procedures used in the study, the results, a set of criticisms, and a rebuttal. For half of the participants, the study described as finding a deterrent effect was described as using a longitudinal methodology and the study finding no deterrent effect was described as using a cross-sectional methodology.[50] For the other participants, the methodologies of the studies were described in the opposite manner. Participants were asked to rate the methodology of each study as well as its persuasiveness.

Studies described as having results that were consistent with the participants' prior beliefs were rated as having higher methodological quality and as being more convincing than were studies described as having results that contradicted these prior beliefs, regardless of which methodology was described. Thus, participants who favored capital punishment and believed it to have a deterrent effect, found the prodeterrence study to be of higher quality and persuasiveness than the antideterrence study and vice versa.[51] Thus, the identical methodologies were evaluated differently depending on whether they produced findings that were consistent with the evaluators' prior attitudes and beliefs. Lord, Ross, and Lepper concluded that "[s]ubjects' decisions about whether to accept a study's findings at face value or to search for flaws and entertain alternative interpretations seemed to depend far less on the particular procedure employed than on whether the study's results coincided with their existing beliefs."[52]

The basic phenomenon of biased assimilation has been widely replicated using different populations of evaluators and a variety of research topic areas.[53] Importantly, for present purposes, Redding and

49. *Id.* at 2098.

50. The longitudinal study was described as involving a comparison of homicide rates in a number of states before and after each state adopted a system of capital punishment. The cross-sectional study was described as involving a comparison of homicide rates across states with and without capital punishment. *Id.* at 2100.

51. *Id.* at 2102.

52. *Id.* at 2106.

53. *See, e.g.*, Peter H. Ditto & David F. Lopez, *Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions*, 63 J. PERSONALITY & SOC. PSYCHOL. 568 (1992) (undergraduate students evaluating medical testing); Kari Edwards & Edward E. Smith, *A Disconfirmation Bias in the Evalu-*

Reppucci found a pattern of biased assimilation for state court judges and law students who were asked to make judgments about social science research on capital punishment and deterrence.[54] The results of the empirical studies that were presented to the U.S. Supreme Court in *Furman v. Georgia*[55] were varied so that they either supported or failed to support the death penalty. Participants judged the research to be more relevant to the legal issues, more likely to be admissible, and to have greater dispositive weight when the results were consistent with their personal attitudes about the death penalty than when the results contradicted those attitudes.[56]

Koehler extended the examination of biased assimilation to additional content areas and to more highly skilled evaluators, finding a pattern of results similar to that of Lord, Ross, and Lepper.[57] In one study, Koehler asked participants, who were graduate students in the natural and social sciences, to read summaries of two fictitious scientific controversies that were designed to induce prior beliefs about the hypotheses at issue. Participants were then asked to evaluate research reports related to the two issues for their relevance, methodological quality, and clarity.[58] Most participants endorsed the scientific norm that the outcome of the studies should not influence their judgments and reported that the outcome did not influence their

---

ation of Arguments, 71 J. PERSONALITY & SOC. PSYCHOL. 5 (1996) (undergraduate students evaluating a variety of arguments); Jonathan J. Koehler, *The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality*, 56 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 28 (1993) (graduate students and scientists evaluating fictional scientific controversy and research on ESP); Charles G. Lord et al., *Considering the Opposite: A Corrective Strategy for Social Judgment*, 47 J. PERSONALITY & SOC. PSYCHOL. 1231 (1984) (undergraduate students); Michael J. Mahoney, *Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System*, 1 COGNITIVE THERAPY & RES. 161 (1977) (journal reviewers evaluating research on behaviorism); Geoffrey D. Munro & Peter H. Ditto, *Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information*, 23 PERSONALITY & SOC. PSYCHOL. BULL. 636 (1997) (undergraduate students evaluating research on homosexuality); Geoffrey D. Munro et al., *Biased Assimilation of Sociopolitical Arguments: Evaluating the U.S. Presidential Debate*, 24 BASIC & APPLIED SOC. PSYCHOL. 15 (2002) (undergraduate students evaluating arguments made in a political debate); S. Plous, *Biases in the Assimilation of Technological Breakdowns: Do Accidents Make Us Safer?* 21 J. APPLIED SOC. PSYCHOL. 1058 (1991) (students, ROTC cadets, and staff of 2 peace organizations evaluating technological breakdowns).

54. Richard E. Redding & N. D. Reppucci, *Effects of Lawyers' Socio-political Attitudes on Their Judgments of Social Science in Legal Decision Making*, 23 LAW & HUM. BEHAV. 31 (1999).

55. 408 U.S. 238 (1972).

56. Redding & Reppucci, *supra* note 54, at 43. The effects were more pervasive for law students (evident in four different ratings) than for judges (evident in ratings of dispositive weight only). *Id.* at 44. Biased assimilation was not found for evaluations of research as to predictions of future violence by psychologists. *Id.* at 46.

57. Koehler, *supra* note 53.

58. *Id.* at 35-36.

judgments.[59] Nonetheless, participants gave more positive evaluations to the reports with findings that were consistent with their induced expectations than to those that were inconsistent with these expectations.[60]

Research replicating the biased assimilation effect and extending the work to explore the mechanisms underlying the effect has demonstrated both motivational and cognitive influences.[61] There is evidence that biased assimilation effects are mediated by the different affective responses that evaluators have to attitude-congruent and attitude-inconsistent research results.[62] Other researchers have fo-

59. *Id.* at 38-39. A majority of participants (64%) indicated that their judgments were not influenced by the outcome of the studies (24% indicated that they were influenced; 10% were not sure) and an ever larger majority (83%) indicated that such judgments should not depend on outcome (12% thought should; 5% not sure). *Id.* Moreover, "[t]hose who believed that the outcome of the study did not influence their quality judgments were actually influenced by the outcome as much as those who admitted some probable influence." *Id.*

60. *Id.* at 37. In a second study, participants, scientists with opposing views on the controversy over extrasensory perception (ESP), were asked to evaluate a research report describing an ESP study. The study was of either high or low methodological quality, and the results were described as supporting ESP, refuting ESP, or no results were described. *Id.* at 41. Again, despite the fact that most participants reported that they were not (and should not be) influenced in their evaluations by the outcomes of the studies, participants gave more favorable evaluations of the relevance, methodological quality, and clarity of studies that reported results consistent with their prior beliefs on the issue. *Id.* at 44. That is, parapsycholgists gave more positive evaluations to studies with results supportive of ESP and "skeptics" gave more positive evaluations to studies with disconfirming results. *See also* Mahoney, *supra* note 53 (finding biased assimilation in scientific journal reviewers).

61. To say that the bias is "motivated" is not to say that it is intentional. *See* MacCoun, *supra* note 45 ("*Motivation* is shorthand for the degree to which the bias has its origins in the judge's preferences, goals, or values; intentional bias is motivated, but not all motivated biases are intentional."). While biased interpretation of empirical research may be intentional in some cases, *see generally* MacCoun, *supra* note 45 (discussing fraud and advocacy as sources of bias); Shari Seidman Diamond & Jonathan D. Casper, *Empirical Evidence and the Death Penalty*, 50 J. SOC. ISSUES 177 (1994) (discussing reasons why the Court might be resistant to empirical research on capital punishment); Michael J. Saks, *Social Psychological Contributions to a Legislative Subcommittee on Organ and Tissue Transplants*, 33 AM. PSYCHOL. 680 (1978) ("These decision makers usually have looked to 'empirical research' not as a source of knowledge that will guide them to an informed conclusion but as a hook on which to hang their presuppositions."), it is unintentional bias that is primarily the focus of my remarks.

62. *See, e.g.*, Munro & Ditto, *supra* note 53. Munro and Ditto found that participants who held highly prejudicial views about homosexuality evaluated studies reporting results consistent with negative stereotypes about homosexuality as being of higher methodological quality than studies that used the same methods but reported stereotype-inconsistent results. These high prejudice participants also reported feeling more negative emotion after reading the non-stereotypical results and more positive emotion after reading the stereotypical results. Conversely, participants low in prejudice found studies with non-stereotypic results to be

cused on differences in the cognitive processes used to evaluate attitude-congruent and attitude-inconsistent information. The findings generally suggest that evaluators engage in a truncated evaluation process in response to results that are consistent with their attitudes, expectations, or theories—essentially being more willing to accept the results at face value. On the other hand, when evaluating information that disconfirms one's theory or expectations or that is not preferred, evaluators appear to spend more time considering alternative hypotheses,[63] expend more cognitive effort,[64] require more information to make a decision,[65] generate and consider more arguments (with higher proportions of them being refutational and redundant),[66] and display increased sensitivity to the quality of the information.[67] Evaluators may also have an overall "propensity to remember the strengths of confirming evidence but the weaknesses of disconfirming evidence."[68] Hence, Lord, Ross, and Lepper note:

> With confirming evidence, we suspect that both lay and professional scientists rapidly reduce the complexity of the information and remember only a few well-chosen supportive impressions. With disconfirming evidence, they continue to reflect upon any information that suggests less damaging 'alternative interpretation.'[69]

Thus, there appear to be parallel tendencies to accept empirical results that confirm previously held attitudes or expectations at face value and to look for ways to reject out of hand results that disconfirm or conflict with those expectations or attitudes. Importantly, neither of these tendencies is likely to be the most useful approach for evaluating empirical studies for use in legal or policy decisionmaking.

---

more methodologically sound than those using identical methods but reporting stereotypic results. These participants reported more negative emotion associated with the stereotypical results and more positive emotion associated with the non-stereotypical results. *Id.* at 641-46. Further analyses demonstrated that participants prior beliefs (i.e., prejudice) influenced their affective reactions to the studies described and that these affective reactions, in turn, influenced participants' methodological evaluations. *Id.* at 643 (study 1), 647 (study 2); *see also* Edwards & Smith, *supra* note 53, at 16 (finding larger effects for those high in emotional conviction).

63. Ditto & Lopez, *supra* note 53, at 575; Edwards & Smith, *supra* note 53, at 10.

64. Peter H. Ditto et al., *Motivated Sensitivity to Preference-Inconsistent Information*, 75 J. PERSONALITY & SOC. PSYCHOL. 53 (1998) (using cognitive load manipulation)

65. Ditto & Lopez, *supra* note 53, at 573.

66. Edwards & Smith, *supra* note 53, at 10, 12 (study 1), 17 (study 2); Ditto & Lopez, *supra* note 53, at 578 (citing more test-affecting irregularities if against preference).

67. Ditto et al., *supra* note 64, at 64.

68. Lord, Ross, & Lepper, *supra* note 46, at 2099.

69. *Id.* at 2099.

B. Methodological Background

A second major obstacle that prevents careful consideration of empirical research stems from the fact that many participants in the legal system, such as judges and legislators, who are called upon to evaluate empirical data, lack the background in science, empirical research methodology, and statistics that would facilitate a nuanced evaluation of social scientific evidence. As Faigman notes:

> Further complicating the law's use of science, legal consumers of scientific research often have little understanding of the product they are buying. In most areas of the law, those using science have little or no training in the subject. This is true for judges, jurors, legislators, and to a lesser extent, administrators. All judges and most legislators and administrators come from the ranks of lawyers. . . . Fewer than 10 percent of all students attending law school have undergraduate degrees in fields that require substantial math and science training . . . . Not only do they not have training in the particular subject, they have a more profound disability: most lawyers have little or no appreciation for the scientific method and lack the ability to judge whether proffered research is good science, bad science, or science at all.[70]

Not only do most law students not come from scientific backgrounds, but (notwithstanding some exceptions) legal education has also not tended to provide training in empirical research methods or statistics.[71] Lehman, Lempert, and Nisbett compared the statistical and methodological reasoning of graduate students in a variety of disciplines.[72] They found that despite few initial differences across disciplines, students in psychology and medicine improved dramatically in their ability to engage in statistical-methodological reasoning in the first three years of graduate school, while students in law did not.[73]

The impact of this lack of background in the scientific method can have important implications for the evaluation and application of empirical research. Research into dual process models of persuasion suggests that not only must an observer be motivated to engage in careful and effortful processing of a message, but the observer must also have

---

70. FAIGMAN, *supra* note 8, at 53-54. "[O]f the 535 members of the United States Congress, fewer than 1 percent have any significant training in science. Today, the route to being a 'statesman' is through law or business." *Id.* at 123.

71. *See* Epstein & King, *supra* note 9 (arguing that law schools ought to teach statistics/scientific methodology); Michael Heise, *The Importance of Being Empirical*, 26 PEPP. L. REV. 807 (1999).

72. Darrin R. Lehman et al., *The Effects of Graduate Training on Reasoning: Formal Discipline and Thinking About Everyday-Life Events*, 43 AM. PSYCHOL. 431 (1988).

73. *Id.* at 437. "Training in the law does not stress rules for dealing with variability or uncertainty in causal relations, and so it is not surprising that it produces no improvement in the ability to apply the statistical and methodological rules of the probabilistic sciences to either scientific studies or everyday-life events." *Id.* at 440.

the ability to engage in such processing.[74] Accordingly, one who is called upon to evaluate an empirical study for use in the legal or policy arena can only engage in a careful consideration of the implications of methodological choices if he or she has the ability to do so. Absent some background in empirical research methodology that enables a more nuanced consideration, one is more likely to fall back on less systematic heuristic processing.[75] Thus, one may be more likely to react to a piece of research in a way that is consistent with one's prior expectations or attitudes than to carefully consider how the methodological and legal tensions play out in the particular instance.

Recent research provides evidence that attorneys, judges, and jurors have difficulty assessing empirical research methodology and are not sensitive to differences in methodological quality.[76] Kovera, McAuliff, and Hebert presented jurors with one of several different descriptions of a study of sexual harassment and its methodology in the context of a civil trial. They found that jurors' assessments of the study were not sensitive to differences in how sexual harassment was measured.[77] In another experimental study, Kovera and McAuliff found that methodological flaws in a study of sexual harassment did

74. *See* RICHARD E. PETTY & JOHN T. CACIOPPO, COMMUNICATION AND PERSUASION: CENTRAL AND PERIPHERAL ROUTES TO ATTITUDE CHANGE (1986); Shelley S. Chaiken et al., *Heuristic and Systematic Processing Within and Beyond the Persuasion Context*, *in* UNINTENDED THOUGHT 212 (J.S. Uleman et al. eds., 1989).

75. *See, e.g.*, Joel Cooper et al., *Complex Scientific Testimony: How Do Jurors Make Decisions?* 20 LAW & HUM. BEHAV. 379 (1996) (noting how an expert's credentials had an effect on juror judgements when the expert's testimony was complex, but not when the testimony was easily understood); Kovera et al., *supra* note 39, at 372 (finding that jurors "relied on heiristic cues when evaluating the validity of scientific evidence," specifically the research's general acceptance and its ecological validity).

76. Margaret Bull Kovera & Bradley D. McAuliff, *The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers?*, 85 J. APPLIED PSYCHOL. 574 (2000); Kovera et al., *supra* note 39; Margaret Bull Kovera et al., *Assessment of the Commonsense Psychology Underlying Daubert: Legal Decision Makers' Abilities to Evaluate Expert Evidence in Hostile Work Environment Cases*, PSYCHOL., PUB. POL'Y, & L. (forthcoming 2002) (citing unpublished studies). There is also a corpus of evidence that laypeople in general have difficulty with statistical reasoning and with identifying flaws in empirical research. *See, e.g.*, RICHARD E. NISBETT, RULES FOR REASONING (Richard E. Nisbett ed., 1993); Christopher C. Jepson et al., *Inductive Reasoning: Competence or Skill?*, 6 BEHAV. & BRAIN SCI. 494 (1983); Richard E. Nisbett et al., *Teaching Reasoning*, 238 SCIENCE 625 (1987); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974) (sample sizes).

77. *See* Kovera et al., *supra* note 39, at 372. They compared a version in which sexual harassment was measured by a single confederate's rating of participants' sexual motivation to a version in which several other measures of sexual harassment were used (e.g., how close he sat, how many inappropriate questions, his evaluations of her qualifications) in addition to the confederate's rating. *Id.* at 367.

not significantly influence trial court judges' decisions about the admissibility of the study as evidence.[78] In a similar study of a national sample of attorneys, Kovera and McAuliff found that attorneys' ratings of the scientific quality of the study at issue were not influenced by flaws in the research method, and their reported cross-examination strategies were relatively unlikely to include issues related to these flaws.[79]

In order for empirical research to be optimally used to inform legal and policy decisionmaking, those called upon to evaluate the research and its implications must be able to identify the tensions between methods and legal theory and to appraise them. While a background in scientific methodology may not be sufficient to eliminate biased assimilation,[80] a basic grounding in the methods of social science is essential to a nuanced understanding of the implications of methodological choices. Those without a basic understanding of methods will be less likely to be able to identify the benefits of a particular methodological approach and will not be attuned to the drawbacks of the approach. Accordingly, they are less likely to be able to engage in a critical evaluation of the interplay between the research methodology and the legal or policy setting.

IV. STRATEGIES FOR EVALUATING EMPIRICAL RESEARCH

It is inevitable that there will be some tension between the strictures of the empirical methodologies employed to study legal or policy questions and the legal theory or setting. But, while there are obstacles to a sophisticated examination of these tensions, these obstacles are not inevitably insurmountable. Decisionmakers can attempt to self-consciously analyze the implications of a particular tension for the specific legal or policy setting. In addition, decisionmakers can endeavor to read and understand empirical research in the context of other relevant studies that are designed to address the same compli-

78. Kovera & McAuliff, *supra* note 76, at 580. In addition, very few judges mentioned internal validity threats (confound (13%), non-blind research assistant (9%), missing control group (8%)) when justifying their admissibility decisions. *Id.* at 581. These findings are particularly troubling given trial judges are now expected to act as gatekeepers and determine the reliability of scientific evidence. *See* Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 589 n.7 (1993).

79. Margaret Bull Kovera & Bradley D. McAuliff, Attorneys' Evaluations of Psychological Science: Does Evidence Quality Matter? (2002) (unpublished manuscript on file with the author), *described in* Kovera, Russano, & McAuliff, *supra* note 76. Social psychologists, on the other hand were sensitive to the differences in the methodological quality of the different descriptions of the study. *See* M. B. Russano & Margaret Bull Kovera, Psychologists' Evaluations of Valid and Flawed Psychological Science, paper presented at the 109th Annual Convention of the American Psychological Association, San Francisco (2001), *described in* Kovera et al., *supra* note 76.

80. *See infra* notes 97-98 and accompanying text.

cated legal or policy questions using both similar and different research methods. Finally, consumers of empirical research can master the basic concepts of research methodology and specific debiasing techniques that will enable them to engage in such a careful consideration of policy relevant studies.

A. Explore the Tensions

Rather than either uncritically accepting the results of any single study as reflecting an unambiguous "truth" about a legal setting or rejecting the study out of hand, those evaluating empirical research ought to self-consciously adopt the goal of careful consideration of the methodological-legal issues. That is, evaluators should attempt to identify the tensions and tradeoffs among empirical methods and the law (such as those identified here) and carefully consider what a specific methodological choice means for the interpretation and application of the research results. For example, consumers of the Nebraska capital punishment study ought to ask themselves how they might expect the details of the culpability measure to affect the results. Consumers of an experimental study of the effects of death-qualification or comprehension of instructions using students as participants should ask how using a different sample of participants would affect the results. If a different measure of culpability or different participants were used, how likely is it that the results would be different? Why? If differences are anticipated, of what magnitude and in what direction are they expected to be? What implications might those differences have for policymaking?

Decisionmakers ought to consider both the nature of the research question being studied[81] and the expected patterns of results. Of particular concern should not be expected differences in magnitude, but differences in the patterns of effects obtained with different measures or participant population (i.e., interactions between the variable of interest and the method variable).[82]

---

81. *See* Diamond, *supra* note 30, at 564 ("A less-than-full-scale trial stimulus can also provide a suitable vehicle if the research is designed to study decisionmaking issues that are not dependent on the nature of the evidence being presented at trial.").

82. *See generally* Bray & Kerr, *supra* note 30; Diamond, *supra* note 30, at 563 ("If differences between student and general adult populations were confined to differences in rates of conviction or willingness to find liability, and did not interact with experimental manipulations, the choice of a subject population would not be troubling."); Greene et al., *supra* note 2, at 253 (stating that, more important than main effect differences, "there might be 'interactive' differences in the way that students versus non-students . . . respond to experimental manipulations . . . . The second difference is more worrisome than the first because it suggests that factors associated with the methods used may limit the generalizability of research findings based on experimental manipulations of variables."); Michael J.

Consider the use of counting aggravating and mitigating circumstances as measures of culpability. It is plausible to expect that some other measure of culpability would result in the same directional pattern of results, but with effects of a somewhat different magnitude. Though somewhat interesting, these types of differences may not have widely divergent implications for policy. It would be more troubling if we were to find different patterns of effects as a result of using different culpability measures. For example, we might detect effects when using one set of measures, but not when other measures are used. Even more troubling might be a pattern of effects in which effects are detected using both measures, but the effects are in opposite directions, a pattern which seems unlikely in the present instance.[83]

In order to understand the implications of a study for law or policy, the tensions between the method and the law need to be identified, the alternative approaches specified, and an explicit theory articulated about why, in what ways, and to what extent the tension is likely to influence the results in ways that have implications for the policy or law at issue. Because such tensions are inevitable, evaluators must do more than superficially discount a piece of research because a tension exists or a tradeoff has been made (such will always be the case) or ignore the tension or tradeoff and accept the results at face value. Rather, evaluators must critically assess the nuances of the methodological choice and its implications for policy.

B. Research in Context

Not only must the tensions between empirical methodology and the law be articulated and carefully considered in the context of a single study, but, even more importantly, such tensions should be considered in the context of a body of research as a whole. It is a widely held view in social science that no single study will be dispositive of a research

---

Saks, *What Do Jury Experiments Tell Us About How Juries (Should) Make Decisions?*, 6 S. CAL. INTERDISC. L. J. 1 (1997).

83. See Robert M. Bray & Norbert L. Kerr, *Methodological Considerations in the Study of the Psychology of the Courtroom*, *in* THE PSYCHOLOGY OF THE COURTROOM (Norbert L. Kerr & Robert M. Bray eds., 1982) for a typology of interaction patterns. The same type of analysis can be conducted for differences between student participants and jury-eligible community residents. Haney and Lynch self-consciously considered the potential impact of various aspects of their methodology on the generalizability of their research. For example, when using students as participants in studies exploring comprehension of penalty phase instructions, they considered how using a broader set of jury-eligible participants would be likely to affect the results. They anticipated that using students would overestimate comprehension of instructions, primarily because of their educational background. *See* Haney & Lynch, *Comprehending Life and Death*, *supra* note 34, at 418; Haney & Lynch, *Clarifying Life and Death*, *supra* note 34, at 577. In subsequent studies, they found that, indeed, using jury-eligible participants resulted in even lower levels of comprehension. Lynch & Haney, *supra* note 34 at 338-39.

question.[84] Instead, researchers strive to conduct multiple studies in an area using multiple methods, participant populations, approaches, materials, and researchers. In doing so, researchers can maximize the inherent advantages of divergent methodologies and minimize the inherent disadvantages of each as well. This is the concept of convergent validity—using a number of approaches researchers hope to converge (or triangulate) on an understanding or explanation of a phenomenon. If archival studies can find an association between two variables in real world cases and experimental studies can isolate that effect in controlled settings, both using different participant populations, different trial materials, different measures, and so on, greater confidence in the results is justified.

Thus, the experimental studies of death qualification effects should be read in conjunction with other, related, studies in which similar effects have been demonstrated in actual cases or in which connections have been made between attitudes about capital punishment and other demographic and attitudinal differences between jurors.[85] Experimental studies of juror comprehension of instructions should be read in conjunction with interviews of jurors who have sat in capital cases.[86] Archival studies of decisionmaking in capital cases should be read in the context of other archival studies as well as the experimental literature.[87] Each study, by itself, is interesting, but it is when a number of studies are read together that a more complete picture of the phenomenon of interest is likely to emerge.

This suggests that it is incumbent upon scientists to strive to build bodies of research that are interconnected using multiple methodologies and conducted by different researchers.[88] While there will al-

84. *See* Edith Greene et al., *supra* note 2, at 247 ("[I]t is a rare individual study that is able to address all of these issues. More importantly, it is not clear that individual studies should address every issue."); Michael J. Saks, *Improving APA Science Translation Amicus Briefs*, 17 LAW & HUM. BEHAV. 235, 240 (1993) (noting the "nearly universal consensus that a single study is not sufficient").

85. *See, e.g.*, Fitzgerald & Ellsworth, *supra* note 33, at 39; Gary Moran & John C. Comfort, *Neither "Tentative" nor "Fragmentary:" Verdict Preference of Impaneled Felony Jurors as a Function of Attitude Toward Capital Punishment*, 71 J. APPLIED PSYCHOL. 146, 152-54 (1986).

86. *See, e.g.*, Eisenberg & Wells, *supra* note 43; Craig Haney et al., *Deciding to Take a Life: Capital Juries, Sentencing Instructions, and the Jurisprudence of Death*, 50 J. SOC. ISSUES 149 (1995); Luginbuhl & Howe, *supra* note 43; A. Reifman et al., *Real Jurors' Understanding of the Law in Real Cases*, 16 LAW & HUM. BEHAV. 539 (1992).

87. For similar studies in other jurisdictions see, for example., DAVID C. BALDUS ET AL., EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS (1990); David C. Baldus et al., *Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview, with Recent Findings from Philadelphia*, 83 CORNELL L. REV. 1638 (1998).

88. It is also incumbent on researchers to conduct well-conceived and executed research. There is some evidence that biased assimilation might be tempered

ways be instances in which legal or policy decisions must be made without a well-developed body of research directly on point, that is not always the case. Moreover, the research community can attempt to anticipate the next important legal and policy issues and begin to build programs of research around those issues.[89] Policymakers, in turn, ought to consider a particular study in the broader context of the literature.

C. Methodological Training

Finally, as noted earlier, in order for members of the legal and policy community to engage in this process of self-consciously scrutinizing empirical research, they must have a basic background in empirical research methodology. There is evidence that training in statistics and research methodology does improve individuals' ability to engage in statistical and methodological reasoning.[90] In particular,

---

somewhat if the studies are difficult to challenge. *See* Rohini Ahluwalia, *Examination of Psychological Processes Underlying Resistance to Persuasion*, 27 J. CONSUMER RESEARCH 217, 222 (2000); Ditto et al., *supra* note 64, at 64. Munro and Ditto found that, "in general, participants showed a good deal of sensitivity to the methodological details of the scientific studies presented" and were "more convinced by the methodologically detailed long descriptions." Munro & Ditto, *supra* note 53, at 649. Pyszczynski and Greenberg argue that information processors have the need to maintain an "illusion of objectivity." Tom Pyszczynski & Jeff Greenberg, *Toward an Integration of Cognitive and Motiviational Perspectives on Social Inference: A Biased Hypothesis-Testing Model*, 20 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 297, 302 (1987). Thus, "[j]udgments seem best characterized as a compromise between the wish to reach a particular conclusion and the plausibility of that conclusion given the available data." Ditto & Lopez, *supra* note 53, at 569; *see also* Ditto et al., *supra* note 64, at 64; B. R. Sherman & A. Kunda, Motivated Evaluation of Scientific Evidence, Paper presented at the American Psychological Society Convention, Arlington, VA (June 1989), *reported in* Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCHOL. BULL. 480, 490 (1990) ("Of importance is that all subjects were also quite responsive to the differential strength of different aspects of the method, which suggests that they were processing the evidence in depth. Threatened subjects did not deny that some aspects were strong, but they did not consider them to be as strong as did nonthreatened subjects. The bias was constrained by plausibility."). Accordingly, the stronger the evidence, the more difficult it is to reject it out of hand.

89. John Darley, *Getting Ahead of the Curve: Anticipating Future Policy Needs in Today's Research*, 15 APS OBSERVER 5 (May/June 2002) (suggesting that psychologists "[t]ake our task as 'getting ahead of the curve,' as identifying what issues will require policy decisions in the next decade, and mobilizing research now to make the discoveries around which polices eventually can be formed.").

90. *See* Geoffrey T. Fong et al., *The Effect of Statistical Training on Thinking about Everyday Problems*, 18 COGNITIVE PSYCHOL. 253, 253 (1986); Richard P. Larrick et al., *Teaching the Use of Cost-Benefit Reasoning in Everyday Life*, 1 PSYCHOL. SCI. 362, 369 (1990); Richard P. Larrick et al., *Who Uses the Cost-Benefit Rules of Choice? Implications for the Normative Status of Microeconomic Theory*, 56 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 331, 333 (1993); Lehman et al., *supra* note 72, at 431; Darrin R. Lehman & Richard E. Nisbett, *A Longitudinal*

mechanisms as varied as undergraduate education in psychology and the social sciences,[91] graduate training in probabilistic sciences such as psychology and medicine,[92] individual statistics courses,[93] and even relatively short, targeted training sessions[94] have been shown to improve individuals' ability to engage in statistical and methodological reasoning.

Kovera and McAuliff found that judges with some scientific background properly judged a study with high internal validity as more likely to be admissible than did judges with no science background. In addition, they judged a flawed study (i.e., one that did not rule out an alternative explanation) as less likely to be admissible than did judges who did not have any background in scientific methods.[95] However, the judges did not differ in their assessments of versions of the study with other methodological flaws.[96]

General training in methodology may be useful in helping legal decisionmakers to identify flawed research. And certainly if one cannot distinguish between well-done and unquestionably flawed research, one is unlikely to be able to engage in the subtle examination of studies that is suggested above. However, the focus here is primarily on biased interpretation of well-done, but inherently imperfect research, and it is less clear that such training will help to ameliorate the effects of biased assimilation. Redding and Reppucci did not find that judges' and law students' background in science had any moderating influence on biased assimilation.[97] In addition, as described above, a number of studies have found that even individuals who have advanced training in science are susceptible to biased assimilation.[98]

However, there are particular (and relatively straightforward) strategies of reasoning that might be fruitfully taught or adopted by evaluators of research. Research exploring the contours of the "biased

---

Study of the Effects of Undergraduate Training on Reasoning, 26 DEVELOPMENTAL PSYCHOL. 952, 959 (1990); Richard E. Nisbett et al., *Teaching Reasoning*, 238 SCIENCE 625, 627-28 (1987).

91. Lehman & Nisbett, *supra* note 90, at 956. And to a somewhat lesser degree undergraduate education in the natural sciences and the humanities. *Id.*
92. Lehman et al., *supra* note 72, at 437.
93. Fong et al., *supra* note 90, at 280.
94. *Id.*
95. Kovera & McAuliff, *supra* note 76, at 580. The Judges' backgrounds in science were based on their undergraduate majors, their graduate science courses, CLEs on scientific methods, and their familiarity with the Federal Judicial Center's *Reference Manual on Scientific Evidence*. *Id.* at 578.
96. *Id.* at 580 (missing control group or the use of research assistant who was not blind to the hypotheses of the study).
97. Redding & Reppucci, *supra* note 54. Background in science was based on the number of empirical social science courses taken, the number of natural science courses taken, and the number of law and social science courses taken in law school. *Id.* at 38.
98. Koehler, *supra* note 53; Mahoney, *supra* note 53.

assimilation" phenomenon has suggested that in evaluating the methodology of an empirical study, it may be useful to adopt a strategy of "considering the opposite." That is, when considering an empirical study, one ought to reflect on how one would have evaluated the study, its methodology, and its persuasiveness if the results had come out differently.

Lord, Lepper, and Preston tested two different mechanisms for debiasing judgments of research quality using the same capital punishment materials as in the original Lord, Ross, and Lepper biased assimilation study.[99] Participants in the debiasing conditions were given either a set of instructions that told them to "be unbiased" by trying to be as objective, fair, and impartial as possible or instructions that directed them to "consider the opposite" by taking into account "whether you would have made the same . . . evaluations had exactly the same study produced results on the *other* side of the issue."[100] Results consistent with biased assimilation were obtained for participants who received instructions identical to those in the original study and for those who received the "be unbiased" instructions; that is, these participants' prior attitudes about the death penalty influenced their evaluations of how well-done and how convincing the studies were. However, this pattern was *not* found for participants who were instructed to "consider the opposite"; that is, the evaluations of these participants were not influenced by their prior attitudes on the topic.[101]

These results suggest that correcting for biased assimilation is more than merely a matter of good faith efforts to be fair and unbiased. Accordingly, efforts to improve the impartiality of those called to evaluate research are unlikely to be successful. Rather, it appears that biased assimilation is the "result of inadequate cognitive strategies."[102] Thus, efforts might be more fruitfully directed so as "to promote an explicit consideration of alternative possibilities, especially those possible outcomes that are diametrically opposed to those expected or perceived."[103]

There are open empirical questions as to how legal actors will respond to more specific training in the scientific method, which types of training are most effective with legal actors, and so on.[104] Whether

99. *See* Lord et al., *supra* note 53.
100. *Id.* at 1233.
101. *Id.* at 1233-34.
102. *Id.* at 1239.
103. *Id.* at 1233.
104. For example, Kovera and McAuliff, *supra* note 76, at 578, included those with undergraduate majors in the natural sciences as part of the group of judges characterized as having some scientific background. However, Lehman et al., *supra* note 72, at 437, found that graduate education in chemistry did not improve methodological and statistical reasoning and Lehman and Nisbett, *supra* note 90,

mechanisms such as reference materials dealing with methodological issues, training sessions targeted at judges or legislators, law school courses, or assistance by experts can improve the evaluation of social science methods by legal and policy decisionmakers is empirically testable.[105] But, certainly, there is no reason to believe that judges, lawyers, and legislators cannot grasp the fundamentals of scientific methodology and strategies such as consideration of the opposite results,[106] and law schools as well as those engaged in policy relevant scientific research ought to facilitate that mastery.[107]

V. CONCLUSION

While I have focused my comments on the Nebraska study of capital punishment and other research on capital punishment decisionmaking, these ideas are much more broadly applicable. Virtually any empirical study that is relevant to law or policy will have relative advantages and disadvantages, and evaluations of the methodological pros and cons can be colored by biased assimilation. Consumers of such empirical research, whether it be on capital punishment or any other legal or policy issue, should be encouraged to step back from their prior beliefs to critically examine the implications of the researchers' methodological choices for any conclusions that might be drawn. Rather than uncritically accepting the results of any study or rejecting them out of hand, policy and legal decisionmakers ought to learn to think deeply about the implications of a study's methods in the broader context of empirical work.

---

at 956, found that improvements in methodological reasoning were greater for psychology and social science students than for students in the natural sciences and humanities.

105. *See, e.g.*, FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (2d ed. 2000); FAIGMAN, *supra* note 8, at 200 (noting training programs by the Federal Judicial Center, the National Judicial College, and the Private Adjudication Center at Duke University).

106. FAIGMAN, *supra* note 8, at 64 (noting that most judges are intelligent and well-educated).

107. *See* Epstein & King, *supra* note 9; Heise, *supra* note 71, at 815. For a casebook addressing these issues in detail see generally DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW: STANDARDS, STATISTICS, AND RESEARCH ISSUES (2002); *see also* Shari Seidman Diamond, *Using Psychology to Control Law: From Deceptive Advertising to Criminal Sentencing*, 13 LAW & HUM. BEHAV. 239, 251 (1989) ("It is up to us to see that the judiciary and other actors in the legal system come to understand the meaning and policy implications of social scientific research. If we do not work to educate legal decision makers so they can distinguish between good research and bad, we cannot complain when courts or legislatures give credence to misleading data or reject our own impeccable research findings.").