

University of Michigan Law School

## University of Michigan Law School Scholarship Repository

---

Articles

Faculty Scholarship

---

2010

### The Inevitability of Theory

Richard O. Lempert

*University of Michigan Law School, rlempert@umich.edu*

Available at: <https://repository.law.umich.edu/articles/2363>

Follow this and additional works at: <https://repository.law.umich.edu/articles>



Part of the [Legal Profession Commons](#), and the [Legal Writing and Research Commons](#)

---

#### Recommended Citation

Lempert, Richard O. "The Inevitability of Theory." *California Law Review* 98, no. 3 (2010): 877-906.

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact [mlaw.repository@umich.edu](mailto:mlaw.repository@umich.edu).

# The Inevitability of Theory

Richard Lempert†

## INTRODUCTION

I wrote this Article in response to an invitation to deliver the keynote address at Berkeley Law School's Jurisprudence and Social Policy conference "*Building Theory Through Empirical Legal Studies*." Lauren Edelman, the intellectual mother of the conference, gently brushed aside my suggestion that I present one of my own attempts to synthesize the results of empirical research to generate theory, and asked that I directly address the conference topic. I am glad that she did.

The first question the conference theme raised for me was what should empirical legal scholars<sup>1</sup> be theorizing? My answer is almost everything law related. The empirical legal studies (ELS) realm includes empirically examining and building theory to explain: extralegal forces that shape legal decision making,<sup>2</sup> people's relations to particular laws<sup>3</sup> and to law in general,<sup>4</sup>

---

Copyright © 2010 California Law Review, Inc. California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the author are solely responsible for the content of their publications.

† Richard Lempert is the University of Michigan's Eric Stein Distinguished University Professor of Law and Sociology, emeritus. He is currently serving as the Chief Scientist and Basic Research Lead in the Human Factors/Behavioral Sciences Division of the Science and Technology Directorate of the Department of Homeland Security. Opinions, factual assertions and other statements in this Article are attributable solely to Professor Lempert and are not those of the Department of Homeland Security.

1. For the most part, I shall use the terms "empirical legal," "ELS," and "sociolegal" to refer not just to the scholars and scholarship associated with the empirical legal studies movement in American law schools but to empirically based sociolegal scholarship whatever the academic background or affiliation of the researcher. This is not only because most of what I will say applies generally to all empirically oriented students of legal phenomena but also because a number of those who participate in the annual ELS conference or who publish or seek to publish in the *Journal of Empirical Legal Studies* have no law school affiliation. Where I mean to refer largely or only to ELS scholars and scholarship as an American law school phenomenon, the context should make this clear.

2. These forces include political values, racial biases, economic self-interest, and other characteristics and commitments that can influence the decisions of judges, juries, and administrative agencies as well as other exercises of legal discretion. *See, e.g.,* Stephanie Lindquist & Pamela Corley, *The Strategies of Judicial Review* (2009) (on file with the *California Law Review*).

3. An individual's social position and personality will influence how a person interprets, is affected by, and responds to particular laws. *See, e.g.,* LAURA BETH NEILSEN, *LICENSE TO*

the public's use of and its treatment by the legal system,<sup>5</sup> the activities of law enforcement agencies,<sup>6</sup> the effects of law and its enforcement,<sup>7</sup> law's influence on the rules and practices of other social institutions,<sup>8</sup> and the behavior of legal institutions such as courts, juries, and administrative agencies,<sup>9</sup> to name just a few targets of sociolegal scholarly attention.

Theory is not always central to this work. Many ELS papers, including some of the most valuable, make their contribution primarily through data

---

HARASS: LAW, HIERARCHY, AND OFFENSIVE PUBLIC SPEECH (2004).

4. Individuals differ in how they conceive of the law and the legal system and what these imply for their own activities and social relationships. *See, e.g.*, PATRICIA EWICK & SUSAN S. SILBEY, *THE COMMON PLACE OF LAW: STORIES FROM EVERYDAY LIFE* (1998).

5. People and organizations differ in the uses to which they put law, their readiness to invoke legal norms and remedies, and in their reactions to attempts to use the law against them. *See, e.g.*, MICHAEL W. McCANN, *RIGHTS AT WORK: PAY EQUITY REFORM AND THE POLITICS OF LEGAL MOBILIZATION* (1994); STUART A. SCHEINGOLD, *THE POLITICS OF RIGHTS: LAWYERS, PUBLIC POLICY, AND POLITICAL CHANGE* (1974); Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 *LAW & SOC'Y REV.* 95 (1974); Stewart Macaulay, *Non-Contractual Relations in Business: A Preliminary Study*, 28 *AM. SOC. REV.* 55 (1963).

6. Legal norms and organizational and ideological variables shape law enforcement routines. *See, e.g.*, RICHARD A. LEO, *POLICE INTERROGATION AND AMERICAN JUSTICE* (2008); SUSAN P. SHAPIRO, *WAYWARD CAPITALISTS: TARGET OF THE SECURITIES AND EXCHANGE COMMISSION* (1984); JEROME H. SKOLNICK, *JUSTICE WITHOUT TRIAL: LAW ENFORCEMENT IN DEMOCRATIC SOCIETY* (3d ed. 1994).

7. Researchers seek to discern whether a law has its apparently intended effects and to identify and understand the mechanisms that mediate the effects of legal commands. *See, e.g.*, GERALD N. ROSENBERG, *THE HOLLOW HOPE: CAN COURTS BRING ABOUT SOCIAL CHANGE?* (2d ed. 2008); LAWRENCE W. SHERMAN, *POLICING DOMESTIC VIOLENCE: EXPERIMENTS AND DILEMMAS* (1992); Laura Nyantung Beny, *Insider Trading Laws and Stock Markets Around the World: An Empirical Contribution to the Theoretical Law and Economics Debate*, 32 *J. CORP. L.* 237 (2007); Robert Weisberg, *The Death Penalty Meets Social Science: Deterrence and Jury Behavior Under New Scrutiny*, 1 *ANN. REV. L. & SOC. SCI.* 151 (2005).

8. Private organizations like businesses and universities often adopt and transform principles like due process that have been first defined by the legal system. *See, e.g.*, FRANK DOBBIN, *INVENTING EQUAL OPPORTUNITY* (2009); Lauren B. Edelman, *Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law*, 97 *AM. J. OF SOC.* 1531–76 (1992); Lauren B. Edelman, *Legal Environments and Organizational Governance: The Expansion of Due Process in the Workplace*, 95 *AM. J. OF SOC.* 1401–40 (1990); Lauren B. Edelman, Howard S. Erlanger & John Lande, *Internal Dispute Resolution: The Transformation of Rights in the Workplace*, 27 *LAW & SOC'Y REV.* 497–534 (1993); Lauren B. Edelman, Christopher Uggen & Howard S. Erlanger, *The Endogeneity of Legal Regulation: Grievance Procedures as Rational Myth*, 105 *AM. J. OF SOC.* 406–454 (1999).

9. Many sociolegal scholars have offered in-depth portraits of specific legal institutions with special attention to the forces that shape them and how they go about their business. *See, e.g.*, JAMES EISENSTEIN & HERBERT JACOB, *FELONY JUSTICE: AN ORGANIZATIONAL ANALYSIS OF CRIMINAL COURTS* (1977); KEITH HAWKINS, *LAW AS LAST RESORT: PROSECUTION DECISION-MAKING IN A REGULATORY AGENCY* (2002); PHILIPPE NONET, *ADMINISTRATIVE JUSTICE: ADVOCACY AND CHANGE IN A GOVERNMENT AGENCY* (1969); H.W. PERRY, JR., *DECIDING TO DECIDE: AGENDA SETTING IN THE UNITED STATES SUPREME COURT* (1991); PHILIP SELZNICK, *TVA AND THE GRASS ROOTS: A STUDY IN THE SOCIOLOGY OF FORMAL ORGANIZATION* (1949); NEIL VIDMAR & VALERIE P. HANS, *AMERICAN JURIES: THE VERDICT* (2007).

collection, organization and description.<sup>10</sup> But even largely descriptive papers usually conclude with a bow to theory and an attempt to explain or theorize what has been found. The only area of the law where ELS scholars<sup>11</sup> seldom venture, and are, for the most part, content to leave the theory building to others, is the interpretation of legal doctrine.<sup>12</sup> This is the enterprise that involves culling statutes, cases, historical records, and other sources to advance a general theory of what cases and statutes mean and/or how the law should be changed or amended.<sup>13</sup> This doctrine-centered approach to understanding law and building legal theory broadly characterizes most legal scholarship. Although it is not the kind of theory building that ELS scholars engage in, it nonetheless has an important empirical component. As I once observed on the ELS blog site:

The most traditional lawyers work with empirical data all the time—case texts. Texts are out there as part of the real world, and from the point of view of empirical scholarship have some substantial virtues. In particular they are transparent (even when opaque), equally accessible to all scholars and open to various methods of interpretation, some of which are better than others. They also often matter. Would that the data “legal empiricists” examined always had

---

10. The work of Theodore Eisenberg is outstanding in this respect. *See, e.g.*, Theodore Eisenberg & Charlotte Lanvers, *What is the Settlement Rate and Why Should We Care?*, 6 J. EMPIRICAL LEGAL STUD. 111 (2009); Theodore Eisenberg, *Use it or Pretenders Will Abuse It: The Importance of Archival Legal Information*, 75 UMKC L. REV. 1 (2006); Theodore Eisenberg et al., *Juries, Judges, and Punitive Damages: Empirical Analyses Using the Civil Justice Survey of State Courts 1992, 1996, and 2001 Data*, 3 J. EMPIRICAL LEGAL STUD. 263 (2006). For other fine examples, see Thomas Y. Davies, *Affirmed: A Study of Criminal Appeals and Decision-Making Norms in a California Court of Appeal*, 1982 AM. B. FOUND. RES. J. 543 (1982); Samuel R. Gross & Kent D. Syverud, *Don't Try: Civil Jury Verdicts in a System Geared to Settlement*, 44 UCLA L. REV. 1 (1996).

11. When I refer to ELS scholars I usually mean to include empirically oriented sociolegal scholars of all disciplinary backgrounds. *See supra* note 1. Many of the latter will see the Law & Society Association (“L&SA”) as their primary associational home, but there is increasing overlap between the attendees at the ELS and L&SA annual meetings and among those who seek to publish in the flagship journals of these two associations. To the extent there are differences in their scholarly style, L&SA members are more likely than those primarily identified with ELS to be concerned with how their work is rooted in and contributes to social theory.

12. ELS scholars are not, however, reluctant to test or challenge understandings of what the law means with empirical data. *See, e.g.*, Daniel Ho, Kevin Quinn & Erica L. Ross, *Did Liberal Justices Invent the Standing Doctrine? An Empirical Study of the Evolution of Standing, 1921–2006*, 62 STAN. L. REV. 591 (2010). Hybrid articles in which the writer begins as a law professor, analyzing cases and interpreting their meaning, then switches to ELS mode and collects data bearing in some way on the law or its effects, and then switches back to law professor mode to argue for a new interpretation of the law are also common. Both these enterprises distinguish ELS scholars from an earlier generation of law and economics scholars who drew on economic theory to argue for or explain particular interpretations of cases and statutes or to advocate particular laws or policies. These scholars too often failed to complement their theoretical grounding with empirically rigorous attempts to look at what was actually happening.

13. *See, e.g.*, Peter Westen, *Confrontation and Compulsory Process: A Unified Theory of Evidence for Criminal Cases*, 91 HARV. L. REV. 567 (1978) (sketching a general theory of the Confrontation and Compulsory Process Clauses).

these virtues.<sup>14</sup>

We should thus be wary of drawing hard and fast lines between traditional legal scholarship and the kind of systematic parsing, assembly, and analysis of textual data that some sociolegal scholars do.<sup>15</sup>

From my perspective then, almost all efforts to construct theories of law and the many things it touches are empirical; so much so that it is natural to ask, “How else can one build theory but through empirical research?” Yet classic theory building is not rigorously empirical. Its model is physics.

## I

### PHYSICS ENVY, OR GRUMBLES ABOUT THEORY

Although empirical information may have inspired physics theorizing—think Newton’s falling apple—the link from empirical evidence to theory building is often loose.<sup>16</sup> Following Thomas Kuhn,<sup>17</sup> the history of many sciences has been understood as one of paradigm shifts. Explanations for observed phenomena are advanced within the parameters of a particular theoretical paradigm, but as our window on the world opens wider, facts that are hard to fit within the dominant paradigm emerge. At some point, the accumulation of anomalous facts makes the inadequacy of the dominant paradigm obvious, and a new paradigm arises to explain everything the prior paradigm could explain, as well as the facts it could not make sense of. The new paradigm then dominates understanding and theory construction until so many new anomalies have arisen that a new overarching theory is needed to reconcile them.

Scientific theories always seek to make sense of the world “out there,” but data may or may not be collected with theory in mind. In physics, unlike much social science, formal theory usually guides the search for data. Albert Einstein, for example, developed his general theory of relativity in an effort to extend the concept to include gravity in a way that was mathematically consistent with the special theory he had developed some years before.<sup>18</sup> The key empirical data

---

14. Posting of Richard Lempert to Empirical Legal Studies Blog, [http://www.elsblog.org/the\\_empirical\\_legal\\_studi/2006/08/day\\_2\\_\\_\\_whats\\_i.html](http://www.elsblog.org/the_empirical_legal_studi/2006/08/day_2___whats_i.html) (Aug. 8, 2006, 00:17 EST).

15. See, e.g., KIM LANE SCHEPPELE, *LEGAL SECRETS: EQUALITY AND EFFICIENCY IN THE COMMON LAW* (1988).

16. I am thinking here primarily of those branches of modern physics that seek to understand the nature of matter, basic forces like gravity and the origin of the cosmos. If, however, theory in physics is often driven by the mathematical implications of relationships, physicists have never lost sight of the importance of testing their theories empirically and revising them in the light of new empirical information. However much confidence they may have in theoretical derivations, physicists as a community properly regard theoretical advances as provisional until confirmed by empirical data. Hence, numbers of Nobel Prizes in physics have gone to experimentalists, and high energy physicists and astrophysicists have successfully lobbied for billions of dollars of investments in telescopes and ever more powerful colliders.

17. THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* (3d ed. 1996).

18. BARRY R. PARKER, *EINSTEIN’S BRAINCHILD: RELATIVITY MADE RELATIVELY EASY!*

indicating Einstein had “got it right” was Arthur Stanley Eddington’s evidence that light was affected by the sun’s gravity. Eddington reported that light bent exactly as Einstein had predicted, when measured during a solar eclipse.<sup>19</sup> Eddington, however, was not examining the behavior of light to build a theory, but rather to test one. His search would not have occurred had Einstein’s theory not existed.<sup>20</sup> Indeed, in modern theoretical physics even the anomalies to be reconciled are often not empirical; rather, they are mathematical inconsistencies or gaps in received theory. Thus, billions of dollars are being spent not to generate theory, but, as in the search for the Higgs boson, to confirm or call into question a theoretical view that is regarded as likely but unproven.

Social scientists are not physicists, and people are much harder to understand and model than even questionably existent subatomic particles. Not only is human behavior context dependent, but people have minds of their own and can behave in unexpected ways. Hence, we should not expect that efforts to understand people and their groups, institutions and social creations (like the law), will be amenable to mapping in such strict mathematical terms that the math itself will reveal gaps in understanding and suggest probable gap-filling explanations.<sup>21</sup>

Further complications arise because not only are most social science theories true only in a statistical sense,<sup>22</sup> but also, even if central tendencies matter, it is often behavior occurring in the tails of distributions that is of most concern. People and groups at the extremes of human distributions (violent terrorists, for example) are usually not there just because of bad luck or other random factors. Many will differ from more typical actors in systematic, caused ways. Efforts to understand the sources of these differences are confounded both because extreme behavior is rare and because actors may exhibit the same extreme behavior for different reasons. One suicide bomber may, for example, be motivated by ideological commitments and another by a desire to avenge a sibling’s death.<sup>23</sup>

---

85–140 (2000).

19. It has been suggested that Eddington’s report of precise coincidence with Einstein’s prediction was not an accurate characterization of his data even though Einstein’s prediction was correct. If so, Eddington’s report indicates a danger of beginning research with deeply held theory; one may see data as more closely aligned with theoretical predictions than is in fact the case. *See* JOHN WALLER, *FABULOUS SCIENCE: FACT AND FICTION IN THE HISTORY OF SCIENTIFIC DISCOVERY* 57–64 (2004).

20. *See generally* WALTER ISAACSON, *EINSTEIN: HIS LIFE AND UNIVERSE* (2007).

21. This, however, does not mean that social science methods and models cannot provide reliable predictions. They can, particularly if the effort is to predict aggregate behaviors like voting in an election or responses to economic incentives.

22. Many physical science theories also allow for only stochastic prediction, but unlike the concerns of social science, stochastic uncertainty does not have obviously important effects at the level at which humans live.

23. In recent years, both social science research and the theories built on it are changing, as increased computing power allows better modeling of complexity and emergent relationships. Most empirical research to date on law and in the social sciences has been either qualitative or

### *A. Grand Theory*

Despite these differences between the social sciences and the physical and natural sciences, the physics model of the primacy of theory and the celebrity status accorded its theoreticians has had substantial spillover effects within the social sciences that only in recent years have subsided.

When I began studying sociology in the 1960s, the most celebrated sociologist in the country was Talcott Parsons, with Robert Merton not far behind. Neither, however, offered a theory that would have been recognized as such by a physicist, or by a chemist, geologist, or biologist for that matter. What Parsons and Merton offered was not so much theory as perspectives for making sense of much that was going on in social life. Their “theories” were not silly, but they lacked precision. In particular, it was hard to draw from these and similar social science “theories” hypotheses that made it possible to conduct rigorous empirical testing.<sup>24</sup> Even operationalizing core theoretical concepts was difficult and open to dispute. Moreover, to the extent these theories had empirical support, their predictions characterized only a portion of cases examined, and, unlike physics, these theories were never adequately captured in the language of mathematics. The same was true of other theoretical perspectives in vogue when I was a student, including Weberian theory,<sup>25</sup> Durkheimian theory,<sup>26</sup> and Marxist theory<sup>27</sup>—theories still invoked, sometimes with continued awe, almost fifty years later. It is also true of grander theories advanced since then, including World Systems Theory,<sup>28</sup> or to give a

based on variations of regression models. Today, new kinds of models, like agent-based models and network models, are gaining prominence. Theory is often incorporated at the micro level and expectations about social behavior and clustering are allowed to emerge, including behavior at the tails of distributions. If emergent patterns are much like what has been known to have happened, the validity of theorized starting points is likely presumed, and model results that project beyond the known to the unknown may be offered as empirical projections. I do not know where these efforts will lead. They may well transform social science research on legal phenomena. I am, however, confident they will not displace the need for good theory and sound empirical research.

24. In fairness, data availability and quality and the statistical methods of the day constituted substantial barriers to rigorous quantitative empirical testing.

25. RANDALL COLLINS, *WEBERIAN SOCIOLOGICAL THEORY* (1986); MAX WEBER, *ON LAW IN ECONOMY AND SOCIETY* (Edward Shils trans., Max Rheinstein ed. & trans., Harvard University Press 1954) (1925); MAX WEBER, *THE PROTESTANT ETHIC AND THE SPIRIT OF CAPITALISM*, (Talcott Parsons trans., Charles Scribner's Sons 1958) (1905).

26. ÉMILE DURKHEIM, *THE DIVISION OF LABOUR IN SOCIETY* (W.D. Halls trans., The Free Press 1984) (1893); ÉMILE DURKHEIM, *RULES OF SOCIOLOGICAL METHOD* (W.D. Halls trans., The Free Press 1982) (1895); ÉMILE DURKHEIM, *SUICIDE: A STUDY IN SOCIOLOGY*, (John A. Spaulding trans., George Simpson ed. & trans., The Free Press 1951) (1897); WHITNEY POPE, *DURKHEIM'S SUICIDE: A CLASSIC ANALYZED* (1976).

27. ÉTIENNE BALIBAR, *THE PHILOSOPHY OF MARX* (Chris Turner trans., Verso 1995); *A DICTIONARY OF MARXIST THOUGHT* (Tom Bottomore, ed., Blackwell 2d ed., 1991); KARL MARX, *CAPITAL: A CRITIQUE OF POLITICAL ECONOMY* (Frederick Engels ed., 1867); EUGENY B. PASHUKANIS, *LAW AND MARXISM: A GENERAL THEORY* (Barbara Einhorn trans., 1978).

28. TERENCE K. HOPKINS & IMMANUEL WALLERSTEIN, *WORLD-SYSTEMS ANALYSIS: THEORY AND METHODOLOGY* (1982).

sociolegal example, Donald Black's "pure sociological" theory of the behavior of law.<sup>29</sup>

Works by these authors and by others writing in the grand theoretical tradition are often admirable for their scholarship—it is, for example, difficult even today to read Weber and not be struck by his erudition, his attempt to deeply root his ideas in empirical fact, and the continued relevance of many of his observations about legal, social and organizational life. Nevertheless, Weber did not construct theories that can be, on the one hand, definitively rejected or, on the other hand, made much more probable by the analysis of empirical data. Perhaps because it is difficult to characterize Weberian theory in formal terms, his ideas have neither been fully cast aside nor fully incorporated into later theoretical work.

### B. Middle-Range Theory

It is not just grand theories that have these qualities. Middle-range theories, such as actor-network-theory,<sup>30</sup> resource mobilization theory,<sup>31</sup> labeling theory,<sup>32</sup> differential association theory,<sup>33</sup> normalization theory,<sup>34</sup> and new social movement theory,<sup>35</sup> are all similar in their resistance to rigorous formalization, uncertainties in testing their predictions, the inconsistency with which they seem to hold true, and the difficulty of encompassing them within larger, more general theories. These middle-range theories can nonetheless be insightful. Although they seldom allow precise predictions, they help make sense, in a narrative way, of particular behaviors or realms of social action, and they have provided the theoretical underpinnings for many empirical studies. Yet, as theory they are ultimately unsatisfying. When different theories are offered to explain the same terrain, empirical evidence seldom forces the

29. DONALD BLACK, *THE BEHAVIOR OF LAW* (1976).

30. BRUNO LATOUR, *REASSEMBLING THE SOCIAL: AN INTRODUCTION TO ACTOR-NETWORK-THEORY* (2005); JOHN LAW & JOHN HASSARD, *ACTOR NETWORK THEORY AND AFTER* (1999).

31. *THE DYNAMICS OF SOCIAL MOVEMENTS: RESOURCE MOBILIZATION, SOCIAL CONTROL, AND TACTICS* (Mayer N. Zald and John D. McCarthy eds., 1979); John D. McCarthy & Mayer N. Zald, *Resource Mobilization Theory: Vigorous or Outmoded?*, in *HANDBOOK OF SOCIOLOGICAL THEORY* 533 (Jonathan H. Turner ed., 2001).

32. HOWARD S. BECKER, *OUTSIDERS: STUDIES IN THE SOCIOLOGY OF DEVIANCE* (1963); EDWIN M. LEMERT, *HUMAN DEVIANCE, SOCIAL PROBLEMS, AND SOCIAL CONTROL* (1967); EDWIN M. SCHUR, *LABELING DEVIANT BEHAVIOR: ITS SOCIOLOGICAL IMPLICATIONS* (1971).

33. DONALD R. CRESSEY, *DELINQUENCY, CRIME AND DIFFERENTIAL ASSOCIATION* (1964); Ross L. Matsueda, *The Current State of Differential Association Theory*, 34 *CRIME & DELINQ.* 277 (1988).

34. Carl May & Tracy Finch, *Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory*, 43 *SOC.* 535 (2009); Carl R. May et al., *Development of a Theory of Implementation and Integration: Normalization Process Theory*, 4 *IMPLEMENTATION SCI.* 29 (2009).

35. *NEW SOCIAL MOVEMENTS: FROM IDEOLOGY TO IDENTITY* (Enrique Laraña et al. eds., 1994); Steven M. Buechler, *New Social Movement Theories*, 36 *SOC. Q.* 441, (1995).

rejection of one in favor of another. Nor do these middle-range theories seem capable of being assimilated into some larger, more complete theoretical framework to give us a more adequate theory of a broader range of social phenomena.<sup>36</sup>

### C. *Micro-Level Theory*

The most specific and in some ways the most successful social science theories are closely tied to narrow studies and aggregates of narrow studies. "Micro-level" theory-building studies come in two flavors. Some focus narrowly on hypotheses implied by larger theories. Others aim at advancing new theory to explain specific, well-defined behavior.<sup>37</sup> Theory at this level forms the spine of articles rather than of books.

### D. *Testing Theories*

Although the first flavor of micro-level research may resemble Eddington's in that a specific hypothesis is drawn from a larger theory, operationalized, and tested, even when this is true such tests seldom go to the heart of what the larger theory is about. Hence it is almost always wrong to say that the larger theory has been supported by or failed a *crucial* test. Typically only a narrow theoretical proposition drawn from the broader theory is substantially supported, and effective tests of the larger implications of sociolegal theories are rare, despite the claims that authors sometimes make. Not only is it difficult to unambiguously identify small-scale implications of larger theories, but even when this can be done it may be difficult, if not impossible, to collect data that reliably represent core theoretical concepts.

Further problems arise because theories are often stated in more or less absolute terms while empirical results, even if they are generally consistent with theoretical expectations, inevitably contain cases and/or variables that do not behave as theory would predict. These difficulties can be seen in efforts by sociolegal scholars to put some of the field's grand theories to the test, such as sociological attempts to use Human Relations Area File data to test Durkheim's ideas about the relationship between the quality of legal institutions and social organization,<sup>38</sup> and a number of efforts to test the theory Black advanced in the

---

36. Empirical evidence is not, however, irrelevant to their persistence, for middle-range theories come into and go out of fashion, and an accumulation of consistent or conflicting evidence, when it exists, can be an important influence on this flux. Nonempirical factors such as ideological preferences, narrative fit to prominent exemplars, and the intellectual attractions of novel ideas can also influence the attention given different explanations.

37. No theory can be reliably established by a single study, and few if any areas of social science investigation explore unstudied phenomena. Thus, authors who do specific studies and contribute to theory on what I call the micro level typically situate their work within an existing literature and their contributions gain credibility from consistency with what others have found.

38. Steven F. Messner, *Societal Development, Social Equality, and Homicide: A Cross-National Test of a Durkheimian Model*, 61 SOC. FORCES 225 (1982); Richard D. Schwartz &

*Behavior of Law* by empirically examining predictions derived from it.<sup>39</sup> The tests of Durkheim's theory were thwarted by insufficient variance in the developmental stages of Area Files societies and too much variance in these societies' other characteristics. Research on Black's theory, whether conflicting or supportive, was unconvincing because the settings used as test beds for the theory were too situation specific to fairly test Black's sweeping propositions.<sup>40</sup>

Articles that claim to test hypotheses derived from middle-range or micro-level theories, especially those that advance their own theory and then purportedly test and find support for it, often have other problems. Such articles typically begin by summarizing or advancing a theory and by rehearsing the evidence for or against it. The instant study is then situated as a test of that theory or an extension of it. Next the author sets out hypotheses implied by the theory followed by a description of the data and methods that enable hypothesis testing. Finally, the author reports the results of the empirical analysis, more often than not confirming many, if not all, of the specified hypotheses. Frequently however, especially in the case of novel theories, it is hard to avoid the suspicion that the research path was more inductive than the author claims, for it is easy to believe that the empirical results did more to shape the theoretical discussion than the theory did to guide the empirical investigation. This feeling is particularly strong when an article advances a novel theory or

---

James C. Miller, *Legal Evolution and Societal Complexity*, 70 AM. J. SOC. 159 (1964); Stephen D. Webb, *Crime and the Division of Labor: Testing a Durkheimian Model*, 78 AM. J. SOC. 643 (1972).

39. Michael R. Gottfredson & Michael J. Hindelang, *A Study of the Behavior of Law*, 44 AM. SOC. REV. 3 (1979); Candace Kruttschnitt, *Social Status and Sentences of Female Offenders*, 15 LAW & SOC'Y REV. 247 (1980–1981).

40. I should perhaps say the research was unconvincing to me. Black wrote his book to invite such testing, and if the theory is taken as he presents it, the tests are valid because Black argues that his theory applies at all levels—from dyads to whole societies—where there are legal relationships. Taking this claim literally, Black's theory should be rejected because a number of studies yield results wholly or partially inconsistent with Black's theoretical predictions. But to investigate the implications of Black's law-like propositions such as “downward law is greater than upward law” in specific settings (for example, bail setting in one jurisdiction) provides a poor test of Black's general theory. Such tests are weak because of difficulties in operationalizing the most theoretically relevant variables, the limited range of variation in test settings, and the frequent presence of unmeasured confounding variables. Yet even if Black's theory did not explain the behavior of law in a particular setting, it might nonetheless be generally true in a statistical sense across many settings, and, in particular, it might explain how law was invoked and applied at the macro level.

My core difficulty with Black's theory, as presented originally and as subsequently elaborated, is that it is far too immodestly presented. By purporting to explain how law is applied at all levels and in all circumstances, the theory overreaches and explains nothing. But treating the theory as a set of propositions that are predicted to hold true statistically at more macro levels, Black makes a contribution. He advances a number of propositions that are generally true of law, and their claimed general truth is not effectively threatened by the failure of some studies to substantiate particular derived predictions. Seen in this light, however, the theory suffers from other problems. It cannot explain why it is true only generally and why there are deviations from its predictions that seem to result from something other than random variation. Not only does it fail to illuminate the mechanisms behind its predicted relationships, it does not seek to do so.

theoretical wrinkle.

## II

### THE INEVITABILITY OF THEORY

Those who conclude from this introduction that I am no great fan of much of what passes for theory in the social sciences are correct. Too often, “theory” is an academic’s word for “explanation,” and too rarely do social scientists other than economists attempt to formalize theory in mathematical terms or to create more encompassing empirically testable formulations. Moreover, except to some extent in economics, efforts to build more formal theory have seldom been a solution. Social life is so complex, so unsystematically studied, and so fraught with data acquisition and measurement problems that there are huge gaps in the knowledge needed for realistic formalization. Where formal theory has been attempted, as in some work in law and economics, the incorporation of unproven or implausible assumptions has led to often-justified suspicion of the purported real-world implications of the theory advanced.

Nonetheless, I am not prepared to write off theory or the theoretical enterprise. Indeed, I fault some empirical work for its ignorance of theory or its acknowledgement of theory with the scholarly equivalent of a nod and a wink. Not only does theory permeate everything an empirical researcher does, whether acknowledged or not, but it is also the basis from which social scientists make policy recommendations and is essential in assessing the likely reliability of empirical results.

All empirical scholarship is theoretically informed, at least in the weak sense that problem selection, model construction, and even the information captured in qualitative research reflect expectations about what matters. These expectations are the products of theories.

Consider research on the legal profession. If theory tells us that gender should influence the job choices of young lawyers, we would not attempt to model those choices without controlling for or measuring the effects of gender. Similarly, if theory suggests that income should increase with experience but at a declining rate, we would include a curvilinear component in our measure of earnings. In short, whatever complaints one might make about the imprecision, inconsistency, over- and under-inclusion, pretentiousness, incompleteness, or limited applicability of social science theories, the answer is not to ignore or dismiss the importance of theory. Rather, it is to combine attention to theory with empirical investigation in ways that will add to our knowledge and understanding.

#### *A. Important Distinctions*

##### *1. Small “t” or Large “T”*

In thinking about how theory informs, and can be tested by, empirical

investigation, some distinctions are fundamental. The most important is between theory (with a small “t”) and Theory (with a capital “T”).<sup>41</sup> Researchers always approach empirical investigations with small “t” theories, that is to say, expectations about how the world works and what is needed to understand phenomena of interest. Quantitative researchers begin their inquiries with a sense of what variables might explain the outcomes they are investigating, and they seek to include them as explanatory or control variables in their models. Qualitative researchers similarly begin their inquiries with a sense of which behaviors are important to observe and what makes for trustworthy informants. These sensibilities reflect both implicit and explicit theories about how the world works. They have many sources, including: (1) common sense (gender affects so many relationships that I should check to see if it affects this one); (2) prior findings (Jones and Smith found that CEO gender affected company share price, so my model of share price movements should control for it); (3) generalized knowledge (since women make different career choices than men, I should control for gender when exploring how law school prestige relates to job choices); and (4) knowledge of large “T” Theory (theories of role strain suggest that women with children will respond differently to time demands than single women or married women without children; therefore, I must control not just for gender but also for the presence of children in explaining lawyer job choices).

The inevitability of theory in these ways need not, however, lead to what I would call “theoretically informed” investigation; that is, research designed to yield results that can be expected to either strengthen a social science theory or call it into question. An investigation should be called theoretically informed only when theory guides research in one of two ways. The first requires social science (or large “T”) Theories to have implications clear enough to allow the derivation of specific hypotheses that can be operationalized to guide investigation. The second occurs when a researcher posits a new theory or an extension of an existing theory *before* data collection and/or analysis begins and derives from the new theory hypotheses that test it. In limiting the concept “theoretically informed” to these circumstances, I do not mean to say that theory cannot be derived or supported inductively in light of what data reveal. But other things being equal, theoretically consistent results support a theory more persuasively if they have been specified in advance.

## 2. *A Priori or Post Hoc*

The distinction between theory as a priori expectation and theory as post hoc explanation matters. When a theory and hypotheses drawn from it are specified prior to research and analysis, explanations in terms of the theory are

---

41. For examples of capital “T” theory see the examples of grand and middle-range theories cited in *supra* notes 25–35.

more credible than when the results of a completed analysis are given a theoretical interpretation. This is true not just of overarching explanatory theory but also of the theories that guide research design.

There are two principal reasons why this is so. First, people are good at finding patterns, even among randomly associated variables,<sup>42</sup> and they are also good at plausibly explaining the patterns they find. Post hoc, it is almost always possible to generate or find some theoretical explanation for empirical results. Second, with enough bites at the data-analytic apple, it is usually possible to find some order in social science data, even if that order exists only by chance. Introducing and omitting control variables, changing variable measurement, and exploring different functional forms may yield results that suggest nonchance relationships when other measurement and model choices would show no associations.<sup>43</sup>

The dangers of post hoc model specification and theorizing are particularly great—and, I suspect, too often realized—in litigation and policy-oriented research where the aim is to support a party’s position or a funder’s policy preferences. These dangers exist, however, whenever a researcher has an interest in advancing or supporting a theory or conclusion. Quantitative researchers often recognize (and criticize) as a problem with qualitative research the possibility that the researcher found not the truth, but what she went looking for. But the danger of a consciously or unconsciously “rigged” analysis is not only shared by quantitative research, it is often more insidious because the influence of potential bias is less transparent.

### 3. *A Personal Example*

To avoid suspicion of their results, litigation- and policy-oriented researchers should take special care to ensure that their choice of data and methods do not favor the outcome they desire.<sup>44</sup> This concern affected, for

---

42. For example, I found this number sequence in a book I was leafing through: 03, 07, 11 . . . . What number comes next? Most people would say 15. But the next number in the text was 20. The sequence came from a portion of a table of random numbers as reproduced in HUBERT M. BLALOCK, *SOCIAL STATISTIC* 437 (1st ed. 1960). Cf. NASSIM NICHOLAS TALEB, *FOOLED BY RANDOMNESS: THE HIDDEN ROLE OF CHANCE IN LIFE AND IN THE MARKETS* (2005) (discusses the way investors and others are fooled by the human tendency to causally interpret random variation).

43. The situation is analogous to the widely recognized problem of significance testing in many variable models. If enough tests are done of enough variables, one can expect some significant relationships to emerge even if all associations are in fact random. There are ways to correct statistically for this possibility, but in epidemiology and in some of the social sciences they are seldom employed. Similarly, if data are measured and examined in enough different ways, it is likely that some approach to the data will yield apparent nonchance relationships even if no such relationships exist. The number of different ways a researcher has examined her data is less visible than the number of variables in a model, and even if known, I know of no way to statistically correct for it.

44. This is not too much to ask of policy-oriented academic researchers, but so long as we have our current adversary system, one cannot expect most litigation researchers to avoid

example, how David Chambers, Terry Adams and I analyzed the data from our study of how Michigan Law School's minority students fared post graduation.<sup>45</sup> We were not concerned with testing theory.<sup>46</sup> Instead we wanted to know whether Michigan's minority students had successful post-law-school careers and how their career success compared to that of Michigan's white alumni. We knew that our work might figure in the politically divisive debate on affirmative action, and because all of us supported affirmative action, we sought to foreclose charges that we had rigged our analysis to favor outcomes that accorded with our values. We also sought to guard against even the unconscious temptation to find results that favored our preferences. Hence, we did something I do not, as a rule, recommend: we avoided almost all preliminary data analysis.<sup>47</sup> We specified our variables and how we would measure them before we looked at the data, and we reported in footnotes all relevant results of sensitivity testing. Still it is natural that some people will suspect our results, particularly our failure to reject the null hypothesis that there is no difference in the career success of Michigan's minority and white graduates—a result that many people (myself included) find surprising. Unfortunately, we have no way of refuting those who will not believe our assertions of how we proceeded.<sup>48</sup>

On this last point, theory could help. To the extent that our results are consistent with well-supported theory, they become more credible.<sup>49</sup> Our work,

---

choosing data and methods likely to support preferred results. These experts are hired with the goal of proving certain facts, and their findings will not be presented unless they support the case of the party that hired them. Even worse, litigators who do not welcome the results of one commissioned study have been known to commission others and to keep the existence of the prior results from those who do the later research.

45. Richard Lempert, David L. Chambers & Terry K. Adams, *Michigan's Minority Graduates in Practice: The River Runs Through Law School*, 25 LAW & SOC. INQUIRY 395 (2000).

46. *Id.* In retrospect, this was to the study's detriment. Had we begun the research with a theoretical as well as an empirical agenda, we would have collected far more data on social class, among other things.

47. We did examine regression models with variables we did not include in our final models, eliminating variables if they were both insignificant and of no apparent theoretical importance. We also added a variable to our model based on a comment received.

48. One way to alleviate suspicion, which we have done, is to make data available to others with different preferences. A problem exists, however, because the same factors that allow an unethical researcher to make data and model choices that appear to lend strong support to a favored hypothesis allow a would-be critic to find apparent relationships in data that call a prior analysis into question.

49. Other empirical research helps as well. Thus, our results gain credibility from their consistency with scholarship in the area, including smaller scale studies of minority physicians. See WILLIAM G. BOWEN & DEREK BOK, *THE SHAPE OF THE RIVER* (1998); Robert Davidson & Ernest Lewis, *Affirmative Action and Other Special Consideration Admissions at the University of California, Davis, School of Medicine*, 278 J. AM. MED. ASS'N 1153 (1997); Stephen Keith et al., *Effects of Affirmative Action in Medical School: A Study of the Class of 1975*, 313 NEW ENG. J. MED. 1519 (1985); Miriam Komaromy et al., *The Role of Black and Hispanic Physicians in Providing Health Care for Underserved Populations*, 334 NEW ENG. J. MED. 1305 (1996); Ernest

however, was not driven by theory in the larger sense, even if our choice of variables and models was motivated by small “t” theoretical understandings of how outcomes relate to success and what variables might influence outcomes. Our contributions to larger theoretical perspectives came in two forms: (1) puzzling results, which, if confirmed in other work, merit more theoretically focused study,<sup>50</sup> and (2) findings that support (but do not prove) theories that elevate law school prestige above most other variables in explaining the career success of law school graduates.

#### 4. *Limits of an Example*

It should be clear from this brief reference to my own work that I am neither arguing that empirical research must be theory driven in the capital “T” sense nor am I suggesting that the hypotheses a researcher poses for testing need have such theoretical roots. I am also not suggesting that quantitative researchers should avoid exploratory data analysis (EDA). My views are to the contrary.<sup>51</sup> EDA is a good way of presenting information about data distribution and comparing data distributions over time or in the context of other variables.<sup>52</sup> EDA also helps to determine what functional form best suits a data set, when collapsing variables into indices makes sense, whether assumptions of particular methods are met, and the most revealing way to distinguish variable states, among other things.<sup>53</sup> Thus, I am not opposed to

---

Moy & Barbara A. Bartman, *Physician Race and Care of Minority and Medically Indigent Patients*, 273 J. AM. MED. ASS’N 1515 (1995).

50. For example, we found to our surprise that although an admissions index combining LSAT scores and undergraduate grade-point averages predicted law school grades, and although law school grades predicted later income, there was no relationship between the admission index and later income even in models where the index might have served as a proxy for law school grades. Lempert et al., *supra* note 45, at 401–02.

51. I have long admired the path-breaking work of John Tukey and his arguments for EDA. See JOHN W. TUKEY, *EXPLORATORY DATA ANALYSIS* (1977). EDA, as introduced by Tukey, involves different relatively simple ways of examining data to check for, capture, and portray patterns with minimal loss of information, usually as a prelude to more complex analysis. Other forms of EDA test whether the assumptions of different approaches to data analysis are in fact met. One must, however, be cautious in choosing models based on what preliminary analysis reveals about apparent relationships, for when variables are dropped because they are insignificant in a preliminary analysis, standard errors on the remaining variables may be distorted. One solution, if the size of the data set allows, is to make model selection decisions based on the analysis of a random subsample of cases and then test the model on the remaining data. Only the results of the test should be considered reliable. For an in-depth discussion of this issue and an illustration of model selection dangers see Richard Berk, Lawrence Brown & Linda Zhao, *Statistical Inference After Model Selection*, 26 J. QUANT. CRIM. 217 (2009).

52. Box and whisker plots, first advanced by Tukey as a form of EDA, are now a standard way of presenting information about data midpoints and spreads in an easy to understand fashion. *Id.* at 39–43.

53. The Michigan study, I expect, would have been better had we engaged in more EDA, including exploring different alternatives for examining our data. Lempert, *supra* note 45. But as already noted, the political sensitivity of our subject matter coupled with our views on affirmative action counseled against looking closely at the data before deciding what relationships to examine

extensive EDA, unmoored from hypothesis testing. Indeed, the clear presentation of variable distributions and simple patterns of associations among variables are often the most revealing products of empirical investigation.<sup>54</sup> Similarly, I am not denying a place in social science research for empirical analysis uninformed by larger theory. In many areas of sociolegal scholarship, larger theories from which testable hypotheses can be derived do not exist, either because a matter is under-theorized or because of insurmountable model and/or data problems. These gaps should be reasons to do more rather than less research on a question. Indeed, curiosity with no expectation of what will turn up is itself a sufficient driver of research. Accumulating information about how the world works, whatever the starting point or motivation, is often what calls existing theory into question and stimulates new theoretical insights. This is true in every empirical discipline.

### *E. Dustbowl Empiricism*

There is, however, one mode of empirically driven data analysis that can seriously mislead and seldom has much to offer. This is to work in ways that are not only uninformed by capital “T” theory, but that also minimize the contributions that small “t” theories make to research design. Rather than formulate specific hypotheses or pare down potential variables to a parsimonious model based on theoretical expectations, some empirical studies include every available variable that can conceivably explain a dependent variable and then let a computer choose what seems important.<sup>55</sup> The problem with this kind of “dustbowl” or “brute” empiricism is that it risks identifying as significant relationships that only exist by chance<sup>56</sup> or which, if statistically significant, are nonetheless too weak to be of either theoretical or practical importance. Moreover, by including variables that proxy for each other, true relationships may be hidden. Hence it is not surprising that the results of brute empiricism are often puzzling. Some statistically significant variables may perform as established theory, common sense, or general knowledge would suggest, but other including some that are statistically significant may have unexpected magnitudes or signs. Even worse, the models that reach print may have been substantially pruned to exclude most of the variables that proved

---

and what methods to use. Moreover, although our study began without litigation on the horizon, a suit had been brought against the law school’s affirmative action program by the time we had reached the data analysis stage. If I were called upon to testify about the study’s results, I wanted to be able to tell a cross-examiner that we had not looked for likely relationships before commencing our analysis.

54. *See supra* note 10; *see also* EDWARD R. TUFTE, *THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION* (2d. ed. 2001) (providing numerous examples of how collecting and creatively organizing data can reveal otherwise difficult to appreciate relationships).

55. This approach commonly relies on stepwise regression models available in most widely used statistical packages.

56. *See supra* note 43.

insignificant, and this pruning is not always transparent. When such work appears theory-guided, empirical results may be accorded more weight than they deserve.

### *1. Ensuring Process Transparency*

My mom-and-apple-pie message here is that regardless of how data have been analyzed and the role theory has played in the analysis, truth and transparency should characterize data presentation. When different ways of coding or analyzing data have yielded different outcomes, the researcher should disclose this, even if she is confident that the analysis she did for publication is the scientifically best way of proceeding.<sup>57</sup> Similarly, research that did not begin as an effort to test theoretically derived hypotheses should not present itself as if this were its pedigree. When a researcher perceives an obvious or best theoretical explanation for results after an analysis has been done, the *post hoc* nature of the explanation should be clear, and the support given the theory not exaggerated.

Humility is also appropriate, for it is easy to overstate the importance of what one has found. Theory is not built from a single study. If a theory is well established, results that only in retrospect appear explicable by the theory are unlikely to do much to make the theory more plausible than it was before the investigation. If new theory is offered to explain results, considerable additional evidence will ordinarily be required for the theory to move from candidate status to likely explanation.<sup>58</sup>

---

57. Readers are justified in presuming that if approaches tried and discarded did not have some justification in the first instance, they would not have been attempted. Moreover, information about what did not work and why it did not work may itself be of scientific interest.

58. This will depend to some extent on the nature of the findings and the theory offered to explain them. Mark Granovetter famously explained why successful job searches tended to owe more to leads provided by people with narrow and discrete links to the job seeker than to leads provided by those with more extensive and more intimate ties. Granovetter's article was almost immediately regarded as reliable theory for two reasons. First, Granovetter had an intuitively plausible explanation for his results: weakly tied people collectively linked the job seeker to a wide array of diverse opportunities while intimates tended to know the same people and possibilities as the job seeker and so provided less added value to the search. Second, Granovetter's theory could not only be represented formally but the intuitive explanation is consistent with what one would expect from network models. For a summary of early empirical tests of the model. For a summary of the original article and a discussion of later research, see Mark Granovetter, *The Strength of Weak Ties: A Network Theory Revisited*, 1 SOC. THEORY 201 (1983).

### III USES OF THEORY<sup>59</sup>

Theory is not just theoretical; it is useful. It is essential to the scientific understanding of social phenomena, to focusing empirical investigation, to evaluating research, and to making responsible policy prescription.

#### *A. Focusing Empirical Investigation*

Theories embody the social science understanding of the phenomena we investigate. Anyone, however, can advance a theory, and implausible theories abound. To provide validated social science understandings, theories must be tested against data and withstand those tests. A virtue of good theories is that they make it easy to derive appropriate tests.<sup>60</sup>

##### *1. An Example from Durkheim*

Arthur Stinchcombe, in his marvelous primer *Constructing Social Theories*,<sup>61</sup> draws on Emile Durkheim's *Suicide*<sup>62</sup> for an iconic example of the logic of scientific inference. The example illustrates how implications derived from well-specified theory allow convincing real-world tests.

Stinchcombe begins by restating the core of Durkheim's theory of egoistic suicide: "A higher degree of individualism in a social group causes a higher rate of suicide in that group."<sup>63</sup> Durkheim drew from this simple theory and his knowledge of French and European society specific expectations about what data would reveal. Protestants, Durkheim argued, were less involved in their churches than Catholics and consequently should have higher suicide rates. Married men with children were linked more tightly to others than bachelors, so marrying and raising a family should diminish the rate of male suicides. Jews in Durkheim's time were forced to cohere, so their suicide rates should be low as well.

Data supported these and other theory-consistent expectations. This left Durkheim justifiably confident that he was onto something. Particularly important to Durkheim's project was the data on suicide among Jews. Durkheim, Stinchcombe tells us, had found that urban location, commercial employment, and education, were all positively associated with suicide. These results were consistent with his theory since these variables were more than likely associated with greater individualism. But commercial employment and

59. In this section I shall be referring only to causal theories. This is, of course, the sense of theory that has predominated in this Article, but I have also referred to theory of at least two other types: theories of meaning (as in the theory of the case), and theories of being (as in the theories of the atom).

60. Whether the tests are easy to accomplish is another matter.

61. ARTHUR L. STINCHCOMBE, *CONSTRUCTING SOCIAL THEORIES* 15–28 (1987).

62. DURKHEIM, *SUICIDE*, *supra* note 26.

63. STINCHCOMBE, *supra* note 61, at 15.

greater education were themselves associated with urban residence, and it was possible that something about urban life other than the lack of social attachments that Durkheim posited increased suicidal propensities. If this was the case, then French Jews who were disproportionately urban, well educated, and engaged in commerce should have suicide rates much like their similarly situated neighbors. But if individualism was the causal mechanism, these markers of urban life would not affect Jews in the way they affected other Frenchmen since they did not undercut solidarity within the Jewish community. When the data revealed that Jews committed suicide at far lower rates than other commercially engaged urban dwellers, Durkheim's theory received substantial support.<sup>64</sup>

## 2. *An Example from Ellickson*

Another example, more familiar to sociological scholars, is Robert Ellickson's investigation into the real-world validity of the Coase Theorem.<sup>65</sup> Ellickson chose to investigate how liability for damages done to or by wandering cattle was apportioned in a California county. The genius of this study was that it looked empirically at a situation that Ronald Coase had used hypothetically in a seminal article to illustrate what later became known as the Coase Theorem.<sup>66</sup> Ellickson found that the Coase Theorem did not hold as an empirical matter. Contrary to the Coasean expectation, the initial placement of liability for damage done by roving cattle affected the final asset distribution. This finding did not directly contradict anything Coase's Theorem had postulated because Coase had made it clear from the start that his theorem presumed a nonexistent world of zero transaction costs. Ellickson's empiricism does, however, undermine policy recommendations justified in part by reference to Coase because it revealed that Coase, unlike Durkheim, had not produced a valid *social* theory. Although as a matter of economic logic the theory may be impeccable, it does not, if Ellickson's results can be trusted, advance our understanding of human behavior.<sup>67</sup> Indeed, even in a zero-

---

64. Durkheim's theory of suicide is a bit more complicated than this. Durkheim recognized that not all suicides could be explained by the forces that led to egoistic suicide. Thus his book defines three other types of suicide: (1) altruistic suicide in which the causes of egoistic suicide are turned on their head, (2) anomic suicide, and (3) fatalistic suicide.

65. ROBERT C. ELICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* (1991).

66. Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 2-8 (1960). Coase's claim (which George Stigler rather than Coase first denominated a theorem) is that when trade is possible and there are no transaction costs, the initial allocation of property rights and liability for their infringement does not matter because bargaining will lead to an efficient outcome regardless of the initial allocation.

67. In fairness to Coase, there is no reason to believe he ever intended his theory to do this. The theory's implication for understanding social transactions is not that the placement of liability burdens does not matter, but that to understand how a situation will resolve itself one must focus on transaction costs. Theorists in law and economics nonetheless used the Coase Theorem to derive policy recommendations without having determined the likely burdens of transaction costs.

transaction-cost world, it is not clear that people would behave as Coase predicts.<sup>68</sup>

Drawing empirical implications from theories is often the easy part. Determining empirically if the implications hold is the more difficult task. Theoretical concepts and predictions must be linked to real-world situations and behavior. This requires finding ways to measure the real-world manifestations of those concepts and predictions<sup>69</sup> and devising ways to assess whether theoretical expectations hold.

Durkheim, despite the primitive state of social science methods in his day, took on a manageable challenge. The theoretical concept of suicide maps directly onto actual suicides, and France had good data on suicide rates. Individualism, which for Durkheim referenced the degree to which an individual's behavior is self directed rather than shaped by group norms, is more ambiguously related to observable conditions. Nevertheless, there was substantial consensus in Durkheim's France that Protestants as a group were less closely regulated by church norms than Catholics or Jews, and that bachelors had generally shallower relations with others than family men. But in other societies connections like these might have been debatable, and had Durkheim's altruistic suicide<sup>70</sup> been as common in France as egoistic suicide, aggregate suicide rates would not have enabled a test of the theory. Whether it is easy or difficult to test a theory is, however, beside the current point, which is that an important virtue of theory is that it focuses empirical investigation.

---

Ellickson's work indicates that this is no easy matter, for the economic costs associated with doing business did not drive the deviations between the Coasean prediction that liability placement would be of no consequence and how people actually behaved. Rather, the relevant transaction costs were social in nature. They were the costs of violating norms and disrupting social relationships.

68. See generally Donald H. Regan, *The Problem of Social Cost Revisited*, 15 J.L. & ECON. 427 (1972) (theoretical questioning). Empirically, game theory experiments indicate that rational actor model predictions do not routinely hold, even in situations where their assumptions appear fully met. Cf. Gary Charness & Uri Gneezy, *What's in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games*, 68 J. ECON. BEHAV. & ORG. 29 (2003) (Finding that when players in the dictator game knew the names of those they were playing against they were more generous, but this did not hold for players in the ultimatum game.); Carolyn H. Declerck, Toko Kiyonari & Christophe Boone, *Why Do Responders Reject Unequal Offers in the Ultimatum Game? An Experimental Study on the Role of Perceiving Interdependence*, 30 J. ECON. PSYCHOL. 335 (2009) (finding that contrary to expectations, interdependence did not affect proposer behavior, although it spurred responders to be more rational); Andrew Schotter & Barry Sopher, *Advice and Behavior in Intergenerational Ultimatum Games: An Experimental Approach*, 58 GAMES & ECON. BEHAVIOR 365 (2007) (finding that the advice of earlier participant in an ultimatum game affects the behavior of replacement participant and that fairness justifications are used to justify offer rejections by receivers).

69. As Ellickson's work indicates, the assessment need not involve quantification; interviewing informants is one way of determining what is happening and why. Ellickson, *supra* note 65, at 6–9.

70. Altruistic suicide is the opposite of egoistic suicide in that it is motivated by such close involvement with a group that one is willing to die for the group or in support of its norms. Suicide bombing is a current example.

### *B. Evaluating Empirical Research*

A second virtue of theory is its value in evaluating empirical research, sometimes in overlooked ways. Theory provides a vantage point for assessing the adequacy of models and the validity of conclusions derived from modeling efforts. If a model omits or poorly measures theoretically important variables, or if some modeled relationships make no theoretical sense, then there is good reason to view skeptically claims based on the model. This is true even if core results are consistent with reasonable theoretical claims. There is also reason to be skeptical of a study's results if outcomes differ when an analysis is altered in theoretically irrelevant ways. In short, one should approach empirical studies with the following expectations: (1) Expect robustness to theory-irrelevant variation. (2) Expect to find theoretically relevant variables included in models or legitimate reasons for their absence. (3) Expect results for variables that have been introduced into a model as controls to make theoretical sense. Research on capital punishment and right-to-carry laws nicely illustrates these points.

#### *1. A Death Penalty Example*

In 1975, the economist Isaac Ehrlich published a time series study of the deterrent effects of capital punishment. Ehrlich analyzed data from 1933 to 1969 using what were at the time advanced techniques of statistical modeling.<sup>71</sup> His results indicated that one execution deterred seven or eight homicides. This was attention grabbing, to say the least, particularly since earlier studies had, with rare exceptions, found no evidence that capital punishment deterred homicides and in some cases suggested brutalization effects.

Ehrlich's work reshaped the empirical debate on capital punishment in three ways. First, it changed the focus of the debate from the deterrent effects of having a death penalty to the deterrent effect of executing convicted murderers. Second, it both marked and encouraged a shift in the methods used to search for deterrence. Before Ehrlich, the most influential studies either compared homicide rates in states that seemed similar except for the presence of the death penalty or looked within states for changes in homicide rates following the adoption or abolition of the death penalty.<sup>72</sup> After Ehrlich, the search for deterrence almost always relies on complex regression models that attempt to control for factors, apart from the death penalty, that might be expected to influence homicide rates.<sup>73</sup> Third, before Ehrlich wrote, the

---

71. Isaac Ehrlich, *The Deterrent Effect of Capital Punishment: A Question of Life and Death*, 65 AM. ECON. REV. 397 (1975).

72. See Richard Lempert, *Desert and Deterrence: An Assessment of the Moral Bases of the Case for Capital Punishment*, 79 MICH. L. REV. 1177, 1197–1206 (1981) (review of studies by Thorsten Sellin, Isaac Ehrlich, and others).

73. Earlier, apparently simpler, studies were not blind to the need for controls. Most eschewed statistical modeling in favor of using the proximity of states or in-state longitudinal designs to control for factors apart from capital punishment that might influence homicide rates. The adequacy of these controls is questionable, but so is a researcher's capacity to specifically

deterrence argument—because of its inconsistency with most extant research—had ceased to be an important element of the contemporary case for capital punishment.<sup>74</sup> After Ehrlich, claims that deterrence justified the death penalty had, for a while at least, more credibility.<sup>75</sup>

Ehrlich's results were immediately controversial, and soon other social scientists reexamined Ehrlich's data using similar statistical approaches.<sup>76</sup> Their efforts revealed that Ehrlich's finding of deterrence was acutely sensitive to the years he chose to analyze. If the last years of the time series—the 1960s—were removed from the analysis, evidence for a deterrent effect disappeared, and in some analyses a possible brutalization effect emerged.<sup>77</sup> During the 1960s, executions for murder dropped almost to zero, and rates for many crimes, including crimes never punishable by death, rapidly increased. It is thus likely that during the 1960s factors other than low execution rates were driving up murder rates since it is plausible to suppose that murders are affected by social

measure and correctly model all factors that substantially affect homicide rates. For a more extended discussion of these issues and a sense that the difference in the quality of controls may not be as great as it might seem, see Richard Lempert, *The Effect of Executions on Homicides: A New Look in an Old Light*, 29 CRIME & DELINQ. 88 (1983).

74. Indeed, it is plausible to think that the failure of earlier research to find that the death penalty deterred and the suggestion that it might brutalize helped promote the focus on retributivism as the primary justification for punishment in general and the death penalty in particular. See, e.g., WALTER BERNS, FOR CAPITAL PUNISHMENT: CRIME AND THE MORALITY OF THE DEATH PENALTY (1979).

75. The most immediate indicator of Ehrlich's influence came soon after his publication, when the Supreme Court, in reinstating the death penalty, refused to consider the implications of its deterrent effect for their decision, noting that the research on deterrent effects was inconsistent. Ehrlich's research was the only research suggesting deterrence that the Court cited. *Gregg v. Georgia*, 428 U.S. 153, 184 n.31 (1976). Responses to Ehrlich, for reasons discussed below, soon muted the deterrence case for the death penalty, but beginning around the mid 1990s a spate of econometric research examining the post moratorium use of the death penalty reinvigorated the deterrence argument to the point where one noted liberal suggested that deterrence might not just provide an adequate moral justification for the death penalty but might also make it morally mandatory. See Cass Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required?: Acts, Omissions, and Life-Life Trade-Offs*, 58 STAN. L. REV. 703, (2005) (citing scholarship on the efficacy of capital punishment). But recent critical research is calling the latest deterrence findings into question just as Ehrlich's eventually were. See John J. Donohue & Justin Wolfers, *Uses and Abuses of Empirical Evidence in the Death Penalty Debate*, 58 STAN. L. REV. 791 (2005); see also Richard Berk, *New Claims About Executions and General Deterrence: Déjà Vu All over Again?*, 2 J. EMPIRICAL LEGAL STUD. 303 (2005); Jeffrey Fagan, Franklin E. Zimring & Amanda Geller, *Capital Punishment and Capital Murder: Market Share and the Deterrent Effects of the Death Penalty*, 84 TEX. L. REV. 1803 (2006). Professor Sunstein later made it clear that his discussion of the morality of the death penalty did not mean that he accepted the evidence indicating that capital punishment deterred homicides. Rather in a column written with Justin Wolfers, one of the prominent critics of the "new deterrence" literature, Sunstein endorsed the view that, "the best reading of the accumulated data is that they do not establish a deterrent effect of the death penalty." Cass R. Sunstein & Justin Wolfers, *A Death Penalty Puzzle: The Murky Evidence for and Against Deterrence*, WASH. POST, June 30, 2008, at A11.

76. For a detailed discussion of Ehrlich's study and the work published in response to it see Lempert, *Desert and Deterrence*, *supra* note 72.

77. See, e.g., William J. Bowers & Glenn L. Pierce, *Deterrence or Brutalization: What is the Effect of Execution?*, 26 CRIME & DELINQ. 453 (1980).

forces in the same way as armed robberies, rapes and other violent crimes. Crime rates in the 1960s, coupled with the failure to replicate Ehrlich's results with data from different time periods, destroyed his study's credibility.

This brief tale may seem to recount a case where one analyst's findings were trumped by contrary findings of many other scholars. But to the best of my recollection, none of the competing analysts hit the nail on the head when it came to explaining why their work refuted Ehrlich's deterrence claim. It is at least possible that the high homicide rates in the 1960s were caused by the scarcity of executions, while increases in other violent crimes had other causes. The reason why this possibility cannot save Ehrlich's argument lies in theory.

Ehrlich's research was not aimed at estimating how many lives each execution saved. Rather, as Ehrlich described his effort, it was designed to see whether deterrence theory could explain homicide rates. An implication of that theory, which accords well with common sense, is that the more likely a person is to be arrested for and convicted of murder, and the more likely he is to be executed if he is arrested and convicted, the less likely he is to kill someone. Ehrlich's execution results gained credibility because they were not alone in their fit with deterrence theory. Consistent with theory and with research on other crimes, Ehrlich found that apprehension probabilities (likelihood of being arrested for and convicted of murder) were more important to deterrence than punishment (execution) probabilities given apprehension. So far, so good. However, nothing about deterrence theory suggests that executions would have one effect from 1937 through 1960 and a different effect from 1937 through 1967. Rather, deterrence theory implies that, other things being equal, the effects of executions on homicides should be the same over the entire period, as well as over any part of it.<sup>78</sup> What destroys Ehrlich's claim that executions deter homicides is not the inconsistency between Ehrlich's results and the results of other analysts who looked at executions over different time periods. Rather, the flaw is that if Ehrlich's theory that executions deter homicide is valid, then there is no good reason to expect that deterrence will disappear when a theoretically irrelevant parameter, like time period, is varied. The existence of such unexpected variance, together with evidence that the 1960s was a time when all crime rates were rising, deservedly discredits Ehrlich's results even if, like the Cheshire Cat's grin, they have long remained in the air.

---

78. Statistically it could be more difficult to spot deterrence over subsets of the period studied because of the decreased number of data points, but this statistical issue does not explain the failure to find deterrence by those who omitted the last years of the time series from their data since they had enough data points for analysis. Ehrlich's control variables were intended to ensure that other factors affecting homicide rates were held constant throughout the period, and while they were no doubt inadequate to ensure this, neither Ehrlich nor any of his defenders offered plausible reasons why if the death penalty deterred during the period 1937-67 it should not also deter during the period 1937-60.

A second article of Ehrlich's illustrates another way in which theoretical sensitivity can attune one to the problematic nature of statistical conclusions.<sup>79</sup> Here the focus is on what theory tells us about model construction. Two years after his initial foray into the death penalty-deterrence debate, Ehrlich published a second study where he again found that executions deterred homicides. Using a cross-sectional rather than a longitudinal design, Ehrlich compared homicide rates across states rather than national data over time.

Although it never received the attention of Ehrlich's first effort, the second study is in some ways the stronger study because it does not depend on the crime rise in the 1960s. The model Ehrlich uses suffers, however, from a glaring omission: it does not distinguish Southern from other states. Yet for the time periods Ehrlich studied there were good reasons to believe that Southern states differed from other states in their homicide rates and in their use of the death penalty, and that these differences were not captured by standard control variables.<sup>80</sup> Prior cross-sectional studies of homicide and other violent crime rates invariably included a dummy variable for states of the former confederacy, and the coefficient on that variable was almost always significant. Ehrlich offered no explanation for deviating from the prior practice. Results from studies that ignore plausible, available control variables are deservedly regarded with suspicion.<sup>81</sup>

## 2. A "Right to Carry" Example

Even if an analysis appears to support the theory it was designed to test,

79. Isaac Ehrlich, *Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence*, 85 J. POL. ECON. 741 (1977).

80. RICHARD E. NISBETT & DOV COHEN, CULTURE OF HONOR: THE PSYCHOLOGY OF VIOLENCE IN THE SOUTH (1996); Dov Coehn, Richard E. Nisbett, Brian F. Bowdye & Norbert Schwartz, *Insult, Aggression, and the Southern Culture of Honor: An "Experimental Ethnography,"* 5 J. PER & SOC. PSYCHOL. 595-60 (1996) (finding that when university students were bumped or insulted, students from the South were more likely than those from the North to become upset and show priming for aggression).

81. Ehrlich's two studies were also suspect because they yielded conflicting estimates of the number of lives each execution saved. In the longitudinal study the estimate was seven or eight lives and in the cross-sectional study the estimate was about eighteen. There is no obvious theoretical or statistical reason to explain the inconsistency in the number of lives saved if the two models were simply different ways of measuring the same deterrent effect. Although states differ in many ways that might be relevant to crime rates, as aggregates they are unlikely to be that different in factors apart from the death penalty that lead to homicide. If each execution really deterred as many as eighteen murders (and some studies published around the same time as Ehrlich's put the number of murders deterred per execution as high as a hundred and fifty!) one might expect some sign of deterrence even in uncontrolled comparisons. See, e.g., James A. Yunker, *Is the Death Penalty a Deterrent to Homicide? Some Time Series Evidence*, 5 J. BEHAVIORAL ECON. 45, 65 (1976). Yet the work of Sellin and others who had looked at homicide rates in neighboring states controlling for time did not find a whisper of deterrence. See Lempert, *Desert and Deterrence*, *supra* note 72. Surely, if each execution saved eighteen lives something would have been whispering, and if executions so strongly deterred, one would expect that the existence of death penalty statutes would by themselves have some deterrent effect even if actual executions were rare, but not even Ehrlich claimed that they did.

one should be suspicious if it also reports results that are theoretically inexplicable or incoherent. John Lott and David Mustard's widely reported finding that right-to-carry laws diminish violent crime provides an example.<sup>82</sup> Lott and Mustard's hypothesis, that people would be less likely to attack others if they had reason to fear that those they assaulted were armed, is reasonable and worth testing. If all Lott and Mustard had found was that, as they predicted, right-to-carry laws were associated with lower murder, rape, and aggravated assault rates, their results would initially have had to be taken seriously. But the co-authors also found that right-to-carry laws were associated with *higher* rates of nonviolent property crimes. This makes no theoretical sense.<sup>83</sup> Other empirical implications of the Lott-Mustard model, like the suggestion in the data that reducing the number of black women over forty would have substantial crime reduction effects, seem not only theoretically inexplicable but weird to the point of being incredible.<sup>84</sup> Given these peculiarities, it is not surprising that others who have examined their research call the Lott-Mustard findings into question.<sup>85</sup> If some of the implications of an empirical analysis make no sense or, worse yet, are obviously false, absent a very good explanation for why this happened, the safest conclusion is that there is

---

82. John R. Lott, Jr. & David B. Mustard, *Crime, Deterrence, and Right-to-Carry Concealed Handguns*, 26 J. LEGAL STUD. 1, 24–25 (1997).

83. The authors would disagree with this judgment. They argued that different types of crimes are substitutes for one another, and that as violent crimes are deterred from fear of meeting someone with a gun, crimes that involve nonconfrontational thefts will be substituted for them. But as I have pointed out, a theory can be constructed to fit almost any data pattern. When authors provide a *post hoc* explanation for a counter-intuitive result, one should be cautious in accepting it. Lott and Mustard's theory was borrowed from economics where it often helps make sense of purchasing decisions, career choices, and the like. It may even sensibly explain choices criminals make between some crimes; for example, efforts to shut down the market for marijuana in the 1970s may have stimulated the importation of cocaine. But the idea that stealthily stealing from another might substitute for crimes like rape or murder ignores what we know about differences between these crimes, the motives for them, and those who commit them. See generally JACK KATZ, *SEDUCTIONS OF CRIME: MORAL AND SENSUAL ATTRACTIONS IN DOING EVIL* 52–79, 274–309 (1988).

84. Reading this sentence, my Department of Homeland Security colleague Rik Legault alerted me to research suggesting a possible explanation for this result, albeit one that runs directly counter to the Lott and Mustard thesis. Data indicate that black women in this demographic are particularly likely to own handguns, largely because of their fear of crime. David J. Bordua & Alan J. Lizotte, *Patterns of Legal Firearms Ownership: A Cultural and Situational Analysis of Illinois Counties*, 1 LAW & POL'Y Q. 147, 159–60 (1979). If these guns are particularly likely to be borrowed or stolen, then black female handgun ownership could proxy for the greater ease of hand gun acquisition by people planning violent encounters or other crimes.

85. See Ian Ayres & John J. Donohue III, *Shooting Down the "More Guns, Less Crime" Hypothesis*, 55 STAN. L. REV. 1193 (2003); Dan A. Black & Daniel S. Nagin, *Do Right-to-Carry Laws Deter Violent Crime?*, 27 J. LEGAL STUD. 209 (1998); Hashem Dezhbakhsh & Paul H. Rubin, *Lives Saved or Lives Lost?: The Effects of Concealed-Handgun Laws on Crime*, 88 AM. ECON. REV. 468 (1998); Jens Ludwig, *Concealed-Gun-Carrying Laws and Violent Crime: Evidence from State Panel Data*, 18 INT'L REV. LAW & ECON. 239 (1998); see also Robert A. Martin, Jr. & Richard L. Legault, *Systematic Measurement Error with State-Level Crime Data: Evidence from the "More Guns, Less Crime" Debate*, 42 J. RES. CRIME & DELINQ. 187 (2005).

something wrong with the model, methods, or data. In such cases, even those study outcomes consistent with plausible theory cannot be trusted.

### *C. Prescribing Policy*

A third use for theory lies in its relationship to policy prescription. Deriving wise policy from empirical research almost always requires considerable confidence in some theoretical perspective. It is this confidence that justifies the expectation that particular policies will have desired results. If the circumstances in which policies were implemented always resembled closely the contexts in which empirical results were derived, and if the empirical exercise were repeated in different settings with similar results, one might prescribe policy on the supposition that what worked in the research setting would work everywhere. Understanding why a treatment worked would not matter.<sup>86</sup> This may on occasion be true of engineering or medical interventions, but it is seldom if ever true of social policies.

The scale on which policies are implemented is usually far greater than the scales on which research has been done. Implementation contexts vary from study contexts in numerous ways, including, usually, time frame and location. In laboratory research, such as experimental law and economics and mock jury studies, the study situation differs in obvious and arguably important ways from the real world. Moreover, the nature of legal policy is such that one size usually must fit all, yet different people may react in different ways to the same policy prescription.

A well-supported theory provides a coherent explanation for why variables produce the effects they do and the conditions in which relationships hold. The explanation will identify those causes that are most important in determining outcomes and how causes relate to outcomes controlling for context. Knowing these relationships allows policy makers and implementers to anticipate problems and to adapt policies wisely to differing situations. These benefits are important enough that we should seldom generalize directly from empirical results to policy prescriptions. Without the intermediate step of theory development, empirical investigators, legal or otherwise, are seldom if ever justified in promoting policy changes or other interventions based on what their data reveal. Theory resting on a bedrock of rigorous empiricism is the key to sound, evidence-based social policy. One study, however well-designed, does not a bedrock make.<sup>87</sup>

---

86. For example, aspirin was used to relieve headaches and fevers long before science provided any theoretical explanation for its association with these effects.

87. One might argue that evidence-based medicine generalizes directly from large-scale clinical trials to what works and what does not work without any need for theory as an intermediary. To some extent this is true, but medical researchers generalizing from treatment effects have a number of advantages over sociolegal researchers generalizing from social science research. Evidence-based medical decisions are rooted in studies that are rigorously controlled and/or involve large populations followed over time. In addition, those studied are, as a group,

*D. Undertheorized Empirics, or Beware the Single Study*

More than two decades ago, Lawrence Sherman and Richard Berk published an article in the *American Sociological Review* exploring the effects of arrest on recidivist spouse abuse.<sup>88</sup> Their work, and the policies that it led to, indicate the dangers of advocating and adopting policies by generalizing directly from an empirical study rather than through well-constructed theory. This is so even when the study is of high quality.

The Sherman and Berk study reports results from a well-designed and effectively implemented field experiment complemented by sophisticated statistical evaluation. The researchers persuaded the Minneapolis Police Department to agree that when police were called to the scene of misdemeanor spouse abuse they would randomly choose whether to arrest, separate, or merely lecture those men whose quarrels with wives or girlfriends had escalated into threatened or actual violence. Outcomes were measured in two ways: (1) subsequent spouse abuse arrests, and (2) follow-up interviews with alleged victims to learn if abuse had persisted. Although the outcome measures were not in perfect accord, they separately indicated that mandatory arrest (followed by at least a night in jail but usually no further punishment) was associated with lower recidivism rates than the other experimental treatments.

If ever one social science study might justify policy prescription, this one seems about as good a candidate as one might find, and it did, in fact, lead to policy change. The study's obvious quality, efforts made by Sherman to publicize its results,<sup>89</sup> and the nice fit between the study's apparent results and

---

likely to be quite similar in biologically relevant ways to the people, as a group, who will receive the clinically proven treatments. Clinical trials have misled when the experimental subjects differed in clinically relevant ways from those later treated, as when trial subjects were almost all males and women were in the treated population. Moreover, it is a mistake to think that theory plays no role in evidence-based medicine. Often there will be strong theoretical reasons for thinking a drug will have certain effects and reasons to believe a treatment will work through well-understood pathways. The empirical evidence gathering is designed to validate expectations theory has generated. When theoretical understandings are inadequate there is often, despite strong experimental designs, a price to be paid. Poorly understood side effects may exist, or a treatment may not work for, or may even cause harm, to some patients. In the usual instance, the likelihood of benefit to a user or the costs of not treating will be so high, relative to harmful side effects or the costs of a failure to treat, that if a treatment is validated in a robust clinical trial it will be allowed on the market. However, there are medicines that work well for some people but never are marketed because their untoward effects are prevalent enough and serious enough to justify withholding the drug from all. With better theoretical understanding of the mechanisms by which such drugs work, they might be given to those who would benefit from them and kept from persons they would harm. This realization has spurred genomic medicine.

88. Lawrence W. Sherman & Richard A. Berk, *The Specific Deterrent Effects of Arrest for Domestic Assault*, 49 AM. SOC. REV. 261 (1984). The presentation and discussion of their research draws heavily on material I presented on the ELS Blog on August 6, 2006. See Lempert, *supra* note 14.

89. See, e.g., Lawrence W. Sherman & Ellen G. Cohn, *The Impact of Research on Legal Policy: The Minneapolis Domestic Violence Experiment*, 23 LAW & SOC'Y REV. 117 (1989) (describing efforts made to publicize the study and the spread of its influence). *But see* Richard

the agendas of both the conservative law-and-order movement and the liberal women's rights movement combined to make it immediately influential. Police departments across the country adopted policies in spouse abuse cases of presumptive or mandatory arrest.

The Sherman and Berk study was, however, as theoretically weak as it was methodologically strong. Although presented as a study of deterrence theory, its results were not tightly linked to theory. Beyond presuming deterrence, no attention was paid to the mechanism by which arrest had its effects. The authors never examined their data with an eye to understanding why some men who had been arrested offended again or why some men who had been randomly assigned to the less drastic treatments nonetheless sinned no more. Also, the authors never addressed inconsistencies between their measures of recidivism,<sup>90</sup> and they presented but ignored data suggesting that arrest may have prevented recidivism by disrupting relationships rather than by discouraging men from abusing women.<sup>91</sup>

The National Institute of Justice, which had funded the original research and which was getting great credit for it, then made a decision as valuable and exemplary as it is rare. They funded five replications. These replications painted a more nuanced picture than the original study—one with much murkier policy implications.<sup>92</sup> Lawrence Sherman, conducted one of the best replications. He explored the effects of mandatory arrest in Milwaukee, a city with a population that was more racially diverse than Minneapolis's and less homogenous in terms of socio-economic status.<sup>93</sup> The Milwaukee data revealed that the effects of arrest depended on who was arrested. Men who had a lot to lose from arrest<sup>94</sup> (a group that was disproportionately white, well educated and with good jobs) seemed to be deterred by arrest in that they were less likely to offend again than those who received less severe treatments. However, men with less to lose (a group that was disproportionately poor, black and poorly employed if they worked at all) became, if anything, more violent toward their spouses or girlfriends after experiencing arrest.

The simple deterrence hypothesis does not fit the data, but a more complex understanding of mechanism, consistent with the deterrence hypotheses and with the results of several of the other replications, offers a more adequate explanation. Deterrence theory postulates that the greater the

---

Lempert, *Humility is a Virtue: On the Publicization of Policy-Relevant Research*, 23 LAW & SOC'Y REV. 145 (1989).

90. Using one measure, separation was the next most effective treatment to arrest, and using the other measure, breaking up the fight without separating the parties was next most effective. Sherman & Berk, *supra* note 88, at 267.

91. *Id.*

92. See *Symposium on Domestic Violence*, 83 J. CRIM. L. & CRIMINOLOGY 1 (1992).

93. SHERMAN, *supra* note 7.

94. These losses might be tangible such as a lost job or wages or intangible such as a debased reputation.

likely cost of a behavior the less likely it will occur. Since an arrest is presumably a more costly consequence of police intervention than a separation or a lecture, one would expect an arrest to most dampen future abuse. The Milwaukee replication and others suggest, however, that to consider only the fact of arrest is a mistake. The consequences of an arrest appear contingent on demographic and other variables, including the period of confinement following the arrest and the threat of prosecution.<sup>95</sup> The contingent nature of an arrest's effects clouds rather than simplifies policy recommendations. Do we want to adopt an arrest policy that might better protect middle class white women while perhaps endangering women who are lower class and black or Hispanic?

Moreover, our knowledge of why an arrest seems sometimes to discourage abuse and sometimes to increase it is still far from complete. It is easy to understand why fear of arrest would discourage violence, but why should arrest in some circumstances increase the chance that a man will be a repeat abuser? One possibility is that arrest creates greater anger than other treatments and that residual anger, even months after an arrest has occurred, increases the probability of abuse more than the threat of another arrest decreases it. Another possibility is that arrest has additional consequences, like job loss, which strain relationships with spouses and that these strains increase the likelihood of repeat abuse. Without well-supported theory, this research cannot tell us how best to respond to spousal abuse of various types, involving different types of abusers in diverse situations.

The premature influence of the Sherman and Berk study reinforces a core belief of mine: *it is always unwise to base important policy changes on a single social science study and, in most cases, on a small number of them.* No matter how well done a particular study is, important theoretical gaps will remain, and filling those gaps is seldom easy. A quarter of a century has passed since Sherman and Berk's path-breaking study, and we still lack a theory of what causes and restrains spousal abuse adequate for policy prescription.<sup>96</sup>

If it is important to attend to theory in doing and evaluating empirical research, it is even more important to acknowledge the primacy of theory in making policy recommendations.<sup>97</sup> Theory is what allows us to generalize from

---

95. In Minneapolis arrested men usually were released after a night in jail and nothing further happened to them. In Milwaukee men often spent more time in jail and prosecution might follow.

96. Policy, of course, must be made and implemented whether we fully understand the policy's dynamics and context. Moreover, even incomplete theories and scattered studies may provide a better guide to policy than untutored intuition. I am thus not advocating ignoring social science research and theories that are only partially grounded. Rather, I am advocating caution in adopting policies with limited research support even if the research that exists appears impressive. In particular, being able to point to a study or two that leads in a particular direction will not make most difficult policy choices easy. Nor should a few studies be treated as trump when their implications conflict with other values.

97. The idea that if something works in one setting it will work in others is itself a theory,

a group of empirical studies to the larger universe in which policies are implemented, and by shedding light on mechanism, a good theory can distinguish between settings where a policy is likely to work as planned and settings where it may have no effects or even backfire.

#### IV

##### A CAUTION ON THEORY

There is, however, also a danger of bending too far in favor of theory. A strong theory can mislead by unduly shaping perceptions. It may lead us to see what is not there, or our confidence in theory may be so great that we see no need for (further) empirical evidence even when gaps in understanding remain. Some authorities considering mandatory arrest for spousal abuse may not have asked whether their cities were like Minneapolis because of their faith in deterrence theory. Legal scholarship is not immune to this phenomenon. Early scholarship in law and economics is replete with policy recommendations built on empirically unproven economic theory and little more.<sup>98</sup> Rational actor models were used to predict behavior even though people's behavior regularly deviates from rational actor expectations,<sup>99</sup> and first best analyses were offered as models for practical solutions even though it was widely known that more closely approaching a first best policy is not necessarily the optimum strategy in a second best world.<sup>100</sup> It is, in short, a mistake to think that policymakers

---

and it is sometimes the only theory offered to justify the leap from empirical study to public policy or a generalized conclusion. This is not always a mistake. If eating a mushroom kills the first ten people who try it, we have good reason to believe the eleventh taster will be killed as well, even if we have no idea why the mushroom is deadly. Similarly, if we were to reliably find that the death penalty deterred homicides in all states that have it, there is good reason to believe that it would have a similar effect in the abolitionist states if they changed their laws. But usually, as with the Sherman and Berk study, there is too great a distance between the conditions of even well done research and the settings in which its findings might be applied, and too great an interaction between treatment and context to justify a social policy merely because a particular treatment worked in a research setting. Before taking the leap from research to action we should understand theoretically why an effect is expected, or we should be able to conclude from large-scale or repeat testing that an effect is likely to be invariant across the settings to which we will generalize results.

98. One classic and still influential example is: HENRY G. MANNE, *INSIDER TRADING AND THE STOCK MARKET* (1966). For an early book full of examples, see RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* (1st ed. 1972).

99. See *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* (Daniel Kahneman et al, eds. 1982); RICHARD NISBETT & LEE ROSS, *HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT* (1980); Mark G. Kelman, *Misunderstanding Social Life: A Critique of the Core Premises of "Law and Economics,"* 33 J. LEGAL EDUC. 274 (1983).

100. See R.G. Lipsey & Kelvin Lancaster, *The General Theory of Second Best*, 24 REV. ECON. STUD. 11 (1956); see also RICHARD S. MARKOVITS, *TRUTH OR ECONOMICS: ON THE DEFINITION, PREDICTION, AND RELEVANCE OF ECONOMIC EFFICIENCY* (2008). A first best policy imagines the optimal solution to a problem and assumes that the closer a policy comes to the optimal the better the solution is, but in the real world it doesn't necessarily follow that coming closer to overall optimality yields better results than more distant solutions. For example suppose the safest and most efficient parameters for a curve in a highway would be an embankment of

should choose between reliance on theory and reliance on empirical investigation. The relationship between empirical research and theories of behavior is reciprocal. Each offers most when it is informed by and informs the other. If theory is the foot that I suggest the policy maker set out with, it is at least as important that the empirical foot also be there to stand on.

#### CONCLUSION

Sociology is not physics; neither is political science, psychology, anthropology, or economics. We may have a good theory of the atom, but there are no empirically validated general theories of human or organizational behavior. Most social science theories resist formalization, and where they are formalized, as is common in economics, they often incorporate empirically questionable assumptions, meaning their correspondence with what occurs in the real world is far from perfect. Thus, one can easily criticize social science theories or even dismiss them entirely. Theory is, however, inescapable in empirical research, for at every juncture theoretical understandings drive choice. Moreover, the enterprise of theory building is itself of value. Theory not only guides research, but it is also a tool for evaluating research designs and results, and theory is essential if we are to generalize wisely from empirical results to policy prescription. Theory is, in short, an inescapable part of the empirical research enterprise, and the enterprise is better for it. As ELS or Law & Society scholars we must attend to theory to do empirical research well. Only by working to build theory will we maximize our scientific contributions.

---

fifteen degrees and a speed limit of sixty-five miles per hour (m.p.h.). We come closer to our ideal parameters if we have an embankment of twenty-five degrees and a speed limit of sixty-five m.p.h. than if we have a curve of twenty-five degrees and a speed limit of fifty m.p.h. yet the latter solution which is more distant from the optimal may be the better one. Or to give a different example, an unregulated market may lead to the most efficient distribution of goods and services if the other defining conditions of a free market are present. But in a world where some sellers or laborers have monopoly power and there are information asymmetries getting closer to the ideal market by abolishing regulation will lead to worse outcomes than wise regulation.