

2003

Empirical Inquiry Twenty-Five Years After The Lawyering Process

Stefan H. Krieger

Maurice A. Deane School of Law at Hofstra University

Richard K. Neumann Jr.

Maurice A. Deane School of Law at Hofstra University

Follow this and additional works at: https://scholarlycommons.law.hofstra.edu/faculty_scholarship

Recommended Citation

Stefan H. Krieger and Richard K. Neumann Jr., *Empirical Inquiry Twenty-Five Years After The Lawyering Process*, 10 Clinical L. Rev. 349 (2003)

Available at: https://scholarlycommons.law.hofstra.edu/faculty_scholarship/155

This Article is brought to you for free and open access by Scholarly Commons at Hofstra Law. It has been accepted for inclusion in Hofstra Law Faculty Scholarship by an authorized administrator of Scholarly Commons at Hofstra Law. For more information, please contact lawcls@hofstra.edu.

EMPIRICAL INQUIRY TWENTY-FIVE YEARS AFTER *THE LAWYERING PROCESS*

RICHARD K. NEUMANN, JR. & STEFAN H. KRIEGER*

One of the many ways in which The Lawyering Process was a pioneering book was its extensive reliance on empirical research about lawyers, lawyering, and activities analogous to some or another aspect of lawyering. To what extent has the clinical field accepted Gary Bellow and Bea Moulton's invitation to explore empirical studies generated outside legal education and perhaps engage in empirical work ourselves to understand lawyering more deeply? Although some clinicians have done good empirical work, our field as a whole has not really accepted Gary and Bea's invitation. This article explains empirical ways of thinking and working; discusses some of the mistakes law scholars (not only clinicians) make when dealing with empirical work; explores some of the reasons why empiricism has encountered difficulty in law schools; and suggests that empiricism might in some ways improve our teaching in clinics.

I. INTRODUCTION

Most novels take us through a single story. A few — *Moby-Dick* and Joyce's *Ulysses*, for example — see so much of the world in so many different ways that we can read them over and over again and each time find new and surprising things that had eluded us on earlier readings. One has that same feeling on rereading Gary Bellow and Bea Moulton's *The Lawyering Process* — the feeling of finding yet *more* in a sprawling book we had read before and thought we understood but had not really because it cannot be completely grasped the first or second or even third time we read it, a book with its own profound point of view, or points of view, not just on some issue or cluster of issues but on a big part of life itself.

One of the things one notices on reopening *The Lawyering Process* is a surprising degree of reliance on empirical research, most of it done by nonlawyers, to explain what really goes on out there in the practice of law and what could go on if we could find ways to lawyer better. In their introduction, Gary and Bea observed,

A course about lawyering, taught clinically should (i) place students

* Richard K. Neumann, Jr. is Professor of Law at Hofstra Law School. Stefan H. Krieger is Professor of Law and Director of Clinical Programs at Hofstra Law School. The authors thank Vicky Ku, Teresa Staples, and Alex Pimentel for research assistance.

with simulated or actual cases in a variety of lawyer roles; (ii) afford them a good deal of practice in describing, evaluating and solving problems connected with that experience; and (iii) encourage them to generalize about what and how they have learned in such circumstances.¹

To help students achieve the third goal of developing generalizations about their practice, Gary and Bea provided excerpts from, and make reference to, a variety of studies from different disciplines: cognitive psychology, rhetoric, economic analysis, labor relations, social work, literary criticism, and a number of others.² Some of these studies introduce students to general theories in regard to particular skills, for example, strategic and game theories for the negotiation process;³ theories on interpersonal dynamics for counseling;⁴ and communication theory for interviewing and argument.⁵ Others give them explicit models for performing a particular skill, for example strategic decision-making⁶ and questioning techniques for interviews.⁷

A final group of studies discussed by Gary and Bea is social science empirical research on issues relevant to the lawyering process. Some of this research concerns the effectiveness of particular skills techniques, for example, in negotiation or argument.⁸ Other empirical studies describe the operation of particular dispute resolution systems, for instance, the judicial decision-making process.⁹ Still others consider the dynamics of psychological and social processes, for example, problem-solving.¹⁰

So much discussion of empirical research in *The Lawyering Process* — in fact, *any* substantial discussion of empirical research in 1978 — was far ahead of its time. When the book was published, most of the existing skills literature relied more on anecdotal evidence and “war stories” than empirical studies of the techniques and processes involved in lawyering.¹¹ Bea’s article in this symposium describes her and Gary’s determination to avoid that and the tremendous amount

¹ GARY BELLOW & BEA MOULTON, *THE LAWYERING PROCESS: MATERIALS FOR CLINICAL INSTRUCTION IN ADVOCACY* xxiii-xxiv (1978).

² See *id.* at 362-63, 445-54, 523-28, 625-28 945-47, 978-91.

³ *Id.* at 464-69, 548-55.

⁴ *Id.* at 978-91.

⁵ *Id.* at 212-23.

⁶ *Id.* at 293-96.

⁷ *Id.* at 208-09.

⁸ See, e.g., *id.* at 544 (research on offer and concession strategy in negotiation); *id.* at 910 (research on sequencing of arguments).

⁹ See, e.g., *id.* at 868.

¹⁰ *Id.* at 362-63.

¹¹ Even the classic trial practice text, ROBERT E. KEETON, *TRIAL TACTICS AND METHODS* (2d ed. 1973), contained little discussion of the social science literature on the trial process.

of work she and Gary did to bring empirical insights into the book.

There has been empirical research on law for at least 80 years, although there has not been very much of it.¹² Peter Schuck wrote in 1989 that “empirical research — the uncovering of facts about how individuals and institutions within our legal culture actually behave — is a decidedly marginal activity in the legal academy today. Quantitatively, at least, it comprises a trivial proportion of the work that most law professors do.”¹³ Schuck added: “I strongly suspect that the marginality of empirical legal scholarship and the growing disjunction between the law school and the rest of society are closely related phenomenon.”¹⁴ In 2002, empirical work in the legal academy could still be described as “marginal, at best” and “the overwhelming exception to a general rule favoring nonempirical research,” although “evidence suggests that the production of empirical legal scholarship is on the rise.”¹⁵ The vast bulk of writing about the law posits theories without testing them empirically. “[T]wo forms of legal writing — doctrinal and theoretical — account for almost the entire corpus of legal scholarship. Only a tiny fraction is devoted to the gathering of *new* facts about how the law actually operates and affects us. The relative paucity of empirical work” in law “is perhaps unique among the policy-oriented disciplines, of which law is and ought to be the preeminent example. Indeed, if our colleagues in those disciplines caught on to our game, the intellectual stature of the law school in the larger university community might become even more precarious than it already is.”¹⁶

Part of the problem is what constitutes “knowledge.” In law, we “know” something if we think it is true or want it to be true and if we can come up with arguments that will persuade others that it is true. We start with the “truth” and then prove it. This is, of course, the opposite of the way science operates. Scientists — including social scientists — start with a question. Then they develop a method of inquiry that might lead to the answer, a self-disciplined investigation that might be an experiment. But when the results are in, a scientist usually does not declare that the answer is known. Instead, the scientist only publishes the methodology and results, implicitly inviting others to explore the same question using other methodologies. That

¹² JOHN HENRY SCHLEGEL, *AMERICAN LEGAL REALISM AND EMPIRICAL SOCIAL SCIENCE* (1995).

¹³ Peter H. Schuck, *Why Don't Law Professors Do More Empirical Research*, 39 J. LEGAL EDUC. 323, 323 (1989).

¹⁴ *Id.* at 325.

¹⁵ Michael Heise, *The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism*, 2002 U. ILL. L. REV. 819, 824.

¹⁶ Schuck, *supra* note 13, at 329 (italics in original).

is because only after several different methodologies are used on different sets of data, producing similar results, can we be confident that the results represent "truth". Paradoxically, it is often easier to prove something than to know it. That is because we can prove something by persuading others that it is true. But we cannot know it, in a social science sense, until we have conducted not just one but perhaps many methodologically sound inquiries.

The Lawyering Process was a massive achievement, a synthesis of the best knowledge available at the time, drawn from a wide range of interdisciplinary sources. Do we know more now empirically than we did when it was published? That is a hard question to answer. This essay assesses whether the treatment of empirical research in clinical scholarship has evolved into a more rigorous discipline over the last twenty-five years. We will first set forth some of the general features of empirical study (Part II) and provide some examples of sound social science research on the lawyering process which reflect those features (Part III), as well as some of the ways in which law teachers sometimes handle empirical research badly (Part IV). We will describe some of the obstacles to quality empirical studies by law teachers (Part V) and some of the benefits of teaching the empirical process to our students (Part VI).

In the twenty-five years since publication of *The Lawyering Process*, clinical education, as a field, has tended not to have accepted Gary and Bea's invitation to pursue avidly what empirical research can reveal about the law and lawyering. That is as true of us (Neumann and Krieger) as it is of the field as a whole. Like other textbooks written since *The Lawyering Process*, ours¹⁷ relies on empirical research only occasionally. For the most part, we, like everybody else, tell students what we have gathered from our own observation and experience and from the debates among clinicians about lawyering (which is not the same as empirical knowledge). Like others, we have been deterred from doing more by at least some of the factors discussed in Part V of this article. We are not happy about that because we would like the greater insights (and confidence in them) that empirical methods make possible. This article is our attempt to understand what has happened to empiricism in our field, and in legal scholarship generally, since *The Lawyering Process*. Some parts of the article are as much a criticism of ourselves as of anybody else.

¹⁷ STEFAN H. KRIEGER & RICHARD K. NEUMANN, JR., *ESSENTIAL LAWYERING SKILLS: INTERVIEWING, COUNSELING, NEGOTIATION AND PERSUASIVE FACT ANALYSIS* (2d ed. 2003).

II. THE NATURE OF EMPIRICAL RESEARCH

Empirical research is the use of generally-accepted methodologies to collect data (facts about the world) and analyze them.¹⁸ Although this definition could be read to include nearly all legal scholarship because it concerns observations about the world,¹⁹ empirical research is generally understood to mean the collection and analysis of data using standard social science methodologies. *Interdisciplinary* and *empirical* are not synonyms. To law faculty, empirical work is interdisciplinary because it requires an understanding of one or more of the social sciences. But much interdisciplinary work is not empirical. Law and literature, for example, is not an empirical field although it is by definition interdisciplinary.

Quantitative empirical research gathers numeric information, often from a large number of cases, and subjects it to statistical analysis. In contrast, qualitative research collects and analyzes nonstatistical data using methods such as case studies, ethnographic field work, and comparative historical analysis.²⁰ The purpose of both types of empirical study is to analyze the data collected to develop explanations of facts about phenomena we cannot observe or inferences about the causes of particular outcomes.²¹

The more common methods of empirical research include field investigations that generate quantitative data; analysis of quantitative data generated by others, such as government agencies; experiments that simulate reality; field observation in sociology or anthropology; surveys; evaluative research; content analysis; and historical or comparative analysis. The last three might require some elaboration. Evaluative research measures the situation before a change in law or administrative policy, measures it again afterward, and compares the two to discover the effect of the change; this may involve using one or

¹⁸ See Thomas E. Willging, *Past and Potential Uses of Empirical Research in Civil Rulemaking*, 77 NOTRE DAME L. REV. 1121, 1126 (2002) ("‘Empirical research’ . . . refer[s] to information collected through systematic observation and experience (in contrast, for example, to information derived through theory or logic)").

¹⁹ See Russell Korobkin, *Empirical and Experimental Methods in Law: Empirical Scholarship in Contract Law: Possibilities and Pitfalls*, 2002 U. ILL. L. REV. 1033, 1035 n.3.

²⁰ See generally Henry E. Brady, et al., *Refocusing the Discussion of Methodology*, in RETHINKING SOCIAL INQUIRY: DIVERSE TOOLS, SHARED STANDARDS (Henry E. Brady & David Collier, eds., forthcoming 2003) (manuscript at 1-2, on file with authors).

²¹ Lee Epstein & Gary King, *Exchange: Empirical Research and the Goals of Legal Scholarship: The Rules of Inference*, 69 U. CHI. L. REV. 1, 35 (2002). Although the Epstein-King article makes some useful observations on law and empiricism and some valid criticisms of some of the empirical research published in law reviews, our citation to it is not an endorsement of the article as a whole. Some of the article's conclusions are questionable and were questioned in reply articles in the same issue of the University of Chicago Law Review. And the Epstein-King article's smug and accusatory tone makes it an example of the very kind of scholarship it criticizes.

more the other methods in a pretest/posttest research design. Content analysis is the systematic analysis of speech or documents for the frequency or depth in which certain topics are discussed or the styles of discourse used. Historical and comparative analyses — also called historical-comparative and H-C — is the oldest cluster of methods, used by Durkheim, Weber, and others during the period when the social sciences were first being created. Historical-comparative analyses declined as the other methods developed, especially the quantitative ones, and then rebounded after about 1950, sometimes in combination with other methods. Historical analysis elucidates a social phenomenon by tracing its history, and comparative analysis does so by comparing the phenomenon in different contexts — for example the role of women in country music compared to the role of women in rap music.

While both quantitative and qualitative researchers gather different types of data, some social scientists have argued that research in both methods should use the same methodology. In a highly-influential 1994 book on social science methodology²² and in a recent law review article based on the book,²³ Gary King, a Harvard political science professor, and his coauthors assert that all “good” social science research requires the use of the tools of the quantitative method. They urge a highly systematic approach to all social science empirical research, “the Rules of Inference,” which focuses on the value — from a statistical standpoint — of increasing the amount of data observed.²⁴ This approach has been the subject of much controversy among social science methodologists.²⁵ As two critics of the approach argue, King and his co-authors

overemphasize[] the strategy of increasing the number of observations, without giving sufficient attention to different kinds of observations and to the different ways that data are used in quantitative and qualitative research. [They] pay[] too little attention to the risk that increasing the [number of observations] may push scholars toward an untenable level of generality and a loss of contextual

²² GARY KING, ET AL., *DESIGNING SOCIAL INQUIRY: SCIENTIFIC INFERENCE IN QUALITATIVE RESEARCH* (1994).

²³ Epstein & King, *Rules of Inference*, *supra* note 21, at 1.

²⁴ See, e.g., KING, ET AL., *supra* note 22, at 18 (1994); Epstein & King, *Rules of Inference*, *supra* note 21, at 24, 102-03.

²⁵ See, e.g., *RETHINKING SOCIAL INQUIRY*, *supra* note 20; Timothy J. McKeown, *Case Studies and the Statistical Worldview: Review of King, Keohane, and Verba's Designing Social Inquiry: Scientific Inference in Qualitative Research*, 53 INT'L ORG. 161 (1999); *The Qualitative-Quantitative Disputation: Gary King, Robert O. Keohane, and Sideny Verba's Designing Social Inquiry: Scientific Inference in Qualitative Research*, 89 AM. POL. SCI. REV. 454 (1995).

knowledge.²⁶

In this essay, we do not intend to become entangled in the fray over the merits of the "Rules of Inference" approach to empirical methodology. We agree with the critics of that approach that empirical methodology cannot be simplified into procrustean formulas and must take into account different trade-offs, such as the size of the sample for observation and consideration of the specific context of the study. But we do recognize that, regardless of the size of the sample collected by researchers, there are a number of common features to "good" empirical social science research. Indeed, even King's critics acknowledge that a common framework can exist for both quantitative and qualitative research.²⁷ We believe that most, if not, all social scientists would agree that empirical research should have the following attributes.

1. *Open-mindedness*: Scientists do not set out to "prove" a theory. Rather, they attempt to learn as much about the world as possible and develop inferences and explanations about it. Most lawyers and many law professors approach their work predominantly in persuasion mode²⁸ trying to impose a "self-interested and largely pre-determined meaning on communication, principally by arguing with others over the authorship of ideas."²⁹

Persuasion-mode habits cause a person to argue forcefully, hyperbolically, and at length [and] to know where one stands on an issue at the moment it is raised. . . . Complex phenomena are needlessly reduced to some of their parts, and tentativeness and uncertainty are unnecessarily suppressed. . . . Persuasion-mode habits predispose lawyers to take evaluative stands automatically, as a first response to others' new ideas . . . [and] may also wear down loyalty to

²⁶ Brady, *Refocusing the Discussion of Methodology*, *supra* note 20, at 4. In fact, the critics argue that "alternative methodological tools are relevant and appropriate depending on the goals and context of the research." *Id.* at 7.

²⁷ *Id.* at 4-5.

²⁸ Robert J. Condlin, *Socrates' New Clothes: Substituting Persuasion for Learning in Clinical Practice Instruction*, 40 MD. L. REV. 223, 228-48 (1981). See also Robert Condlin, *The Moral Failure of Clinical Legal Education* in *THE GOOD LAWYER: LAWYERS' ROLES AND LAWYERS' ETHICS* 318, 326-332 (David Luban ed., 1983); DONALD A. SCHÖN & CHRIS ARGYRIS, *THEORY IN PRACTICE: INCREASING PROFESSIONAL EFFECTIVENESS* 63-95 (1974, 1992).

Schön and Argyris were the first to describe these two ways of thinking and communicating; they called them Type I and Type II. Condlin was the first person to discuss them in the clinical literature; he called them persuasion mode and learning mode. We use the term inquiring mode instead of learning mode because it is more descriptive.

²⁹ Robert J. Condlin, *Learning from Colleagues: A Case Study in the Relationship Between "Academic" and "Ecological" Clinical Legal Education*, 3 CLIN. L. REV. 337, 354 (1997). See also Lee Epstein & Gary King, *Exchange: Empirical Research and the Goals of Legal Scholarship: A Reply*, 69 U. CHI. L. REV. 191, 192 (2002) ("Lawyers, judges, and hence law professors . . . specialize in persuasion.").

accuracy . . . [and] may weaken one's general disposition to seek and speak the truth.³⁰

In persuasion mode, the goal of fact discovery is primarily to support a particular position. Scientists, however, operate in an "inquiring mode": open-ended curiosity, an interest in exploring things regardless of their consequences, and acknowledgment of ambiguity.³¹ They attempt to approach the discovery of facts with an open mind.

In fact, scientists often purposefully select theories that could be wrong.³² As a Nobel physicist puts it,

Experimenters search most diligently, and with the greatest effort, in exactly those places where it seems most likely that we can prove ourselves wrong. In other words, we are trying to prove our theories wrong as quickly as possible because only in that way can we find progress.³³

A primary goal of any scientific research is to eliminate rival explanations for outcomes. The primary purpose of this quest is not to strengthen support for researchers' theories or challenge a rival scientist's theories but rather to test the validity of any inferences or explanations they make about the data observed.

Even when a particular study is complete, scientists do not consider their work finished. Research is viewed as an on-going and evolving process. If, for example, the prediction of the researcher's theory was proven correct but was based on several variables, the scientist will want to drop one of the conditions in the next study to further fine-tune the analysis. If, on the other hand, the original theory was proven wrong, the scientist most likely will want to propose future empirical studies in the area to explore alternative theories.³⁴ A researcher, then, tries to keep an open mind even after a specific project is brought to closure.

2. *Concrete and Narrow Theory Design:* Approaching their task with an open mind, scientists attempt to design hypotheses and theories as concretely as possible. In this way, after the data are gathered, they will more easily be able to analyze critically whether their predictions are right or wrong.³⁵ Vague or ambiguous theories or hypotheses obscure issues and make it more difficult to control for competing hypotheses. We might want to do that when advocating in persuasion mode, but we should not when playing the role of the empirical re-

³⁰ Robert Condlin, *Moral Failure*, *supra* note 28, at 330.

³¹ Robert J. Condlin, *Learning from Colleagues*, *supra* note 29, at 354, 357-59.

³² Ronald Rogowski, *The Role of Theory and Anomaly in Social-Scientific Inference*, 89 AM. POL. SCI. REV. 467, 469 (1995).

³³ RICHARD PHILLIPS FEYNMAN, *THE CHARACTER OF PHYSICAL LAW* 158 (1965).

³⁴ See KING, ET AL., *DESIGNING SOCIAL INQUIRY*, *supra* note 22, at 21-23.

³⁵ *Id.* at 20.

searcher in inquiring mode.

Moreover, scientists usually attempt to narrow the scope of their inquiry to take into account, as much as possible, the different variables at play and the full context of the processes being observed. Expansive hypotheses involving processes with complex variables cannot easily be evaluated. Research based on such hypotheses do little more than set the research agenda for future, more limited studies.

3. *Unbiased Data:* Similar to open mindedness, another feature of empirical research is the attempt to eliminate bias in data collection. Unlike lawyers in the discovery process, scientists do not gather data with the purpose of proving a point or developing a particular inference. Rather, they try to select a sample of observations that will remove as much as possible the likelihood of biased inferences. Different methods have been developed to avoid sampling bias.

If, on the one hand, researchers are collecting an array of scores on specific variables for a sample of cases, they usually try to increase the number of observations and to use a random probability sample. In this way, they gather sufficient information about the sample of cases "to ensure the absence of omitted variable bias."³⁶ The larger the data pool, the more opportunity researchers have to consider rival hypotheses.

On the other hand, if the researcher seeks to gather data concerning the relationship between different variables in a particular process or context — what scientists call "causal-process observations" — increasing the size of the sample may not be very helpful.³⁷ In the context of such investigations, "[i]ncreasing the number of observations . . . may push scholars to compare cases that are not analytically equivalent."³⁸ Especially, in studies of human behavior, the precise nature of the cultural, economic, interpersonal, and historical settings of different interactions may have a profound effect on the processes studied. A large, randomized study may result in inferences that incorrectly ignore that context.

The danger remains, however, that small-case studies have the potential for greater selection bias than large sample, randomized re-

³⁶ Epstein & King, *Rules of Inference*, *supra* note 21, at 77; see generally *id.* at 24, 108; Brady, et al., *Refocusing the Discussion of Methodology*, *supra* note 20, at 9.

³⁷ See generally Henry E. Brady, *Doing Good and Doing Better: How Far Does a Quantitative Template Get Us?* in *RETHINKING SOCIAL INQUIRY*, *supra* note 20 (manuscript at 6, on file with authors); Brady, *supra* note 20, at 8-9; David D. Laitin, *Disciplining Political Science*, 89 AM. POL. SCI. REV. 454, 456 (1995). For an example of the difference between these two types of data collection, see Henry E. Brady, *Data Set Observations Versus Causal Process Observations: The 2000 U.S. Presidential Election*, in *RETHINKING SOCIAL INQUIRY* *supra* note 20 (manuscript, Appendix A, on file with authors).

³⁸ Brady, et al., *Refocusing the Discussion of Methodology*, *supra* note 20, at 8.

search. To address this problem, researchers who perform these studies attempt to develop case selection rules that are not related to the dependent variable.³⁹ Just as important, they attempt to make their criteria for choosing the sample case or cases transparent so that other researchers can assess the validity of any findings.

4. *Valid and Reliable Rules for Measuring Data:* The value of the findings of any empirical study depends, in large part, on the validity and reliability of the measurement methods. The measurement process needs to be approached with the same open mindedness as theory design and sample selection. Obviously, measurement methods are only valid if they properly measure the data observed. And they are only reliable if the same measurement standard is applied to all the data collected.

While some researchers have urged that human judgment should be removed as much as possible from measurement,⁴⁰ this goal is very difficult, if not impossible, in much social science and other nonquantitative research. It is one thing to measure variables such as "heat," "weight," and "speed." It is quite another thing to measure variables such as "utility," "democratic values," "nationalism," or, for that matter, "effective legal practice." Indeed, even in the medical sciences, it is often difficult to measure certain phenomena, such as "recovery." For such qualitative concepts, valid standards of measurement cannot be created out of whole cloth by individual researchers but must be based on prior qualitative and quantitative research on similar issues, impressionistic evidence in the field, and theoretical models.⁴¹ Valid measurement standards, therefore, develop as a product of a social enterprise of scientists building on each other's work. While precise measurement of certain qualitative variables may be impossible, empirical researchers consider these ongoing efforts as a way of continuously developing more accurate standards.

Because the development of measurement rules is a social enterprise, scientists place great importance on the clear articulation of these rules in research studies. Such transparency gives other researchers the opportunity to assess rules in light of standards used in other studies and suggested by theoretical models and to evaluate their relative strengths and weaknesses. In this way, ever-evolving measurement standards develop.

5. *Critical Analysis of Data:* Once the data are collected and measured, scientists continue in the inquiring mode critically analyz-

³⁹ See Epstein & King, *Rules of Inference*, *supra* note 21, at 113.

⁴⁰ See *Id.* at 85.

⁴¹ See Henry E. Brady, *Doing Good*, *supra* note 37, at 11-12; Epstein & King, *Rules of Inference*, *supra* note 21, at 89.

ing their findings. Instead of trying to fit the data into the original hypothesis, they try to evaluate it in light of rival hypotheses.⁴² They acknowledge instances when their hypotheses are not borne out and try to explain the reasons for the divergence.

Even when the data support the original hypothesis, researchers are cautious in reaching conclusions. They avoid "reaching conclusions about causation when the data prove only a correlation, and avoid[] reaching conclusions about the world writ large when the data support only a conclusion about a subset of the world."⁴³ Especially in research concerning social processes, researchers are constantly aware of the problems that human subjects are complex and, in experimental conditions, can easily change their behavior, censor their observations, or simply drop out of the study.⁴⁴

Good empirical research, therefore, forthrightly recognizes the degree of uncertainty in any conclusions.

In any empirical research project, there always exist uncorrected methodological problems. The sign of good science is not that every problem be fixed in every project, since that is impossible and trying to do so would mean we would not accomplish much of anything, but rather that the problems be noted and the uncertainty in substantive conclusions be responsibly and honestly reported.⁴⁵

As with any circumstantial evidence, the point is not that it is impossible to draw any reasonable inferences; rather, it is that there is some uncertainty with any inference that is made.

6. *The Public Nature of Research:* Throughout our description of the features of empirical research, a constant thread has been the social nature of the enterprise: in designing their studies, scientists consider rival hypotheses of other researchers; sample selection criteria are made transparent for critical assessment by other scientists; measurement rules develop from past research experiences; and analyses of data are always undertaken with an eye to hypotheses of other researchers and theorists. The quality of empirical research is ideally not assessed based upon the reputation and status of the author but upon the contribution of the data and analysis to the social enterprise of inquiry.⁴⁶

Moreover, the value of any empirical research depends, in large part, on the ability of other researchers to replicate the research.⁴⁷

⁴² See Epstein & King, *Rules of Inference*, *supra* note 21, at 77.

⁴³ Korobkin, *supra* note 19, at 1050-51.

⁴⁴ Brady, *Doing Good*, *supra* note 37, at 9.

⁴⁵ Epstein & King, *Reply*, *supra* note 29, at 208.

⁴⁶ Epstein & King, *Rules of Inference*, *supra* note 21, at 45.

⁴⁷ See generally Gary King, *Replication, Replication*, 28 PS: POL. SCI. & POL. 444 (1995). One study attempting to replicate articles published in an economics journal over

For quantitative and qualitative analyses alike, we need the answers to questions such as these: How were the respondents selected? Who did the interviewing? What was the question order? How did you decide which informants to interview or villages to visit? How long did you spend in each community? Did you speak to people in their language or through an interpreter? . . . How knowledgeable were the coders [of data]? How frequently did the coders agree? Exactly what codes were originally generated and what were all the recodes performed? . . . What were the exact rules for content analysis? . . . Which computer program was used? How did you fill in or delete missing data?⁴⁸

The answers to these and other questions allow other researchers to continue the study of the issues involved, perhaps with different sample selection or measurement rules, to either confirm or fine tune the findings. And even if exact replication is not conducted, answers to these questions allow for a full assessment of the methods of data collection and analysis by other scientists.

Because of the importance of the ability to replicate, empirical scientists are encouraged to disclose publicly not only their methods for sample selection, data collection, encoding, measurement, and analysis, but also the precise wording for questions and response rates. For survey research involving data-gathering orally, researchers are urged to provide transcripts of interviews and video or audio tapes. For both quantitative and qualitative social science studies, professional archival databases have been established for scholars to share their research data with other scientists.⁴⁹

III. HOW EMPIRICISM CAN ELUCIDATE LAW AND ITS PRACTICE

Empirical research has the capacity to test aspects of law and its practice that are based on assumptions and "common sense" to see whether they reflect reality. This section describes some examples. If your favorite empirical study is missing, we mean no disrespect. We are only describing a sample of published empiricism to illustrate the wide range of empirical methods available and the ways in which they can help us avoid being fooled by what we think we see.

The law has long assumed that a trial judge or jury who sees and hears the witnesses first-hand is better able to assess their credibility than appellate judges who read a cold transcript. Several concepts governing appellate practice are based on this assumption.⁵⁰ Olin

the course of a year found that inadvertent errors are commonplace. *See id.* at 445.

⁴⁸ *Id.*

⁴⁹ *Id.* at 446.

⁵⁰ For example, see Rule 52(a) of the Federal Rules of Civil Procedure: Judicial "[f]indings of fact . . . shall not be set aside unless clearly erroneous, and due regard shall

Guy Wellborn surveyed the published experiments conducted by social scientists to learn whether a fact-finder's first-hand observation of a witness's demeanor actually helps in determining whether the witness is testifying honestly as well as whether the witness was able to see and hear and remember accurately the things the witness is testifying to.⁵¹ Wellborn did not conduct experiments himself; he searched the social science literature for experiments already conducted by others. The only advantage first-hand observers of the witness have is the opportunity to see the witness's facial expressions and other body language and hear the witness's tone of voice. In nearly all the experiments, this nonverbal material did not help fact-finders identify truthful and accurate witnesses, and in a number of experiments it seemed to confuse the fact-finders. Where readers of cold transcripts were compared with observers of live witnesses, the cold transcript readers tended to be better judges of witness credibility, apparently because they can concentrate on the logic of the testimony and its consistency with the testimony of other witnesses without being distracted by myths about what credible witnesses look and sound like. This opens up profound questions, not only about appellate review of trial court fact-finding but also about how trial courts should go about finding facts.⁵²

Gay Gellhorn, Lynne Robins, and Pat Roth developed an interdisciplinary clinical course in which graduate anthropology students conducted a linguistic analysis of client interviews conducted by the clinical law students who were also enrolled in the course.⁵³ Here is their description of the methodology:

The primary technique of linguistic anthropology that we used was the process of microlinguistic analysis of these interview tapes and transcripts. . . . [W]e reviewed the recorded and transcribed text over and over, and through that process discovered otherwise hidden dimensions of meaning, or meanings that might otherwise be taken for granted. . . . The review may include word searches, counts of recurrent linguistic or behavioral patterns, notations to indicate which speaker is controlling the choice of topic, counts of particular forms of questions, identification of moments when body language and spoken words are incongruent, notation on recycling of topics, or any number of other inquiries. . . . The participation by the law students as well as the anthropology students and the

be given to the opportunity of the trial court to judge the credibility of the witnesses."

⁵¹ Olin Guy Wellborn III, *Demeanor*, 76 CORNELL L. REV. 1075 (1991).

⁵² For nonempirical reasons, Wellborn does not favor changing much of the law on fact-finding and its appellate review. See *id.* at 1091-104.

⁵³ Gay Gellhorn, Lynne Robins & Pat Roth, *Law and Language: An Interdisciplinary Study of Client Interviews*, 1 CLIN. L. REV. 245 (1994).

authors in the process of microlinguistic analysis of the data placed all of us in the ethnographer's role of participant observer⁵⁴

The anthropology students wrote up in term papers their work and the inferences they drew from it.⁵⁵ In a follow-up project, Gay Gellhorn studied videotapes and transcripts of client interviews, some conducted by clinical law students and others by lawyers and paralegals in a legal aid office.⁵⁶ She also reviewed the published studies of medical interviewing.⁵⁷ Combining insights from both of these empirical sources, she developed a model for effectively opening a client interview.⁵⁸

Using linguistic content analysis of simulated client interviews conducted by clinic students who had been taught client-centered interviewing, Linda Smith tested whether "the techniques we teach produce the client empowerment we seek."⁵⁹ Looking for "linguistic indicia of power and dominance" in the interviews, she identified interviewing techniques that were "crucially important in 'client-centered' representation" as well as other interviewing behavior that was "shown to be in need of improvement."⁶⁰ In a subsequent article, she surveyed the published studies of medical counseling to develop a model of how to give "bad news" to law clients.⁶¹

Austin Sarat and William Felstiner observed and tape-recorded office conversations between divorce lawyers and their clients to discover how the lawyers explained the law, the legal system, and judges to their clients.⁶² "This effort consisted of following one side of 40 divorce cases" in California and Massachusetts, "involving 20 different lawyers, ideally from the first lawyer/client interview until the divorce was final," for a total of 115 lawyer/client conversations.⁶³ The lawyers typically described the law as arbitrary and judges as motivated by many factors other than a good faith attempt to achieve justice. Sarat and Felstiner speculated that "[l]awyer cynicism and pessimism about legal actors and processes is a means through which they seek

⁵⁴ *Id.* at 252-53.

⁵⁵ *Id.* at 264-67.

⁵⁶ Gay Gellhorn, *Law and Language: An Empirically-Based Model for the Opening Moments of Client Interviews*, 4 CLIN. L. REV. 321 (1998).

⁵⁷ *Id.* at 335-44.

⁵⁸ *Id.* at 344-58.

⁵⁹ Linda F. Smith, *Interviewing Clients: A Linguistic Comparison of the "Traditional" Interview and the "Client-Centered" Interview*, 1 CLIN. L. REV. 541, 544 (1995).

⁶⁰ *Id.*

⁶¹ Linda F. Smith, *Medical Paradigms for Counseling: Giving Clients Bad News*, 4 CLIN. L. REV. 391 (1998).

⁶² Austin Sarat & William L.F. Felstiner, *Lawyers and Legal Consciousness: Law Talk in the Divorce Lawyer's Office*, 98 YALE L.J. 1663 (1989).

⁶³ *Id.* at 1669.

to control clients and maintain professional authority”⁶⁴ by presenting themselves to clients as experts not on the law, but “on local knowledge, insider access, connections, and reputation. Lawyers often suggest that their most important contribution is knowledge of the ropes, not knowledge of the rules. They describe a system that is not bureaucratically rational but is, nonetheless, accessible to its ‘priests.’”⁶⁵ Some of the conversations they quote are eerily similar to the one K. has with his lawyer in Kafka’s *The Trial*. In a follow-up study of the same data, Felstiner and Sarat analyzed the way lawyers and clients decide what goals they will pursue and a division of power and responsibility between themselves.⁶⁶

Douglas Maynard did a content analysis of tape-recorded courthouse plea bargaining in 52 misdemeanor cases.⁶⁷ Because these conversations were short, totaling less than ten hours, they would predictably have a high ratio of content to effect. But in fact little negotiation happened because

many of the sample cases were settled simply by one party proposing a disposition and the other agreeing to it. . . . In most cases where the two sides advanced different positions this did not prompt proposals for an intermediate solution but instead simply raised the issue of which original proposal would be accepted. In the literature on negotiation, this is considered to be . . . at least a rare phenomenon.⁶⁸

Maynard’s study could not determine why this happened — whether the lawyers involved did not have time for true negotiation, whether they did not know how to do it, or whether the stakes seemed to them too low to justify it.

Milton Heumann and Jonathan Hyman looked at negotiation from a different angle.⁶⁹ They surveyed New Jersey lawyers on what tends to happen when they negotiate nonmatrimonial civil disputes as well as what the lawyers would like to have happen, and they followed up with interviews of 78 lawyers and observations of 71 settlement conferences. Among other things, they asked respondents how often negotiations were positional or problem-solving and whether respondents liked things that way. In the questionnaires, they told respon-

⁶⁴ *Id.* at 1665.

⁶⁵ *Id.* at 1685.

⁶⁶ William L.F. Felstiner & Austin Sarat, *Enactments of Power: Negotiating Reality and Responsibility in Lawyer-Client Interactions*, 77 Cornell L. Rev. 1447 (1992).

⁶⁷ Douglas W. Maynard, *The Structure of Discourse in Misdemeanor Plea Bargaining*, 18 LAW & SOC’Y REV. 75 (1984).

⁶⁸ *Id.* at 98.

⁶⁹ Milton Heumann & Jonathan M. Hyman, *Negotiating Methods and Litigation Settlement Methods in New Jersey: “You Can’t Always Get What You Want,”* 12 OHIO ST. J. ON DISP. RESOL. 253 (1997).

dents to classify negotiation as positional if "the negotiators stake out bargaining positions [and n]egotiation consists of one or more moves and countermoves in which the parties may grant concessions to the other party, and seek agreement by the reciprocal exchange of positions" ⁷⁰ They told respondents to classify a negotiation as problem-solving if there is "mutual discussion of the underlying needs and interests of each side [and a]greement results not as much from an exchange of concessions as from new proposals that both parties think meet their needs." ⁷¹ The respondents reported that "most cases were settled using only the positional method." ⁷² But 61% of the respondents wanted to use problem-solving more, and 47% wanted to use positional negotiating less. Why would so many lawyers do something they dislike and consider counterproductive? Heumann and Hyman saw some evidence that "[l]awyers seem to lack a rich vocabulary of problem-solving that would enable them to implement that method. And they may also be blocked by the time and energy they would have to use to change their positional negotiating habits." ⁷³

Criminal law is based on the assumption that punishment discourages law-breaking. A research team administered anonymous questionnaires periodically to Stockholm teenagers. ⁷⁴ The questionnaires asked whether the teenagers had committed various crimes and if so, whether they had been caught by the authorities and what the authorities had done about it. The questionnaires did not identify the crimes as crimes. Instead, using everyday language the questionnaires described activities in a value-neutral way and asked whether the teenagers had done them. The anonymity and the straight-forward and value-neutral language were essential to validity. The results showed that, in a city that at the time had a very low crime rate, the overwhelmingly majority of teenagers — regardless of social, economic, and academic background — were criminals. The lowest rates of recidivism were among those who had never been caught by the authorities. The highest rates of recidivism were among those who had been caught and punished. In between were those who had been caught but not punished. The research suggests that, among the teenagers studied, episodes of criminality were an ordinary part of growing up, and that the criminality tended to disappear spontaneously as the

⁷⁰ *Id.* at 254.

⁷¹ *Id.* at 255.

⁷² *Id.*

⁷³ *Id.* at 309.

⁷⁴ FAKTISK BROTTSLIGHET BLAND SKOLBARN, SOU 1969:1. The title translates as Actual Criminality among Schoolchildren. SOU is the abbreviation for *Sveriges Offentliga Utredningar* (Official Reports of Sweden), which publishes the results of government commissions and investigations intended to inform legislation and administrative activities.

teenagers matured. In other words, among teenagers criminality came and went in ways that were impossible for the authorities to control, and the only effect of the authorities' intervention was to make things worse.

Substantive criminal law — definitions of crimes and defenses as well as grading of crimes and sentencing ranges — is adopted by legislators who are confident that they are giving the public what it wants. Paul H. Robinson and John M. Darley conducted eighteen separate surveys of public opinion to find out whether the public has gotten what it wants in criminal code provisions regarding “the justified use of force, insanity, causation, complicity, risk-creation, omission liability, culpability requirements, duress, entrapment, multiple offenses, and criminalization matters such as felony murder and sexual offenses. The questions selected for investigation [were] those for which the community’s intuition would be of importance to legal decision-makers.”⁷⁵ The surveys did not ask respondents directly what they wanted. Instead, respondents were asked how they think a series of hypotheticals should be resolved. The hypotheticals were constructed so that a respondent’s answers would illustrate the principles to which the respondent adhered, even if the respondent was not consciously aware of them.⁷⁶ This technique, called a factorial survey, “combines elements from experimental psychology with those of survey sociology. From experimental psychology, one takes the notion of the balanced multivariate experimental design and from sociology the method of sample survey procedures.”⁷⁷ Its value is to prevent the respondent from giving glib or self-flattering answers and to get the respondent to work out how she really feels by responding to situations rather than theorizing. In a number of respects, the respondents’ intuition differed considerably from the code drafters’ intuition.

These are examples of methodologically sound and empirically imaginative research that has creatively explored whether what we or the law assumes to be true really is true. But an empiricist would quite rightly say that few of these studies “prove” anything because few of them have been explored in replication studies by later researchers who would either apply the same methodology to different populations or develop a different methodology to explore the same questions. (Of course, that is not so for the Wellborn and Smith

⁷⁵ PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* xv (1995). This book aroused considerable comment and controversy. See *Is Justice Just Us? A Symposium on the Use of Social Science to Inform the Substantive Criminal Law*, 28 HOFSTRA L. REV. 601-797 (2000).

⁷⁶ ROBINSON & DARLEY, *supra* note 75, at 217.

⁷⁷ *Id.*

surveys of research already published in social science and medical journals. There was enough research to be surveyed because people in those fields had tried to replicate each other's findings, which is a routine part of social science.)

This illustrates a failure, to which we will return in a moment, in law teachers' understanding of the scientific method. A law teacher looking for a research topic would typically read one of these studies and think, "Well, I shouldn't do anything remotely like this because it's already been done." A scientist would think, "There's been just one study, and only attempts at replication can show whether it describes reality. Let's try to replicate and see what happens." And deep down, every one of us knows that the scientist is right. If we are in persuasion mode, the lawyer in us might use a single study to argue that a proposition we like is true. But if we reflect quietly, a nagging doubt sneaks in so that we wonder, "What if these results are an anomaly? What if other populations don't behave that way? What if the same population behaves one way when the researchers are looking and another way a week or so later, when the researchers aren't looking?" How can we really know what's going on out there? By trying to replicate the results until we reach the point of diminishing returns.

IV. WHAT GOES WRONG IN EMPIRICAL LEGAL SCHOLARSHIP

Social scientists complain about how law teachers do and use empirical research. According to Lee Epstein and Gary King, much empirical work appearing in law reviews "ignores the rules of [scientific] inference and applies instead the 'rules' of persuasion and advocacy."⁷⁸ Even some law teachers complain about it. Deborah Rhode says that "[r]esearch by law students and professors with no formal training in social science methodology provides constant reminders of the limitations of armchair empiricism. Sloppy survey techniques, skewed samples, and sweeping generalizations from unrepresentative findings are among the most common problems."⁷⁹

Because *The Lawyering Process* often referred to studies on particular skills without evaluating the strengths or weaknesses of the research, the empirical insights can appear to be arguments in support of a particular theory or model. At times, Gary and Bea draw inferences from studies without describing their specific contexts or identifying the limitations in their conclusions. That may have resulted from the editing needed to bring a huge manuscript down to a size

⁷⁸ See Epstein & King, *Rules of Inference*, *supra* note 21, at 9.

⁷⁹ Deborah L. Rhode, *Legal Scholarship*, 115 HARV. L. REV. 1327, 1343 (2002).

that could be published in one volume. In discussing the "law of primacy" in argumentation, for instance, Gary and Bea say that

If one assumes a "law of primacy" in persuasion, an anti-climax order (strongest arguments first) would seem to be preferable, and this position, too, has its supporters among trial lawyers. Such an assumption is warranted by most research on learning and recall. When presented with a series of statements under experimental conditions people seem to remember the statements placed at the beginning of the sequence, and at least one experiment purports to demonstrate larger shifts of opinion in an audience presented arguments in an "anti-climax rather than a climax order."⁸⁰

While this suggests that lawyers should lead with their strongest arguments in trials or oral arguments, the studies cited by Gary and Bea did not focus on legal advocacy before judges or juries. Rather, they concerned formal speeches on topics of general public interest presented to undergraduate introductory public speaking students.⁸¹ The differences between these two audiences — legal decision-makers and a public audience — and the nature of the argument — issues related to a specific case and questions of public interest — are significant. The author of one of the two studies warned against the "general application" of his research to other contexts.⁸²

Although Gary and Bea sometimes evaluated the methodological problems with the studies they cited,⁸³ they often did not. That omission is understandable in light of the tremendous task they had undertaken, surveying virtually all knowledge then published relevant to lawyering, as well as the massive cuts they must have had to make in their manuscript to bring it down to publishable size. For example, in discussing concession strategy in negotiation, they observe that

there is . . . evidence that small infrequent concessions combined with an unyielding or tough approach produce far more favorable settlements than a conciliatory strategy. One study of bargaining behavior concluded that a more inflexible concession posture produced more favorable outcomes than one more prone to compromise, although in both situations concessions were, in fact, reciprocated. Another experiment suggests that the longer one party is able to maintain a posture outside of even the minimal

⁸⁰ BELLOW & MOULTON, *supra* note 1, at 910 (footnotes omitted).

⁸¹ Harvey Cromwell, *The Relative Effect of Audience Attitude of the First Versus Second Argumentative Speech of a Series*, 17 *SPEECH MONOGRAPHS* 105 (1950); Harvey Sponberg, *A Study of the Relative Effectiveness of Climax and Anti-Climax Order in Argumentative Speech*, 13 *SPEECH MONOGRAPHS* 35 (1946).

⁸² Cromwell, *supra* note 81, at 121.

⁸³ For example, see BELLOW & MOULTON, *supra* note 1, at 867-69 (identifying the methodological problems with studies attempting to create a science for predicting judicial decisions and concluding "a 'science' of such judgments has a very long way to go.").

point at which the other party would settle, the more likely he or she is to obtain a disproportionately favorable settlement.⁸⁴

But the evidence is subject to significant limitations in methodology. Both of the studies concerned simulated negotiations, not real-life bargaining in which the parties had a stake in the outcome.⁸⁵ The "laboratory" nature of the research clearly may have affected the outcome.⁸⁶ Moreover, the subjects in each of the studies were students. (In one study, introductory psychology students bargained for the sale of a used car, and in the other MBA students negotiated for the purchase of information to maximize returns on an investment.)⁸⁷ The limited experience of the negotiators may have affected the results. Finally, the size and nature of the sample selection in both studies may have skewed the results. The study for the information bargaining had only 24 subjects,⁸⁸ and the used car negotiation used only female subjects.⁸⁹ These problems weigh against sweeping conclusions about the results.

The creation of a book as encyclopedic as *The Lawyering Process* — at a time when very little empirical work was being done in law schools and when efficiencies like word-processing and web research were unknown — was such a heroic achievement that flaws like these seem understandable and only faintly troubling in retrospect. But in the 25 years since then, have empirical techniques become widely understood and commonly used on law faculties? The answer, embarrassingly, is no. Some good work has been done, but so has some not-so-good work.

How do law teachers go wrong when working with empirical material? Here are the most common difficulties:

1. *Inference problems:* Perhaps the most prominent fault in empirical research conducted or interpreted by law teachers is overgeneralizing from the data. Here is a simple example, where law teachers are only reporting research done by others:

Although forms of summary proceedings existed in the United

⁸⁴ *Id.* at 544 (footnotes omitted).

⁸⁵ Bernard L. Hinton, et al., *The Influence of Reward Magnitude, Opening Bid, and Concession Rate on Profit Earned in a Managerial Negotiation Game*, 19 BEHAV. SCI. 197, 198 (1974); Jerome M. Chertkoff & Melinda Conley, *Opening Offer and Frequency of Concession as Bargaining Strategies*, 7 J. PERSONALITY & SOC. PSYCHOL. 181 (1967).

⁸⁶ The author of one of these studies recognizes this problem but asserts that "the behavior of most subjects indicated the instructions emphasizing the experiment as a test of bargaining skills produced strong competitiveness." Chertkoff & Conley, *supra* note 85, at 185.

⁸⁷ See Hinton, *supra* note 85, at 198 (negotiation for the sale or purchase of information); Chertkoff & Conley, *supra* note 85 (negotiation for sale or purchase of a used car).

⁸⁸ Hinton, *supra* note 85, at 198.

⁸⁹ Chertkoff & Conley, *supra* note 85.

States as early as 1769, several states enacted summary judgment statutes based on the English model in the late 1800s. . . . Initially, judges expressed reluctance in granting summary judgment motions, viewing summary judgment as a drastic remedy. Yet by the mid-1920s, judges granted more than half of such motions before them.⁹⁰

The only support for the last sentence is a footnote citation to a 1929 law review article “relying upon a study of motions for summary judgment in New York.”⁹¹ This is the kind of thing that makes a social scientist grimace in pain. The sentence can be substantiated only by empirical evidence because it is a statement not about the law but about what people (judges) were doing. The only ways to find out what people were doing were to have watched them doing it or to collect after-the-fact evidence of what they did.

The social scientist’s gasps of pain are caused by two things. First, the authors of the block quote above do not cite to the study itself and do not discuss its methodology so that they and we can ponder its evidentiary value and limitations. And second, they make a statement about the entire United States, or at least all the states that had adopted these statutes, based on data from only one state, despite the differences in legal culture and judicial thinking from one region to another — which were much greater then, before the eras of easy communications and uniform statutes. In fact, how can we be sure the data were not drawn from a single courthouse? The authors have not told us enough to preclude that possibility. And when we go back to the source they cite, we find that in fact that most extreme of all possibilities is the case: the data came not only from a single courthouse, but from one that during the period in question may have been among the most atypical in the country.⁹²

So in a law review, we read that “by the mid-1920s, judges granted more than half of such motions before them.” Sweeping statements like that commonly appear in law reviews supported by a citation to something empirical. A social scientist working with the same raw materials might make this claim: “Although we have data for only a single county, and although that county was at the time the single most urbanized county, by a very wide margin, in the United States, the data do show that judges in that county, by the mid-1920s granted more than half of such motions before them.” And a social

⁹⁰ Jack H. Friedenthal & Joshua E. Gardner, *Judicial Discretion to Deny Summary Judgment in the Era of Managerial Judging*, 31 HOFSTRA L. REV. 91, 97 (2002) (footnotes omitted).

⁹¹ *Id.* at 97 n.24.

⁹² See Charles E. Clark & Charles U. Samenow, *The Summary Judgment*, 38 YALE L.J. 423, 455 n.230 (1929). It was the county courthouse on Foley Square in Manhattan.

scientist would be satisfied with this sentence only if it were followed by a discussion of where these numbers came from and why we can trust them to be accurate for the county they purport to represent.

That is a relatively simple example. Let's move to a more complex one that is closer to clinical teaching. In clinics, one of the most influential empirical studies is Douglas Rosenthal's book *Lawyer and Client: Who's in Charge?*⁹³ Rosenthal studied some personal injury disputes and in each one estimated the degree to which the plaintiff was involved in preparing the case, on a scale ranging from a passive client deferring to the lawyer as an authority figure (the traditional model of the lawyer-client relationship) to an active client working together with the lawyer to solve problems (the participatory model). Rosenthal came up with a creative research design, which, as a good social scientist, he described fully so that anybody who wanted to do a followup replication study could do so.⁹⁴ He asked "two plaintiff attorneys, a 'switch hitter,' and two insurance representatives" to assess the value of each case, out of which Rosenthal computed a single value using appropriate statistical tools.⁹⁵ To learn how the client dealt with the lawyer and vice versa as well as how both the clients and the lawyers viewed lawyer-client relationships in general, Rosenthal interviewed all of the clients and about a third of the lawyers and sent questionnaires to substantially all the lawyers. He learned the settlement amounts from court records.

Rosenthal found that, as a group, the more participatory clients got more money than the more passive ones. Rosenthal did not average the participatory clients' awards and compare that number in a single table with the average of the passive clients' awards. That is frustrating to the law-trained reader, but the omission is good social science. It avoids misleading precision (see below), which is a genuine risk because it is clear from the data that a different group of evaluators could easily have produced different assessments and because Rosenthal's assessment of the clients' degree of activity was necessarily subjective. Moreover, enough passive clients did better and active ones did worse than the evaluators' assessment of the cases that a table of their averages could have oversimplified the results. Instead, he differentiated only between cases where the client had gotten a good award in the evaluators' assessment and cases where clients had gotten a poor award. An award was good if it equaled 70% or more of the evaluators' mean assessment of the case, and it was poor if it was less than that. (The evaluators' means were more often than not

⁹³ DOUGLAS E. ROSENTHAL, *LAWYER AND CLIENT: WHO'S IN CHARGE?* (1974).

⁹⁴ *Id.* at 179-208.

⁹⁵ *Id.* at 202-07.

greater than the actual settlement.) By that very rough gauge, 75% of the active clients got good awards, and 59% of the passive clients got poor awards.⁹⁶ And Rosenthal went out of his way to treat it as a rough gauge.

Rosenthal also carefully admitted the limitations of his methodology and the tentativeness of the results. All of the cases came from a single atypical county. The clients tended to be more educated than the general population. Subjective judgments by Rosenthal, his five case evaluators, and the clients themselves permeate the data. Perhaps most importantly, the sample size was small, 57 clients from cases settled in 1969. Rosenthal also pointed to other limitations on the data.⁹⁷ He reported the results in a section of his book titled "*The Limits of Client Participation in Explaining Case Outcome.*"⁹⁸ And in good social science tradition, he published the book more to raise and define issues and invite further research than to provide conclusive answers.

Rosenthal's book was published in 1974. What has happened since then would, in two ways, strike social scientists as remarkable. First, when law teachers cite to this part of Rosenthal's book, they usually do so as though he has conclusively settled the question and without any hint of any limitations on the data or the inferences that can be drawn from it. Here is a typical law review reference to the book:

See, e.g., DOUGLAS ROSENTHAL, *LAWYER AND CLIENT: WHO'S IN CHARGE?* (1974) (demonstrating with empirical data that clients who actively participate in their cases often achieve better outcomes).⁹⁹

Second, there has been no followup replication study. Rosenthal was both a lawyer and a social scientist, and the book appears to have been written primarily for a law audience. If a replication study were to be done — and social science would not be satisfied without several of them — it would be by law teachers. The lawyer-client relationship is not something that many social scientists could be expected to understand well enough to research, depending as it does on an appreciation of ethics law and the practicalities of daily law practice. The absence of replication studies does not mean that we do not care about this issue. The participatory client is a core part of the idea of

⁹⁶ *Id.* at 56-57.

⁹⁷ For example, see *id.* at 186-190.

⁹⁸ *Id.* at 56 (*italics added*).

⁹⁹ Carrie Menkel-Meadow, *Toward a Theory of Reciprocal Responsibility Between Clients and Lawyers: A Comment on David Wilkins' Do Clients Have Ethical Obligations to Lawyers? Some Lessons From the Diversity Wars*, 11 *GEO. J. LEGAL ETHICS* 901, 904 n.13 (1998).

client-centered lawyering, which itself is a core part of clinical education.

The problem is that we have all been persuaded by Rosenthal's 57 clients who settled their cases in 1969 in a very atypical county. And we have ignored all of his warnings that his data and inferences were limited.

Why do lawyers and law teachers do these things? There are at least two possibilities. One is that we are in the habit of proving things. We like participatory clients, and our case in their favor is bolstered by a study that seems to show that they get better litigation settlements. Another example — maddening to social scientists — is the tendency of lawyers and law teachers to mistake correlation for causation. "Even in the first course in statistics, the slogan 'Correlation is no proof of causation!' is imprinted firmly in the mind of the aspiring statistician or social scientist."¹⁰⁰ The fact that two things usually occur together does not mean that one causes the other or that both are caused by some third thing. Causation might be present, but the correlation alone does not prove it. Nevertheless, the lawyerly urge to prove causation, so useful in trial work, leads us to miss all the precision and nuance by treating correlation as though it were proof of causation. After all, we have to have proof. We have an idea, either because we believe it or because our client needs us to persuade somebody of it, and we gather evidence to support it. In that frame of mind, lawyers use evidence to support their points. But an empiricist is obligated to consider the whole picture with an open mind. An empiricist investigates.

The other possibility is that many law teachers do not understand how empiricism works and in particular empirical standards of inference. We can explore this through one of the core issues of negotiation research. None of the following statements is accurate now, nor is any of them likely to become accurate during our lifetimes:

1. Empirical research on negotiation shows that making the first offer tends to cause a worse outcome from the offeror's point of view.
2. Empirical research on negotiation shows that making the first offer tends to cause a better outcome from the offeror's point of view.
3. Empirical research on negotiation shows that making the first offer tends not to affect the outcome.

There are arguments in support of all three of these positions,¹⁰¹ but

¹⁰⁰ Herbert A. Simon, *Spurious Correlation: A Causal Interpretation*, in *CAUSAL MODELS IN THE SOCIAL SCIENCES* 5, 5 (H.M. Blalock Jr. ed., 1971).

¹⁰¹ Generations of lawyers have believed that if you make the first offer in negotiation,

even if these arguments were supported by empirical research, none of them could settle the question. Five studies in which first-offerors did worse do not prove the first proposition. Nine studies in which first-offerors did better do not prove the second proposition. Twenty studies in which first offers did not correlate with results do not prove the third proposition.

Why will empirical studies not settle this? There are so many different kinds of negotiation — and each kind of negotiation is different from others in so many ways — that what is true in one kind of negotiation may easily be false in another. To find out whether making the first offer affects the result, we would have to study each different kind of negotiation separately — and not just once but several times in replication studies. But we have not even mapped all the different kinds of negotiation. And we might never be able to, remembering how varied the practice of law is among specialties, regions, and the various levels of stratification within the legal profession; how a category of negotiation can be defined not only through its subject matter but also through the needs of the clients and the characteristics of the negotiators; and how negotiation permeates not only the practice of law but also the lives and businesses of our clients.

Empirical studies could reveal some things about first offers, even

you are at a disadvantage because your offer will reveal your expectations before you have discovered the other side's expectations, which might be more favorable to you than your own. But that argument makes at least two unexamined assumptions. One is that revealing expectations hurts more often than it helps, or, put another way, that the other side is willing to give us more than we expect to get frequently enough that it pays not to make the first offer. The other assumption is that we cannot learn how to make first offers that are high enough to avoid this problem most of the time.

The argument in favor of making the first offer is that in many negotiations a first offer can anchor the rest of the negotiation to the offeror's favorite number or concept. In fact, anchoring first offers are all around us. For example, in a car dealership, an anchoring first offer is physically attached to every car in the form of the sticker price, an artificially inflated number, and the dealership maximizes revenue when people feel compelled to negotiate down from it. If you persuade the dealer to give you a "discount" of \$2,000 off the sticker price, you feel as though you have gotten a good deal even if you haven't. You can defeat the anchoring effect of this first offer by learning what the dealer paid wholesale for the car, estimating the dealer's overhead and a fair profit, totaling up those numbers, and offering to buy the car for that total — in other words, by substituting your own anchor and ignoring the dealer's first offer. The argument in favor of making the first offer, like the argument against doing so, makes at least two unexamined assumptions. One is that negotiators generally know how to make first offers that successfully anchor. The other is that the risks we run by making a first offer usually are outweighed by the anchoring benefit.

The argument that first offers do not affect the outcome is based on the fact that some studies show first offerors doing better; some studies show them doing worse; and some show mixed results. This argument is based on a fundamental misunderstanding of empirical methods. Equivocal results from studies do not necessarily mean that it does not matter what you do. Equivocal results from studies can mean that we still do not know.

if they will never be able to tell us whether first offers are productive (or counterproductive) in negotiation generally. A narrowly framed empirical inquiry might be able to reveal the effects of first offers in, for example, misdemeanor plea bargaining in high-volume urban courts with defendants represented by a public defender organization and where the judges and the prosecutor's office have developed certain specified attitudes toward plea bargaining. (But even with all those limitations, it would not be surprising to find the results of such a study cited in law review articles as "proof" of what happens to first offers in *all* plea bargaining or even in all negotiation.)

Each empirical inquiry can answer only one or a few questions, and often — though not always — the answerable questions are narrow ones. This can make lawyers and law teachers impatient. We are used to getting there quicker, although through a much looser kind of knowing. Not only do we want answers faster, but we want bigger answers than empiricism is ready to provide. But in all the hard and soft sciences, we can know more if we proceed more carefully and more slowly. Scientists can be comfortable saying "We don't know yet" for decades until empirical standards of inference are satisfied. Lawyers and law teachers might be uncomfortable saying those words, but that does not mean that empirical standards of inference are suspended where matters of concern to law and lawyers are concerned.

"A true causal relationship may have an extremely complex structure."¹⁰² Lawyers and law teachers often oversimplify this and look for one or two events or conditions that "cause" X. An empiricist would see a much wider range of factors that social science would say "could have a causative effect" on X because X could be the result of one of them, more than one acting separately and causing separate parts of X, more than one acting together to cause X as a whole, and so on. An empiricist would treat each of these factors as a variable and construct a methodology that strips out each variable one at a time to measure the causative effect of each of them on X. "One set of data alone will rarely, if ever, indicate the existence of a causal relationship."¹⁰³ And life being what it is, the bigger X is, the less likely is the possibility that any one variable could cause it.

2. *Using methodology to prove a point:* The inference problems discussed above are caused by a failure to turn off persuasion mode so we can think empirically. Lawyers and law teachers who use methodology to prove a point, however, go further; they construct a methodology intending to get certain results. But the goal of empirical work is not to prove a point. It is to find out. Empirical inquiry is detached.

¹⁰² W.J. REICHMANN, *USE AND ABUSE OF STATISTICS* 124 (1961).

¹⁰³ *Id.* at 126.

We should enter it because we find a problem or a question fascinating and want to know the answer. Doing empirical work is like reading a mystery novel: we can get pleasure by being surprised by the ending or by guessing it from clues along the way, but we are open-minded all along and are willing to accept any ending that is supported by the plot (or by the data and methodology).

3. *Failure to observe appropriate methodological controls:* These may include — depending on the research method being used — pretesting, control groups, double-blind data gathering, randomization, creating representative samples, and so on. Some empirical research methods, such as anthropological or sociological field work, are open-ended inquiries. In others, such as surveys and other forms of quantitative work, the researcher is testing an hypothesis. That hypothesis must be identified precisely. Rival hypotheses must be identified and tested as well. All the variables involved must be identified precisely, and the methodology must accurately measure them.

In virtually every form of scientific research, there is a risk that results could be anomalies caused by chance. Good empiricists worry a lot about that. They go to much effort to insure that samples are representative of larger populations, for example, and that experiments are conducted with subjects similar to people in the real world whose behavior the experimenters are trying to learn about. Many negotiation experiments have revealed little or nothing about real life negotiations because the experimenters used students as subjects rather than lawyers or business people or diplomats. It is hard to say whether students were used as subjects because they were cheap, because they are convenient, or because the researchers were not able to imagine much outside university settings. But the idea that 20- or 23-year-old students will see a negotiation the same way that 50-year-old professionals or business people will is surprising and is never proved empirically before these experiments are carried out.¹⁰⁴ Many other

¹⁰⁴ There seems to be much less cause for concern where law school clinical students are used in interviewing or counseling experiments, even where the client is role-played by an actor. Interviewing and counseling are not as complicated as negotiation, and clinic students seem to reach professional competence, particularly in interviewing, much more quickly than they do in negotiation. In most published negotiation experiments where the subjects were students, they were not clinic students or enrolled in some other post-graduate program, such as a business school, where they might have learned some negotiation skills. Moreover, in a negotiation experiment, the people on both sides of the conversation are usually subjects, while in an interviewing experiment only one person is while the other is either a genuine client or someone providing a standardized persona. Standardized patients, role-played by actors, are widely used in medical education and, starting in 2005, will be used in the national medical licensing exam. See Lawrence M. Grosberg, *Medical Education Again Provides a Model for Law Schools: The Standardized Patient Becomes the Standardized Client*, 51 J. LEGAL EDUC. 212, 213-19 (2001); Katherine S. Mangan, *Medical-School Graduates Will Face New Clinical-Skills Test*, CHRON. HIGHER EDUC., Feb. 6, 2003,

negotiation experiments are equally unrevealing because they ask the subjects to do things that are unrealistic, like negotiate for five minutes to divide tokens worth two cents each.¹⁰⁵ Although these negotiation experiments enter the legal literature either directly through publication in law reviews or indirectly by publication elsewhere and citation in law reviews, the experimenters tend to be either people with economics backgrounds or people associated with the law and economics movement.

Experiments are a popular way of studying negotiation. One justification is that a "controlled laboratory setting [allows] experimenters [to] hold constant the variables they are not testing."¹⁰⁶ But the results of these experiments often seem sterile and unpersuasive because laboratory conditions are often so oversimplified when compared to real negotiations where real people have to resolve real problems that affect them or their clients or both personally. It is possible to use real-world data to get at the same questions these experiments claim to settle, but it often takes a lot more effort as well as a more creative and subtle research design.¹⁰⁷

accessed at <http://chronicle.com/daily/02/2003020605n.htm>. Finally, interviewing experiments — such as Gellhorn, Robins & Roth, *supra* 53; Gellhorn, *supra* note 56; and Smith, *supra* note 59 — tend to use content analysis rather than quantification of results, as in negotiation experiments. In content analysis, allowances can be made for behavior that seems more student-like than professional. And in the Smith study, students were especially appropriate subjects because one of the purposes was to evaluate the effectiveness of a particular pedagogy.

¹⁰⁵ This example comes from Don L. Coursey & Linda R. Stanley, *Pretrial Bargaining Behavior within the Shadow of the Law: Theory and Experimental Evidence*, 8 INT'L REV. L. & ECON. 161, 167 (1988).

¹⁰⁶ Russell Korobkin & Chris Guthrie, *Psychological Barriers to Litigation Settlement: An Experimental Approach*, 93 MICH. L. REV. 107, 119 (1994).

¹⁰⁷ For example, compare the Korobkin & Guthrie study, *id.*, which used students in experiments, with Samuel R. Gross & Kent D. Syverud, *Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial*, 90 MICH. L. REV. 319 (1991) and Leandra Lederman, *Which Cases Go to Trial?: An Empirical Study of Predictors of Failure to Settle*, 49 CASE W. RES. L. REV. 315 (1999). Gross and Syverud studied empirically the same issue — the Priest-Klein theory of why disputes do not settle — by working backward from 529 civil trial verdicts and analyzing the offer and demand histories of the negotiations that preceded trials. Among the kind of results that Gross and Syverud obtained was this: in their sample, "only 15% of the trials . . . produced awards that were greater than the offer but less than the demand," Gross & Syverud, *supra*, at 379, which means that in the other 85% of cases trial did not find a resolution between positions that were too far apart to settle; instead, trial found a resolution outside the territory staked out by the plaintiff's final demand and the defendant's final offer. No experiment — no matter how well designed — could have produced that insight because no experiment can include a real trial. Lederman did something similar with 400 cases in the Tax Court, which were particularly useful "because, unlike most courts, the Tax Court keeps records of settled cases." Lederman, *supra*, at 327. In Lederman's sample, the cases that went to trial had identifiable characteristics that separated them from the cases where a negotiated settlement was possible. On the issue being studied by Korobkin and Guthrie, by Gross and

On similar grounds, Stewart Macauley criticized "research based on simulated trials before mock juries," not only because "it is difficult to recreate the essential features of the legal process without an expensive investment in theater," but also because "people in these experiments are playing the role of 'subject' rather than making decisions that count. While they may become involved in their experimental task, they never confront someone who may be hurt by their decisions and at some level they know that they are playing a game."¹⁰⁸

Good empiricists also attack their own hypotheses with the data to see if the hypotheses will hold up. "[I]n empirical research, challenging [one's] theory with the best possible opposing arguments is what makes the strongest case for a theory."¹⁰⁹ Why do law teachers sometimes have trouble doing this?

While a Ph.D. is taught to subject his or her favored hypothesis to every conceivable test and data source, seeking out all possible evidence *against* his or her theory, an attorney is taught to amass all the evidence *for* his or her hypothesis and distract attention from anything that might be seen as contradictory information. An attorney who treats a client like an hypothesis would be disbarred: a Ph.D. who advocates a hypothesis like a client would be ignored.¹¹⁰

And to study a given topic, some empirical methods will work better than others. Using a survey to ask what people do or why they do it will often not tell us what is really going on because people frequently flatter themselves by assuming that they behave consistently with norms they respect or believe that others respect.¹¹¹ The only reliable methods of finding out what people do is to watch them do it or develop objective after-the-fact evidence of what they have done. Surveys will much more accurately reveal what the respondents have seen others do or what the respondents want.

4. *Failure to recognize the value of replication:* This takes at least three forms. One is interpreting a study to have conclusively decided an empirical question, even if it is the only research that has addressed the question. A second form is the failure to attempt replication on the ground that the first person to have performed research has preempted the topic. And a third form is deriding other's attempts at

Syverud, and by Lederman, see George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984); George L. Priest, *Reexamining the Selection Hypothesis: Learning from Wittman's Mistakes*, 14 J. LEGAL STUD. 215 (1985).

¹⁰⁸ Stewart Macauley, *Law and Behavioral Sciences: Is There Any There There?*, 6 L. & POL'Y Q. 149, 158 (1984).

¹⁰⁹ See Epstein & King, *Rules of Inference*, *supra* note 21, at 10.

¹¹⁰ See *id.* at 9.

¹¹¹ SCHÖN & ARGYRIS, *THEORY IN PRACTICE*, *supra* note 28, at 6-7.

replication on the theory that replication studies are copycat work. We discussed these at the end of the preceding section of this article and will return to them later.

5. *Failure to explain one's methodology completely:* It is not enough to have perfect methodology. Published empirical research must describe it in exhaustive detail. In both the hard and the soft sciences, a description of the methodology may take up most of the article reporting the work. There are two reasons.

The first is that methodology is what makes the research credible. The reader who understands empiricism is on the road to being convinced when, nearing the end of the description of the methodology, that reader starts thinking things like "they really covered all the angles, accounted for all the variables, incorporated testing all the competing hypotheses," and so on. Although an empirical reader scrutinizes the methodology carefully, many readers who think in adversarial terms search the text for arguments and find the methodology boring.

The second reason is that an empirical researcher owes to everyone who might do replication studies a full explanation of the methodology so that later studies can really reexplore the same or similar ground.¹¹² A first study is an invitation to other researchers to attempt replication.

6. *Failure to identify the limitations of the study and of the inferences that can be drawn from it:* Here is another thing that readers and writers who think adversarially can find boring: all the nitpicky qualifying language that takes a potentially Big Empirical Inference and cuts it down to a small and undramatic one. Why can't we have a Big Empirical Inference so we can use it to support an Important Argument? Because science does not work that way. In both the hard and soft sciences, a researcher can look back on a lifetime of productive work if she has produced many studies that have added to knowledge in small, even tiny, but discrete ways. The handful of researchers who ever develop work that supports a Big Empirical Inference might as well be candidates for a Nobel or one of the lesser prizes. There are few such studies and few such researchers. The hard and soft sciences both progress in frequent but small steps.

There is still another reason for stating a study's limitations: "a basic premise of all empirical research — and indeed of every serious theory of inference — is that all conclusions are uncertain to a degree" because "the facts we know are related to the facts we do not know but would like to know only by assumptions that we can never

¹¹² Epstein & King, *Rules of Inference*, *supra* note 21, at 38-47.

fully verify.”¹¹³ For that reason, a good report of empirical research “estimate[s] the degree of uncertainty inherent in each conclusion.”¹¹⁴

Consider this, from the Sarat-Felstiner study of how divorce lawyers explain courts and the law to their clients:

[Our] lawyer samples have two obvious biases. . . . [T]hey involve a higher proportion of women than exists either in the bar or among divorce lawyers generally. Nevertheless, the samples contain more male than female lawyers. More importantly, the samples appear not to include many lawyers high in income, experience, and status; relatively few represented doctors, lawyers, businessmen, or others with substantial income and assets. As a general matter, these lawyers also attended less prestigious law schools than the lawyers considered to be at the top of local divorce practice. Thus the findings of this project should not be considered representative of all divorce lawyers. However, other than their relative status within the local bar, we know of no other relevant trait on which these lawyers are different from the rest of the divorce bar, and consider it fair to say that the findings do originate with a sample that is characteristic of lawyers that most people with ordinary financial resources are likely to consult.¹¹⁵

By law review standards, this is pretty good, but by social science standards it is still incomplete. Why should we not be troubled by the fact that the sample is not representative of the bar in gender terms? There probably is a good reason, but the researchers need to tell us about it rather than assume it away. They nicely confess that their sample is “not . . . representative of all divorce lawyers.” But why should we accept that “the findings do originate with a sample that is characteristic of lawyers that most people with ordinary financial resources are likely to consult”? The researchers are so careful about other things that we are tempted to assume this, but we should not have to. Good empiricism requires not an invitation to assume a proposition like this but instead data that really do substantiate it.

One of the ways that a researcher can set out a project’s limitations is to itemize the questions it does not answer, sometimes in the form of suggestions for further study. Here is the beginning of a good example, from the Smith client-centered interviewing experiment:

This study attempted to capture a typical successful interview by an inexperienced law student instructed in the client-centered approach. While some tentative conclusions can be derived from the findings here, they are incomplete in many respects.

Following this are five paragraphs setting out a host of questions the

¹¹³ See *id.* at 50.

¹¹⁴ See *id.*

¹¹⁵ Sarat & Felstiner, *supra* note 62, at 1670 n.41.

experiment did not answer but might profitably be explored in later research.¹¹⁶

7. *Confusing anecdote or nonmethodological observation with data:* Social scientists sometimes joke that data is the plural of anecdote — but that *is* a joke. To a social scientist, anecdotal evidence is not evidence at all because anecdotes have not been subjected to a methodology that tests them for accuracy and representativeness. Put another way, methodology is the reason why data is *not* the plural of anecdote.

That is hard for lawyers to accept because we prove our clients' cases through stories. Thus anecdotes, in the long form called stories, are our bread and butter. As lawyers, we naturally believe that when we have told an anecdote that illustrates our point, we have thus proved our point. When we are doing lawyer work, we are right in thinking that because proving is persuading and stories persuade. But when we are doing empirical work — when we are trying to discover what is really going on out there in the practice of law — the proof is in the data. Anecdote does have some value, but primarily as illustration and not as proof. Used for illustration, it helps the data come alive, but that role is supplementary.

Nonmethodological observation is observation without the benefit of methodological controls. Without methodology, generalizations drawn from observation have no assurance of validity or reliability. On cross-examination, lawyer A asks a question to which he does *not* already know the answer, and he hits the jackpot (the answer wins his case). He tries it again a few weeks later and gets good results once more. He sees a few other lawyers do the same thing and also get good results. Now he knows for sure that asking on cross questions to which he does not already know the answer is a good thing, and he advises all young lawyers and law students to do it. Over in the next county, lawyer B's experience is exactly the opposite, and lawyer B is telling everyone who will listen never to ask such questions. But neither A nor B really knows in an empirical sense because each of them have seen a tiny fraction of reality, and nothing in this picture has heightened the likelihood that what either of them saw was representative of trial practice in general. Nothing did that because neither of them used a methodology. They just saw some incidents and drew broad inferences from them.¹¹⁷

¹¹⁶ Smith, *supra* note 59, at 587-89.

¹¹⁷ In baseball, a prospect is a high school or college player who might be hired by a professional team. Scouts employed by major league teams are unable reliably to differentiate between talented prospects and untalented ones because of "the tendency of everyone who actually played the game to generalize wildly from his own experience. People

8. *Misleading precision:* Here is an example, involving government agencies and statistics. A fishery is the industry of catching a specific form of sea life in a particular region, such as the Alaskan salmon fishery, the Grand Banks cod fishery, or the Antarctic blue whale fishery.¹¹⁸ Many fisheries are legal and political battlegrounds between environmentalists, who protest that overfishing depletes sea life, and those who earn their livings in the fishery and who want it to continue unrestricted. The power to restrict a fishery might be exercised by government agencies, such as the regional fisheries management councils that answer to the National Marine Fisheries Service, or by an international organization, such as the International Whaling Commission.

The regulating entity will try to manage a fishery based on statistical evidence, the goal often being to permit the largest amount of fishing that will not cause the sea life involved to become commercially extinct. Commercial extinction is the disappearance of enough of a species that it no longer pays to try to fish it. (Biological extinction is the complete disappearance of the species.)

The statistical evidence is usually the tonnage caught the preceding fishing season or some shorter period of time, although for large creatures, like whales, it might also include the number of individuals caught and their average size. Most fish, however, are caught by the thousands in nets by factory trawlers, and it is impractical to count and measure them individually, although the total weight of a trawler's catch has to be determined to establish how much the total catch can be sold for in port. The regulatory process often involves interpretation of the statistics by marine biologists who specialize in the species being fished.

Some fisheries crash. Suddenly none of the creatures that were previously fished can be found anywhere in the fishery, and overnight the species has become commercially extinct there, or nearly so. The

always thought their own experience was typical when it wasn't. . . . [In addition,] there was the bias toward what people saw with their own eyes, or *thought they had seen*. The human mind played tricks on itself when it relied exclusively on what it saw." MICHAEL LEWIS, *MONEYBALL: THE ART OF WINNING AN UNFAIR GAME* 18 (2003). Baseball people in general subscribe to a number of myths about what indicates talent in a ballplayer, not very different from the myths about body language and truth-telling exposed in the studies reported by Wellborn. And just as reading the cold transcript has turned out to be a better way to separate accurate and truthful from inaccurate or untruthful witnesses, studying a baseball prospect's performance statistics — without actually seeing the prospect pitch, bat, or field — is a more reliable method of identifying a prospect who can play in the major leagues. See *id.* at 37-40. Seeing the prospect play can fool scouts, and "we aren't allowing ourselves to be victimized by what we see," declared the general manager of the first team to recognize this fact. *Id.* at 37.

¹¹⁸ See MARK KURLANSKY, *COD: A BIOGRAPHY OF THE FISH THAT CHANGED THE WORLD* (1997); GEORGE L. SMALL, *THE BLUE WHALE* (1971).

fisheries workers, now out of work, blame the regulators; the environmentalists blame both the fishing industry and the regulators; and the regulators point to statistics showing larger and larger catches every year until the point where the fishery crashed. Before the crash, the size of the catches had seemed to show that there were abundant fish in the sea. After all, to support catches of the size reported statistically, the species biomass — the total weight of all the individuals of that species under the surface of the sea — must have been huge. If there were trouble — if overfishing had been happening — the regulators would have expected to see declining catches for several years as fish became scarcer. Why didn't they? And what happened to the fish?

Each year, the statistics reported precisely, down to a fraction of a ton, the size of that year's total catch. Those numbers were probably accurate, down to a fraction of a ton. Part of the problem was exactly that precision — not because precision is bad (it never is), but because of the assumptions we can make when we see a precise number. One of those assumptions is that the precise number represents what is really going on out there, in this case under the ocean surface. Precision can induce the illusion of completeness — the illusion that the precise number tells the whole story. Here, the important number was not the one being reported (the total weight of a season's catch). The important number was the average size of the fish being caught, which would reveal their average age. But because it is impractical to measure individual fish size when thousands of fish are being hauled in huge nets into factory trawlers, nobody knew what was really going on.

What was really going on was that every year the fish caught were, on average, younger than the year before until, in the crash year, there were no fish of reproductive age and few fish of any age left. The catches had been abundant — but not because the fish had been abundant. In fact, the fish had stopped being abundant years before. The catches had been abundant because the new and relentless efficiency of modern factory fishing was cleaning out the fishery until, in the last year before the crash nearly all the remaining reproductive fish were captured — but nobody knew it at the time. The beautiful precision of the statistics deluded everyone, except perhaps the environmentalists, and even the people on the trawlers would not have noticed that the fish were getting smaller. They would not have noticed because so much of the process is now automated.

How can we avoid being fooled by precise numbers? Perhaps the most effective method is to hypothesize all the possible reasons why the numbers could have come to exist, and then to seek confirmation

or disconfirmation of each hypothesis. One of those hypotheses should almost always be that the numbers are inaccurate (even though that hypothesis would have been disconfirmed in the fishing example if catches reported to the regulators corresponded to the transaction records when fish were sold in port). A precise number is not necessarily an accurate number.¹¹⁹

V. WHAT PREVENTS US FROM DOING BETTER?

Why have law teachers (including us) produced so little empirical research on the skills of law practice? And why has some of what has been published included methodological faults? Here are some possible reasons, which we do not assert to be true but offer only as hypotheses that could be tested empirically:

1. *Lack of resources*: Computers and software such as SPSS¹²⁰ make quantification easier than it was in the past, and datasets are being generated and made available by governments and other researchers. But despite all this, "data are a luxury good"¹²¹ because "data-gathering is frequently labor-intensive and time-consuming, and consequently, often quite expensive."¹²² And very few law schools and very few grantors are willing to pay for it. In fact, "[w]hat limited funding is available is heavily skewed toward law and economics."¹²³

We also lack research assistants with the kind of empirical skills normally found among social science graduate students. For a statistical study of women in legal education,¹²⁴ one of us (Neumann) did all

¹¹⁹ Baseball is the most statistical of all sports, but of the precise and apparently complete offensive statistics published every day in newspapers — batting average (carried out to three decimal points), home runs, runs batted in, runs scored, doubles, triples, and stolen basis — none of them is as reliable an indicator of a player's offensive value as some statistics that are almost never reported in the press — on-base percentage, slugging average, and number of pitches per plate appearance. LEWIS, *supra* note 117, at 34, 57, 127. A team that ignores the published statistics and plans its operations based on the unpublished ones (the Oakland Athletics) can have one of the best winning records in baseball with one of the smallest payrolls. *Id.* at xi-xii. Nearly everyone else in the sport is fooled by the precision and apparent completeness of the published statistics. "[I]f gross miscalculations of a person's value could occur on a baseball field, before a live audience of thirty thousand, and a television audience of millions more, what [does] that say about the measurement of performance in other lines of work?" *Id.* at 72. How certain are we that we can *really* tell an effective lawyer from an ineffective one, or a promising law student from an unpromising one? Our standards of measurement — law school grades, verdicts won, apparent client satisfaction, reputation, income, status of employer — could be just as faulty as batting average (carried out to three decimal points), etc.

¹²⁰ *Statistical Package for the Social Sciences*.

¹²¹ Rhode, *supra* note 79, at 1353.

¹²² Heise, *supra* note 15, at 829.

¹²³ Rhode, *supra* note 79, at 1353. Rhode identifies the Olin Foundation as "an extraordinarily generous funder of law and economics research." *Id.*

¹²⁴ Richard K. Neumann, Jr., *Women in Legal Education: What the Statistics Show*, 50 J.

the coding and nearly all the data collection personally because law school research assistants, though admirably careful and diligent, lack the background for such a project and could help only with limited and narrowly defined tasks. Coding is placing something into a category so it can be counted, usually by marking it with a letter or number code. It may be the most tedious work in quantitative research. Although there may be projects where coding can be left to students, it is often risky to do so because coding errors introduced by students can destroy the validity of data. Coding over 20,000 faculty entries from three years of AALS Directories of Law Teachers according to detailed criteria¹²⁵ could be done reliably only by someone with experience in and instinct for the nuances of faculty rank and job titles across the broad range of legal education. (They are not as standardized as they might seem.) It is hard to imagine any student at any law school with that kind of experience and instinct. Law school research assistants are good for what we train them to do: finding authority in the library, reading it to see whether it's relevant to a doctrinal project, and telling us about it orally or in writing. We do not train them in empirical methodologies or in the details of many things we might study empirically. "Unlike our colleagues in other disciplines, we cannot easily slough" the details of empirical work "off on acolytes" who would want to learn empirical methods from us so they themselves can get academic jobs practicing them.¹²⁶

2. *Lack of time:* To achieve a publishable result, empirical work requires tremendously more time and effort than doctrinal work. And a lot of it "is grunt work"¹²⁷ and "drudgery."¹²⁸ Compared to empirical work, doctrinal scholarship can seem quick, with a clear end in sight. Empirical work, on the other hand, can expand and keep on expanding. Deborah Rhode's "desperate efforts to compile computer printouts under a Christmas tree remain a still-resented entry in family folklore."¹²⁹

Our professional lives are built around the courses we teach and our service obligations, with an allotment set aside for scholarship. The size of that allotment depends on our faculty status and the traditions of our law school. It is a variation, in some ways, of a standard-size allotment initially determined generations ago when all faculty taught doctrinal courses and nearly all legal scholarship was doctrinal.

LEGAL EDUC. 313 (2000).

¹²⁵ They are itemized, *id.* at 330-32.

¹²⁶ Schuck, *supra* note 13, at 331.

¹²⁷ *Id.*

¹²⁸ Rhode, *supra* note 79, at 1353.

¹²⁹ *Id.* at 1354.

It is measured and thought of in integers of doctrinal articles produced in a certain time-frame. Doctrinal analysis "is the cheapest, easiest, and quickest way for authors to get into print," and it "requires no special expertise beyond what is available through traditional law school training."¹³⁰

An interesting empirical project would be to test with data the hypothesis that empirical articles take more author time to produce than doctrinal articles and, if that hypothesis turns out to be verifiable, to measure the difference. Nobody has published the results of such a study. But if that hypothesis were confirmed and the difference measured, and if law schools and scholars were willing to make adjustments in job definitions and personal lives to accommodate the greater time needed to produce empirical scholarship, what would be adjusted to make room. Would professional standards allow us to publish less frequently? Would we teach fewer courses or do less service or both? Would law schools hire nonteaching social science specialists to help us do the research? Or would we end up computing data under the Christmas tree?

3. *Lack of respect*: What generates prestige in the legal academy is the doctrinal or theoretical article that causes other academics to say "Wow!" even if no judge, legislator, or other policy maker would care about it or even understand it. "In most law schools," empirical research "is endorsed in principle but devalued in practice."¹³¹ According to Deborah Rhode, that is why law "has almost no academic journals specializing in empirical work on the legal profession and legal process."¹³² Rhode identifies some "prominent exceptions" as *Law and Social Inquiry* (previously known as *American Bar Foundation Research Journal*), the *Law & Society Review*, and the *Journal of Legal Studies*. Some people would add the *Journal of Law & Economics* and the *Journal of Law, Economics & Organization*,¹³³ but, for reasons stated later in this article, law and economics scholarship is not genuinely empirical. To some extent, we might add the *Clinical Law Review* and the *Journal of Legal Education*. Perhaps the largest difficulty in operating an empirical scholarly journal is that the standard scholarly journal staffing model in law is student editing and student management, and an empirical journal cannot be staffed by students. Students have no training in empirical scholarship because the standard law school curriculum does not give it to them. And even if the curriculum were otherwise, a student-edited empirical

¹³⁰ *Id.* at 1339.

¹³¹ *Id.* at 1342.

¹³² *Id.* at 1343.

¹³³ See Epstein & King, *Rules of Inference*, *supra* note 21, at 16.

journal would present quality control problems of the kind that caused the Supreme Court to list “[t]he fact of publication (or lack thereof) in a *peer reviewed* journal” as one of the factors to be considered in separating admissible scientific evidence from junk science in a federal trial.¹³⁴

Law faculties tend not to respect the small steps that are the hallmark of empirical thinking. Because empiricism cannot answer — and nothing can answer — the big questions like whether making the first offer in negotiation generally gives you an advantage or a disadvantage or neither, empirical research will focus instead on much smaller inquiries, such as the effect of first offers in plea bargaining in a medium-sized suburban county with a moderate crime rate and no single provider of legal services to indigent defendants. A significant portion of one’s faculty colleagues, naive about empiricism and used to the quicker payoffs of doctrinal scholarship, will deride this research for failure to Answer a Big Question Conclusively. Some of them will see the narrowness of the inquiry not as an understanding of and respect for the method and process of science, but as evidence of limited intellect. Consequently, “a single empirical study might be less fulfilling to the author than an article laying out a theory that seamlessly and completely resolves a legal problem on its own, with no questions left open”¹³⁵ — even if no policy maker ever considers adopting that theory and even if nobody (including the author) really knows how that theory would work in real-world conditions.

And the payoff can be unpredictable. With doctrinal research, you can tell early in the research whether the idea or angle you are pursuing will lead to interesting results. But with empirical research, you will not know until near the end what your results will be or whether others might consider the project a worthwhile investment. If your results resemble those of an earlier researcher who studied a different population and used a different methodology, at least some doctrinal colleagues will dismiss what you have done on the ground that it has not added to knowledge — even though science would consider you to have done something valuable. If your results are inconsistent with those of an earlier researcher, some colleagues may consider you to have proved the other person wrong and may be disappointed if you point out that the scientific method does not necessarily consider you to have done that, but only to have helped point the way to further research and a refinement in research methods. You and the earlier researcher might both be right, but in different ways or

¹³⁴ *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 594 (1993) (italics added).

¹³⁵ Korobkin, *supra* note 19, at 1055.

about different studied populations. Or, worst of all, your results might be inconclusive, which could lead some colleagues to say that you have wasted your time, even though inconclusive results are an ordinary risk of science.

The ease with which the law and economics movement has been received into legal education illustrates the extent to which the legal academy does not demand empirical evidence in support of assertions that should be empirically verifiable. Neoclassical economics (together with its offshoot law and economics) is not empirical because it does not insist that its assumptions — such as that economic actors make their decisions rationally and not out of emotion, altruism, or foolishness — be verified through replicated empirical studies. It instead simply makes those assumptions and draws inferences from them. The result is what Deidre McCloskey and others have called “blackboard economics” and “imaginary economies.”¹³⁶

Economists enamored of pure markets begin with the theory and hang models on assumptions that cannot themselves be challenged. The characteristic grammatical usage is an unusual subjunctive — the verb form “must be.” For example, if wages for manual workers are declining, it *must be* that their economic value is declining. If a corporate raider walks away from a deal with half a billion dollars, it *must be* that he added that much value to the economy. . . . How do we know that? Because to do anything else would be irrational. And how do we know that individuals always behave rationally? Because that is the premise from which we begin.¹³⁷

Science operates differently. In both the hard and soft sciences, a scholar is required to generate the largest possible number of explanations for a phenomenon (why manual wages are falling, why the corporate raider got the half billion dollars), and then to examine each possible explanation through experimentation or other methodologically controlled investigation. In science, there is no *must be*. Science left it behind when it outgrew superstition (“the gods have ways of showing displeasure; if the sky is filled with lightening and thunder, it *must be* because the gods are angry with us”). If the history of thought had been otherwise, a substantial amount of the scientific knowledge and technological advancement we now enjoy would not exist. Thus, to a person who has absorbed scientific standards of knowledge, a field that claims to describe how people behave in the real world but never subjects its views to empirical testing is either

¹³⁶ Peter Monaghan, *Taking on “Rational Man”: Dissident Economists Fight for a Niche in the Discipline*, CHRON. HIGHER EDUC., Jan. 24, 2003, at A12.

¹³⁷ ROBERT KUTTNER, *EVERYTHING FOR SALE: THE VIRTUES AND LIMITS OF MARKETS* 9 (1997) (*italics in original*).

amateurish or anti-intellectual.

Not only is a vast amount of market behavior irrational, but entire industries are devoted to taking advantage of that fact. What percentage of the population never carries credit-card debt at interest rates once considered usurious but instead does the rational thing: live within a household budget and defer gratification for any purchase that cannot be folded into an original or refinanced mortgage, with its low interest rate and tax advantages? Academics who believe in rational choice theory appear never to have heard of the millions of investors who buy into mutual funds after those funds have peaked, at a time when funds generally decline, even though those investors have all been told many times that past performance is no guarantee of future results. The eternal springing of hope is a genteel way of considering oneself subject to Barnum's first law of markets (the one that postulates a birth rate for irrational suckers).

Law and economics adherents sometimes sound remarkably naive when they discuss how real markets actually have behaved or will behave.¹³⁸ People who do very well in the stock market sometimes say that when everyone is buying, it is time to think about when to start selling, and when everyone is selling, it is time to think about when to start buying — because the crowd, after a certain point, is usually wrong, buying in a frenzy at more than stocks are worth or selling in a frenzy at less than stocks are worth. In fact, the economic phenomenon that most characterized the last decade — one that was obvious to any academic who compared her retirement account statements during the boom and its aftermath to the average price/earnings

¹³⁸ See, for example, *Union Carbide Corp. v. U.S. Cutting Service, Inc.*, 782 F.2d 710, 716 (7th Cir. 1986) (Posner, J.) ("most financial planners believe that tax-free [retirement] assets should be invested in interest-paying securities [bonds, money-market funds] rather than in common stock" because of the higher tax rates applied to interest). The words "most financial planners believe" represent something unknowable without empirical evidence such as survey research, of which Posner had none to cite. Nor could he have, at least if he meant to refer to financial planners who actually understand how investment markets work in the real world. First, tax treatment is only one of many factors that inform a sound investment strategy. Some taxed or more highly taxed investments can produce returns so much greater than untaxed or lightly taxed alternatives that the latter's tax advantage disappears. That is often, though not always true when taxable stocks or corporate bonds are compared with tax-free municipal bonds. Second, since the beginning of reliable investment market statistics in 1926, stocks have always outperformed bonds over the long periods of time that typically elapse between investing for retirement and living off the investment during retirement. And third, even an ordinary investor such as a university professor or a federal judge who acted on Posner's idea during the 18-year-old bull market in stocks that began in 1982 would retire with a fraction of the assets of a similar investor who was mostly invested in stocks throughout that period, even taking into account the downdraft that began in 2000. By 1986, one might have suspected not only ignorance but incompetence in a financial planner who suggested investing retirement assets *entirely* in bonds or money market funds.

ratio of the companies in the Standard & Poor's 500 — was summed up by Alan Greenspan as “*irrational exuberance*.”

In 2000, French economics students organized large protests and complained — with the French knack for being simultaneously politically correct and politically incorrect — that neoclassical economics is so unrealistic as to be “autistic” and engages in “the continuous construction of imaginary worlds.”¹³⁹ Since then, the post-autistism economics movement has spread to Britain and the United States.¹⁴⁰ A number of books have been published attacking neoclassical economics on similar, though less rhetorical grounds.¹⁴¹ And while all of this has been happening not a single published article by a law and economics adherent has acknowledged, much less tried to deal with the controversies now going on inside the field of economics. At the same time, a subfield called behavioral economics has developed to explore the extent to which psychology can explain economic behavior.

4. *Career pressures*: Until we are tenured and promoted to full professor, we are under pressure to produce standard-sized scholarship allotments in the time-frames recognized by our law school's tenure and promotion rules, whether they be written or local custom. And afterward, most deans will measure our performance for research grant and merit pay purposes according to those same standard-sized scholarship allotments and time-frames — which originally developed as units of measurement for the typical doctrinal article in size and speed of production. Thus, the law school academic career path, is not very amenable to the investment in time, effort, and delayed results inherent in scientific investigation.

In the face of these pressures, many faculty reasonably consider the ratio of time and work to published pages in the social sciences to be unsatisfactory. “A review of the projects launched during the height of legal realism at the Yale Law School found that the median duration was five years. Particularly for younger faculty, who often face pressure to demonstrate immediate productivity, studies requiring several years are scarcely attractive competitors to doctrinal

¹³⁹ James K. Galbraith, *A Contribution on the State of Economics in France and the World* in *THE CRISIS IN ECONOMICS: THE POST-AUTISTIC ECONOMICS MOVEMENT: THE FIRST 600 DAYS* 47, 48 (Edward Fullbrook, ed. 2003). See also *The French Student's Petition: Open Letter from Economics Students to Professors and Others Responsible for the Teaching of this Discipline*, in *id.* at 13 (“We wish to escape from imaginary worlds”).

¹⁴⁰ The documents and an on-line scholarly review are at www.paecon.net. See also Monaghan, *supra* note 136, at A12.

¹⁴¹ See, for example, GEOFFREY M. HODGSON, *HOW ECONOMICS FORGOT HISTORY: THE PROBLEM OF HISTORICAL SPECIFICITY IN SCIENCE* (2001); TONY LAWSON, *ECONOMICS AND REALITY* (1997); DEIRDRE N. MCCLOSKEY, *THE RHETORIC OF ECONOMICS* (2d ed. 1998); DEIRDRE N. MCCLOSKEY, *THE VICES OF ECONOMISTS — THE VIRTUES OF THE BOURGEOISIE* (1996).

analysis."¹⁴²

5. *Lack of training:* The social science research methods described earlier in this article demand complex and subtle research skills as well as time and money. Some of those skills can be learned by a law school teacher with a social science background and aptitude. But in the social sciences, a researcher is not considered credible without a Ph.D. because post-graduate education trains not only in the topics addressed by a given social science but also — and perhaps primarily — in the methods of research used in that social science. In many social sciences and with some — but not all — empirical research methods, the work of using those methods to produce a successful Ph.D. dissertation tends to be what turns an aspirant into a professional researcher.

Without that background, we are apt to make methodological errors with some — but not all — empirical research methods. On the other hand, researchers with only a social science background (and no legal training and experience) often produce scholarship that cannot be understood outside their social science or that suffers from naiveté about law and its institutions. If a researcher is not both a law teacher and a social scientist, collaboration between social scientists and law teachers can enrich almost any empirical project, although some empirical research techniques can be conducted in a methodologically sound manner by a law teacher who is willing to study methodology skills before using them.¹⁴³

6. *Fear of numbers:* You start to read an article with an interesting title on an interesting subject. You read the author's footnote and the rest of the first page and are still hopeful that the article might be worthwhile. You turn the page and find Table 1, which is a sea of numbers. If at this point you feel any of a range of negative emotions associated with numerophobia or innumeracy, that alone does not mean that you are unfit for empirical work. Even though numbers

¹⁴² Rhode, *supra* note 79, at 1353-54, citing to John Henry Schlegel, *American Legal Realism and Empirical Social Science: From the Yale Experience*, 28 BUFF. L. REV. 577, 582 (1979). See also SCHLEGEL, *supra* note 12.

¹⁴³ One way to get a sense of whether you feel capable of learning a particular social science method would be to go to a university library, pull a lot of books on methodology off the shelves and spend a few hours going through them. The following are the Library of Congress catalog numbers where we found books on various social science methods clustered in our university library, although cataloguers in other libraries might have handled some things slightly differently: data collection and management at H 61 to H 62; survey methods at HA 31 and HN 29; social science experiments (that try to simulate reality) at H 61 to H 62; obtaining empirical data through interviewing at H 61 to H 62, HM 48, and HN 29; sociological fieldwork methods at H 61 to H 62 and HM 48; anthropological fieldwork methods at GN 345 to GN 347; experimental economics at HB 131; statistics and statistical inferences at HA 29.

can often tell us things that words cannot, there are several empirical methods — such as field observation, content analysis, and historical or comparative analysis — in which numbers are not dominant. (This is not the same problem as when the reader turns a page and finds formulae filled with Greek letters that the author assumes need little or no explanation. If those formulae are derived solely from the writer's hypotheses and assumptions, as they usually are — rather than on methodologically organized observations of events or conditions — it is the *writer*, not the reader, who is having trouble with empiricism.)

And if you are neither numerophobic nor innumerate but are not skilled at statistical techniques such as regression analysis, that lack of skill does not foreclose all quantitative work. In a number of instances, valuable quantitative work can be done without regression analysis. Social scientists say that in some situations regression analysis adds nothing to the inferences one would get from a simpler presentation of the data. And occasionally it leads in false directions when a researcher substitutes statistical gymnastics for an astute understanding of the human and social elements of a problem.¹⁴⁴ Consequently, “an informed consumer would approach studies relying on multiple regression cautiously because too many researchers apply the technique mechanically without concern for its limitations.”¹⁴⁵ Peter H. Schuck wrote that a reader of a draft of his article on why law teachers do not do more empirical research “observed that people who do regressions are often insensitive to” what Schuck called the “messiness” of facts.¹⁴⁶ “If that is true,” wrote Schuck, “it serves to remind us that any methodology is simply a tool, which can be employed well or poorly.”¹⁴⁷

7. *Lack of sympathy with the scientific desire to know, even if the results conflict with our preconceptions:* The state of knowing that one does not know is valuable in a scientist, to whom it can be a happy frame of mind because it is the beginning of identifying a problem to be solved and developing a methodology for solving it, leading to years of productive work. But that seems less true of lawyers. We are supposed to know everything and be prepared for everything. And if we have preconceptions — if we want the research to prove something — our desires can deprive us of the frame of mind needed to inquire scientifically. That can account for methodological errors. Or,

¹⁴⁴ For example, see Aaron Wildavsky, *What Would One Have to Believe to Believe in Multiple Regression*, 27 CONTEMP. PSYCHOL. 903 (1982).

¹⁴⁵ Macauley, *supra* note 108, at 160.

¹⁴⁶ Schuck, *supra* note 13, at 325.

¹⁴⁷ *Id.* at 325 n.6.

as in law and economics, it can be one of the causes of a complete aversion to empirical research.

8. *The culture of argument:* It is often easier to prove something than to know it, and we are in the habit and business of proving. Law teachers spend so much time and effort arguing and teaching the skills of argument that, unless we can turn these dynamics off — unless we can completely leave persuasion mode and enter and be comfortable in a parallel universe where everything is done in inquiring mode — we will not be able to do empirical research in a methodologically sound way. Part IV of this article illustrates several ways in which the urges to argue and to assemble proof from which to argue can prevent us from knowing in a scientific sense. Arguments tend to oversimplify complex situations, and the urge to argue prevents the constructive awareness of ignorance.

This problem permeates law school faculties. It would be unempirical, however, to support that statement by referring to the universal experience of listening over and over again in faculty meetings to people stating opinions as though they were established fact. The empirical way to support it would be not to set out to prove it, but instead to wonder what faculty meeting talk reveals about law teachers; tape-record a number of faculty meetings at a variety of law schools, with controls to insure representativeness of the sample; and do a content analysis, coding each comment into categories of discourse, one of which would be persuasion-mode; quantify the results by number of comments, number of speakers, and time consumed in each category; subdivide the quantifications to show the proportion of statements in each category of discourse made by senior tenured faculty, mid-level tenured faculty, untenured faculty, off-tenure-track faculty, etc.; select representative comments to illustrate each of the categories of discourse; and then do the whole thing over again with several other kinds of meetings — of physicists, of nuns, of drama teachers, of business people, and so on — as controls. (This example, by the way, illustrates two things that are not often mentioned in the hard and soft science literature on methodology. First, there is such a thing as empirical imagination. The ability to put together a methodology that explores something in a novel and revealing way is as creative as anything one encounters in the arts. And second, it is not inevitable that published reports of empirical work have to be boring. Some research designs can produce material that Twain, Veblen, or Mencken would have been happy to work with.)

Is it possible for law teachers to turn off argumentation and enter the parallel universe of inquiry? Yes, but it takes self-discipline and a willingness to stop thinking like a lawyer and to start thinking like a

scientist.

9. *In our daily work, we draw inferences about the nature of lawyering and are satisfied by those inferences — even though they are anecdotal.* Our standard for “knowing” puts us into a position where we can be fooled by what we see. What medicine calls the clinical experience of the doctor is enormously valuable in diagnosis. The more a doctor sees of disease, the more a doctor gains an intuitive understanding of what inquiries, examinations, or tests will most quickly and reliably confirm or disconfirm diagnoses as well as how to identify illness that is so dangerous as to require emergency treatment.

But in prescribing treatment, clinical experience can hurt as well as help. Suppose some doctors have claimed they get good results when they use medication X to treat disease Y. A few months ago, your doctor heard these reports and decided to try X the next time she encounters Y. She does so four or five times, and in each instance the patient recovers. Now you are sitting in the doctor’s office with symptoms of disease Y, and your doctor says that in her experience medication X cures Y. Is that really her experience? How does she know that medication X actually caused the recoveries? How does she know that the earlier patients would not have recovered anyway and just as easily without X? How does she know that the patients did not recover *in spite of* X, which might in fact have prolonged the disease? How does she know that the patients even recovered at all? If X failed to cure the disease, they might have given up on your doctor and consulted someone else, who prescribed medication Z instead. And even if medication X did actually cure those four or five patients of disease Y, does that mean that it is more likely than not to cure you? If X cures Y only a small fraction of the time — say, 20% — the four or five preceding patients may not predict continued success but only the beginning of many failures.

You would want to hear from your doctor something like “I’ve read a half dozen studies, using all the appropriate controls and involving a total of several hundred patients, and the studies show that between 80% and 90% of the time, X cures Y.” Although that would increase enormously your confidence in the treatment, you probably will not hear it or anything like it, however.

If physicians can be fooled by what they see, we can, too. We tell students not to ask, on cross-examination, questions to which they do not already know the answer. That makes sense. We have probably seen a cross-examiner defeated by an unexpected answer, and we are persuaded by the stories, probably apocryphal, about the severed

ear¹⁴⁸ and about the moonlight.¹⁴⁹ But all that tells us is that there is risk when we ask a question on cross without already knowing the answer. It does not tell us whether the risk is worth running — or, more precisely, *how* to tell *when* that risk is worth running.

Why are we rarely haunted by the fear that we might not know what we are talking about in these situations? Because we are willing to accept anecdotal confirmation of what we believe or want to believe is true. The progress of medicine is a continual abandonment of treatments that doctors believed were justified by their clinical experience in favor of other treatments that were proved effective in empirical studies. Writing about psychology, Carol Tavis pointed to

two core elements of the scientific method . . . [that] are almost entirely lacking in the training of most psychotherapists The first is skepticism: a willingness to question received wisdom. The second is a reliance on gathering empirical evidence to determine whether a prediction or belief is valid. You don't get to sit in your chair and decide that autism is caused by cold, rejecting, "refrigerator" mothers, as Bruno Bettelheim did. But legions of clinicians (and mothers) accepted his cruel and unsubstantiated theory because he was, well, Bruno Bettelheim. It took skeptical scientists to compare the mothers of autistic children with those of healthy children, and to find that autism is not caused by anything parents do; it is a neurological disorder.

The scientific method is designed to help investigators overcome the most entrenched human cognitive habit: the *confirmation bias*, the tendency to notice and remember evidence that confirms our beliefs or decisions, and to ignore, dismiss, or forget evidence that is discrepant. That's why we are all inclined to stick to a hypothesis we believe in. Science is one way of forcing us, kicking and screaming if necessary, to modify our views.¹⁵⁰

10. *In our daily work, we are required to have knowledge that cannot wait for empirical investigation.* We have to teach something every day to students who will soon need that knowledge to practice

¹⁴⁸ Q: You said on direct that the defendant bit off the victim's ear. Did you see him bite the ear?

A: No.

Q: So how do you know he did it?

A: I saw him spit the ear out. [Or: The defendant bragged to me afterward that he bit it off.]

¹⁴⁹ Q: You said on direct that you are sure you saw the defendant. How could you have seen him, when you admit that you were 30 feet away and it was midnight?

A: It was a full moon, which was shining directly on the defendant's face. There wasn't a cloud in the sky, and we were standing in an open field. Besides, he called out to me and told me who he was.

¹⁵⁰ Carol Tavis, *Mind Games: Psychological Warfare between Therapists and Scientists*, CHRON. HIGHER EDUC., Feb. 28, 2003, at B7 (*italics in original*).

law, and we can't say to them, "We're doing a lot of experiments, and in ten years we might know whether it is true that you should never ask a question on cross-examination to which you do not already know the answer, and if you sometimes can, we might even know when those times are." The same problem infects even medicine, which, though scientific, is still a profession that must act now.

VI. TEACHING STUDENTS EMPIRICAL METHODOLOGY

While this essay has illustrated how the use of empirical methods can provide important insights for clinical scholars into the practice of law, the teaching of these methods to students in clinical courses can also enhance their understanding of their practical experiences. Some commentators have encouraged the training of students in the standards and norms of empirical research to provide resources to law school faculty in their research and to increase the opportunities for publication of empirical pieces in student-run journals.¹⁵¹ They also suggest that expanded course offerings in empirical methods would require the hiring of empirical methodologists in law schools who would also serve as a resource to the entire faculty in improving their research.¹⁵² But besides the crass self-interest of benefitting faculty in their empirical scholarly pursuits, in several ways the training of students about empirical methodology could improve their learning about the practice of law.¹⁵³

First, empirical training would provide students with the tools for working with expert witnesses and studies. The use of expert consultants plays an ever-increasing role in both dispute resolution and transactional practice.¹⁵⁴ Knowledge of general empirical methodology and standards lessens an attorney's learning curve in communicating with both one's own experts and those employed by the other party. Moreover, even in those cases in which a lawyer does not retain an expert consultant, training in empirical methods can assist practitioners in assessing the validity of government, scientific, and

¹⁵¹ Epstein & King, *Rules of Inference*, *supra* note 21, at 117.

¹⁵² *Id.* at 117-18.

¹⁵³ In a recent essay, two commentators question the importance of empirical methodology to legal scholarship because "[t]he legal academy supplies vocational rather than scientific training; law schools usually produce lawyers, not graduate students" Jack Goldsmith & Adrian Vermeule, *Exchange: Empirical Research and the Goals of Legal Scholarship: Empirical Methodology and Legal Scholarship*, 69 U. CHI. L. REV. 153, 155 (2002). This criticism ignores the fact, however, that scientific training can be very useful in the vocational setting of the practice of law.

¹⁵⁴ Even after *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), the Technical Advisory Service for Attorneys lists more than 9,400 categories of expertise on its website. Technical Advisory Service for Attorneys, *Directory of Expertise*, available at <http://www.tasanet.com/expertis.cfm>.

policy reports which are relevant to the particular proceeding.

Second, exposure to empirical methodology can assist lawyers in their continuing education both about their profession and their areas of practice. With the proliferation of social science studies on a wide variety of areas of the law,¹⁵⁵ it is important that attorneys have the basic knowledge to evaluate the strengths and weaknesses of their findings. Much legal empirical scholarship is prone to “armchair empiricism”: “[s]loppy survey techniques, skewed samples, and sweeping generalizations from unrepresentative findings.”¹⁵⁶ Without training in empirical methodology, attorneys are unable to examine critically studies about their professional practice.

Finally, training of students in the basics of empiricism in a clinical setting can be very helpful in teaching them effective methods for fact investigation. In our own clinical teaching, we have found that fact investigation is one of the most difficult skills for students to learn. They usually want a quick answer (often from an Internet search), become frustrated when they are unsuccessful, and, even if they are somewhat successful, terminate their investigation prematurely. Moreover, they often approach fact investigation in persuasion mode, tailoring their inquiries to the one or two theories they have identified without considering alternative hypotheses.

Instruction in rudimentary empirical methods can address many of these problems. For example, training in the importance of open-minded, scientific inquiry can help students learn how to consider rival hypotheses, reject easy answers, and approach fact investigation as an on-going process in which results should not be accepted until they have been proven. Knowledge of empirical methods for obtaining unbiased data can assist students in considering credibility issues raised by facts they have discovered. Instruction in methods of critical empirical analysis can help students become more adept at assessing the inferences they draw from circumstantial evidence and for considering the strength of their “chains of evidence.” And learning the value of replication for empirical research can assist students in understanding the importance of corroborating evidence supporting the facts discovered. In short, educating students in the basic features of empirical methodology helps them to approach the fact investigation process as scientists, not advocates. (Obviously, as lawyers, we are advocates, and at some point in the fact investigation process, we need to consider discovery of those facts most persuasive for our clients and dam-

¹⁵⁵ See generally Margaret Brinig, *Symposium: Empirical and Experimental Methods of Law: Empirical Work in Family Law*, 2002 U. ILL. L. REV. 1083; Korobkin, *supra* note 19, at 1033.

¹⁵⁶ Rhode, *supra* note 79, at 1343.

aging to the other party. Training in empirical methodology should help students understand that this process is best accomplished in both persuasion and inquiring modes.)

CONCLUDING THOUGHTS

Many clinicians recognize the importance of the inquiring mode to good planning in lawyering, and many of us try to teach it, implicitly or explicitly, to our students. But a number of dynamics discourage us from infusing our scholarship with the empirical form of the inquiring mode. Paradoxically, many of those dynamics originate on the doctrinal side of law school faculties and not with us, even though, as clinicians, we are the ones who teach persuasion mode skills.

But we could begin to think empirically ourselves. When we hear a generalized assertion about lawyering or teaching, we might ask, Is this proposition testable? What methodology would confirm or disconfirm it? Has anybody already studied it? If no, why not develop the first study? If it has been studied before, why not do a replication study? Empirical imagination could help us learn a lot more about what's really going on out there.

Thanks to Gary and Bea for pointing the way.

