

A Reply

Lee Epstein[†]
Gary King^{††}

Receiving three detailed responses in this exchange is gratifying and confirms our view of the importance of including articles devoted primarily to methodology and research design in law reviews. Our primary goals in writing *The Rules of Inference*¹ were to adapt the rules to research in legal scholarship, to explicate and elucidate them, and to argue that they should be followed in all future empirical studies. We are pleased, despite the fireworks in these reviews (about which we say more below) that all six of our distinguished commentators appear to stand unanimously with us in support of these key claims. Indeed, none of the reviewers question a single rule we offered or any general methodological advice we gave, and all agreed that following the rules we proposed will help researchers in this field to learn about the world. We are also encouraged that the reviewers appear to believe that our suggestions for improving the organization of legal scholarship's support for empirical work at least "warrant debate and discussion."² These were precisely the goals we set out to achieve. We hope the legal community will join together with us and our discussants to continue to foster the development of a new subfield devoted to the methodology of empirical research in the law, to further adapt and extend the rules of inference in this area of scholarship, and to continue to integrate this aspect of legal scholarship into the rest of

[†] Edward Mallinckrodt Distinguished University Professor of Political Science and Professor of Law, Washington University in St. Louis; <<http://artsci.wustl.edu/~polisci/epstein/>>; epstein@artsci.wustl.edu.

^{††} Professor of Government, Harvard University and Senior Science Advisor, Evidence and Information for Policy Cluster, World Health Organization; <<http://GKing.Harvard.edu>>; King@harvard.edu.

For helpful comments, we thank Jim Alt, Brian Glenn, Fred Schauer, Jeff Staton, and Dennis Thompson. We thank the National Science Foundation (SBR-9729884, SBR-9753126, and IIS-9874747) and the National Institutes of Aging (P01 AG17625-01) for research support; Micah Altman, Anne Joseph, Susan Appleton, Stuart Banner, Shari Diamond, Bill Eskridge, Malcolm Feeley, Barry Friedman, Brian Glenn, Jack Goldsmith, Andrew Holbrook, Dan Keating, Bob Keohane, Jack Knight, Bert Kritzer, Ron Levin, Joan Nix, Eric Posner, Dan Schneider, Nancy Staudt, Sid Verba, and Adrian Vermeule for helpful comments and discussions; Dana Ellison, Andrew Holbrook, and Jeff Staton for research assistance; and Scott Norberg for his data.

¹ Lee Epstein and Gary King, *The Rules of Inference*, 69 U Chi L Rev 1 (2002).

² Frank Cross, Michael Heise, and Gregory C. Sisk, *Above the Rules: A Response to Epstein and King*, 69 U Chi L Rev 135, 147 (2002).

the academy for the purpose of advancing our knowledge of the empirical world.

I. PERSUASION VERSUS INFERENCE

Legal scholars are conducting more and more empirical research, but as is evident from the reviews, inference is hardly the only purpose of legal scholarship. We thought this point was sufficiently obvious and well accepted (and in any event it was not the point of our Article) that we did not emphasize it. In retrospect, however, doing so might have made the reviewers feel less defensive. For example, Professors Goldsmith and Vermeule seem to accept our characterization of the purposes of legal scholarship, but tax us for putting it in a “buried footnote.”³ We have no objection to a change of emphasis, and so in this response, we begin with this footnote. In our Article, we raised some of the obvious training differences between law professors and social scientists and then wrote:

These training differences have led to numerous misunderstandings and accusations being lobbed back and forth across the disciplinary divide. What critics miss, however, is that these differences in perspective are consistent with the markedly differing goals of the two sides. Among other things, [social and natural] scientists aim to conduct good empirical research, to learn about the world, and to make inferences. Lawyers and judges, and hence law professors, in contrast, specialize in *persuasion*. Lawyers need to persuade judges and juries to favor their clients, and the rules of persuasion in the adversary system are different from the rules of empirical inquiry. As actors who lack the power of enforcement, judges attempt to enhance the legitimacy of their actions by persuading the parties to lawsuits, the executive branch, the public, and so on, that judicial decisions have a firm basis in the established prior authority of law rather than in the personal discretion of judges—even when that authority is inconsistent, illogical, historically inaccurate, or nonexistent. Law reviews, in turn, are filled in part with shadow court opinions (with many articles written by former law clerks), rearguing, supporting, or practicing this art of political persuasion. For the purposes of political persuasion, “judges can make claims about history, philosophy, economics, and political science that would be regarded as shallow or discreditable by practitioners of those disciplines but that do not offend the minimal standards of acceptabil-

³ Jack Goldsmith and Adrian Vermeule, *Empirical Methodology and Legal Scholarship*, 69 U Chi L Rev 153, 156 (2002).

ity for performance of their own distinctive craft.”⁴ We can see the lack of contradiction only by recognizing that the Ph.D.s’ goal of learning about the empirical world differs from the J.D.s’ goal of political persuasion. Of course, lawyers also understand that ideas about the world can be wrong, and Ph.D.s use persuasion to convince others of the importance of their ideas, but the institutional differences in the fields nevertheless remain stark.⁵

The overlap in the goals of the two sides briefly outlined in this footnote is substantial. Indeed, the purpose of this response is to persuade readers that following the rules of inference will help them learn about the empirical world. The veracity of our point has been demonstrated innumerable times in different disciplines, and so in some sense we think we are “right,” but our goals of persuading readers to listen to our points about these rules and to act on that basis is no less an act of persuasion than a law professor trying to convince readers that his or her ideological position is normatively preferable.

Yet, there is a crucial difference. To see this, compare the extreme case of pure normative philosophy with most of the sciences. Much progress has been made in philosophy at posing interesting questions and working out illuminating ways of thinking about them, but in twenty-five hundred years there has been relatively little discernible progress at achieving consensus on the “right” normative positions (whatever that might mean). The sciences have done no better at deciding the “right” empirical questions to ask, but they have made stunning progress at learning about the world and often even at engineering it to serve human purposes, at least for some questions. This difference is neither a surprise nor a criticism of the venerable field of philosophy. We obviously need both scholarly endeavors, and it would make no sense to ask which is “better.”

However, the massive difference in results does suggest that if some part of a research program can be posed as an empirical question then it seems worthwhile to do so—not to suppress the philosophical question (which would be unproductive), nor to turn a normative question into an empirical one (which is usually impossible)—but to take advantage of the well-developed rules of inference and the methods of the natural and social sciences to make progress where progress is possible. No amount of empirical research will convince everyone that capital punishment is or is not morally justifiable, but empirical research should eventually be able to tell us whether it has a deterrent effect on crime. The empirical fact could be useful for help-

⁴ Richard H. Fallon, Jr., *Non-Legal Theory in Judicial Decisionmaking*, 17 Harv J L & Pub Pol 87, 93 (1994).

⁵ Epstein and King, , 69 U Chi L Rev at 9–10 n 23 (cited in note 1).

ing to make public policy, for learning about the world for its own sake, or for helping to inform the normative debate. One key point to take from all this is that progress in learning about the world in specific, perhaps narrow, areas will be delayed unless we are able to identify empirical questions so that we may follow the standard procedures of science.

Similarly, although almost every predominantly empirical article engages in some persuasion and nearly every article aiming to persuade readers of an ideological position uses some empirical evidence, most scholars recognize the advantages of keeping the two goals at least conceptually distinct (outside of some extreme nihilistic, relativistic, or postmodern schools of thought). We therefore feel comfortable with our claim that the vast majority of articles in law reviews make at least some empirical claims, even if that is not the main purpose of their work. The same key point applies here too: legal scholars will be hindered in making progress in learning about the world unless they are able to identify their empirical questions and distinguish them from their efforts at persuasion. Such a move does not imply that science is value-free, that normative concerns are unimportant, or that persuasive rhetoric can ever be fully separated from empirical research. It only means that some progress can be made if we keep our empirical focus, at least temporarily.

Goldsmith and Vermeule respond to this notion when they write:

Epstein and King's complaint [about current practices in legal scholarship] elides a critical possibility: the contest of "particular versions" of the truth ventilated by legal articles that are tendentious when taken separately may, at the systematic level, produce increasingly accurate approximations of truth, as scholar-advocates criticize the work of opposing camps. That possibility transposes the standard justification for the adversary system from the courtroom to the faculty common room, collapsing one of the loadbearing walls in [Epstein and King's] construct.⁶

Goldsmith and Vermeule are correct in one narrow sense: that "adversary scholarship" *might* work is indeed a "possibility." However, we have no examples of academic disciplines making sustained progress in learning about the world with such an approach. At best, the hypothesis must come with considerably more uncertainty than is reflected in Goldsmith and Vermeule's conclusion.

Thus, if learning about the world remains a goal of empirical research in the law—in addition to the goals of argumentation and persuasion—the legal community has a choice. It may either encourage

⁶ Goldsmith and Vermeule, 69 U Chi L Rev at 156 (cited in note 3).

scholars to assume the veracity of Goldsmith and Vermeule's untested hypothesis, where we mislead each other in pursuit of personal fame and fortune and hope that the collective result will benefit everyone, or we can encourage practices in this one area of the field (that is, when, and perhaps only when, scholars are doing empirical research for the purposes of learning about the world) to follow the norms in the dozens of academic disciplines that have decades of experience making progress in learning about the world. The evidence is strongly in favor of encouraging legal scholars who wish to contribute to society's collective knowledge about the world to work as these other fields do—in competition and cooperation with, at its best, norms of honesty, openness, free exchange of scholarly information, neutral reporting of empirical results, disincentives to mislead and distort the facts for one's personal gain, etc.⁷ Even if the adversary system is optimal for the courtroom, there exists no serious basis for thinking it would work better in academia than the system—indeed the only system—that has worked so well for learning about the world in so many other disciplines.

Goldsmith and Vermeule defend current practices by explaining that “legal scholars often write in the lawyer's style rather than in the empiricist's because they are participants in, not just students of, the legal system's practices.”⁸ When scholars are participants in the legal system's practices, even when they are writing in law reviews for this purpose, we have little reason to object as long as they clearly state these purposes. But we need to recognize that the change in legal scholarship toward doing serious empirical research for learning about the world demands a concomitant change in attitude and norms for at least this new area. Would you frequent a physician who is “tendentious” in his recommendation for medical treatment, who recommends medicines for you to take based on whether he makes a financial profit or improves his academic reputation? Of course not. Well, then why should society, the legal community, or the rest of the academy pay attention to law professors who draw conclusions about the empirical world based on biased, advocacy-style “tendentious” scholarship? The legal profession has gained the new power of empirical research. It should not be controversial, when offering empirical analyses as the basis for drawing conclusions about the world, to suggest that legal scholars follow the physicians: First of all, do no harm.

⁷ Many scholars in these other fields do not live up to these utopian ideals, but at least the ideals are widely accepted and the institutions in these fields are set up to reward scholars for adhering to them. And in any event the only question at issue presently is whether we should encourage people to follow these norms or to give up and say “anything goes.”

⁸ Goldsmith and Vermeule, 69 *U Chi L Rev* at 155 (cited in note 3).

Which brings us to our discussants. In addition to agreeing with the central points of our Article, these scholars are indeed also “playing a different game.”⁹ They are engaging in persuasion, and, to be clear, they are exceptionally good at what they do: on first reading even we found their rhetoric persuasive and thought we had made a list of embarrassing mistakes. But upon rereading, carefully parsing what they wrote, and sometimes checking the sources they gave, we came to realize that the cases they made were persuasive only when we were not paying close attention. But, as Goldsmith and Vermeule explained in the quote we gave above, the norms of disinterested scholarship, impartial analysis, and the absence of intentionally misleading arguments are clearly missing from these reviews. It is not as hard as it should be to find similar displays in the social sciences, but we know of no exchange like this that is also accompanied by rhetoric (such as the quote from Goldsmith and Vermeule above) that openly seeks to justify deception in scholarly discourse.

To take one example of the tactics used in this exchange, we used an article by Professor Revesz to explain what it takes to make an empirical analysis replicable. With the information he provided in his article, we explained, no one else would be able to reproduce his results. The article contained much interesting analysis, but we found that this one critical feature was missing. We were then obviously surprised to read his Response, which indicates to the contrary:

In fact, any researcher seeking to replicate my study would not have any such problem. My published work contains references to both the coding protocol used for my study and to the notes explaining the manner in which cases were coded, and indicates that these documents are both publicly available from the journal that published my article.¹⁰

On reading this passage, we thought we had made a serious error. So we went back to his article and looked for all this information that we had apparently missed. We found nothing: the information is not in his article. It took a good deal more effort on our part to discover that Revesz did not lie in his Response to our piece. The only way to know, however, is to parse his words exceptionally carefully and track every footnote down to its original source: when Revesz refers to “my article,” it turns out he is not referring to the article we cited, but to one published two years after the article in question!¹¹ Apparently, after

⁹ *Id.* at 153.

¹⁰ Richard L. Revesz, *A Defense of Empirical Legal Scholarship*, 69 U Chi L Rev 169, 183 (2002).

¹¹ The original article, which we cited and contains no replication information, is Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 Va L Rev 1717 (1997).

being criticized in a different law review years after he published the article we discussed, he provided additional information in another article with that journal.¹² We applaud him for taking these steps, and we can only marvel at his masterful use of persuasive tactics, but it doesn't change the fact: anyone reading only Revesz's original article (or Revesz's article and everything cited in the article) would not have a clue as to how he conducted his research, on what basis he drew his conclusions, how to follow up on this research, or how to replicate his results. Our purpose in discussing Revesz's article was not to make moral judgments about the character of someone we have never met, and readers should have no interest in whether Revesz has ever followed this particular rule in other instances. We only sought to offer one prominent and clear example of an article that did not follow this particular rule of inference, so readers would be able to understand the rule and be able to follow it in the future. It now should be obvious that we accomplished that task, and Revesz's response was irrelevant.

We need to understand when it is appropriate to use the various tools that lawyers need and when use of those tools holds back progress. It would be an impossibly idealistic position to think that deception will never occur, but it is worth saying that we should all try to take the high road whenever possible. Revesz could easily have said: "(a) I agree with the rule that research ought to be replicable, (b) my article was not replicable when originally published and so the use of that example is appropriate, but (c) the information necessary to replicate my piece is now on the public record." This approach would have enabled him to defend his reputation, if he thought that necessary, without misleading readers. Taking the high road in this way not only would have been better for everyone else, but it probably would also have been better for Revesz's reputation.

What Revesz misses here is that the stakes for the academic community and society at large in learning about the world are a lot higher than anyone's personal reputation. This "game" we are playing is not about any of us—in other words, professors who work in cushy jobs, on beautiful campuses, heavily subsidized by the rest of society. When everything works, we are returning the favor to society and fulfilling our obligation by contributing to a common good. Although personalities make up the contributors and come with every manner of human frailty, personal comments, attacks, and defenses are beside

The article containing some of the information necessary for replication is Richard L. Revesz, *Ideology, Collegiality, and the D.C. Circuit: A Reply to Chief Judge Harry T. Edwards*, 85 Va L Rev 805 (1999).

¹² See Harry T. Edwards, *Collegiality and Decisionmaking on the D.C. Circuit*, 84 Va L Rev 1335, 1339–54 (1998) (criticizing Revesz's article and methodology).

the point. Similarly, what Goldsmith and Vermeule miss is that the stakes are also a lot higher than worrying about group-centered concepts like “disciplinary hegemony.” In fact, neither political science nor any other discipline has an exclusive claim on the methods of science used in learning about the world. The theory of inference underlying the rules adapted for particular applications is explicitly nondisciplinary and applies to all those engaged in making empirical claims.

As another example of how legal scholars confuse the goals of their enterprise, Professors Cross, Heise, and Sisk conclude their criticisms with this remarkable sentence: “Notwithstanding the shortcomings of *The Rules of Inference*, Lee Epstein and Gary King remain outstanding scholars.”¹³ We appreciate the generosity, but if there was ever a non sequitur, this is it. Our Article was about the rules of inference, not Epstein and King, nor about Cross, Heise, and Sisk, nor about any individuals for that matter. Our reviewers would have accomplished more if they had kept their focus on the subject matter rather than the authors or themselves.¹⁴

Traditional academic fields in which empirical research follows the rules of inference and scholarly exchange contain much advocacy among scholars and their ideas. But the fight is for the truth or progress towards achieving consensus about the truth, and not about people. If the legal profession is interested in incorporating advocacy and persuasion into the traditions and norms of scientific discourse, precedents exist for doing this while also cooperating and increasing what we learn about the world. For example, Mellers, Hertwig, and Kahneman have demonstrated a model of “adversarial collaboration” in psychology whereby competing parties to an academic debate agree in advance on the set of empirical analyses necessary for resolving the dispute, with the help of an arbiter, and then the arbiter and both parties collaborate on a joint publication to report the results.¹⁵

II. SOME SPECIFICS

We now turn to a few other specific points raised by our discussants. In each case, we discuss the flawed inference or logic as well as the persuasive tactic employed.

¹³ Cross, Heise, and Sisk, 69 U Chi L Rev at 150 (cited in note 2).

¹⁴ This point is addressed explicitly in our Article. See Epstein and King, 69 U Chi L Rev at 42 (cited in note 4).

¹⁵ Barbara Mellers, Ralph Hertwig, and Daniel Kahneman, *Do Frequency Representations Eliminate Conjunction Effects?: An Exercise in Adversarial Collaboration*, 12 Psych Sci 269, 269–75 (2001).

A. The Research Design for Our Empirical Study

In order to justify the inclusion of a methods piece about the rules of inference in a law review, we felt it necessary to verify that there was indeed a problem to be solved. Reading and evaluating all articles in the nation's 474 law reviews now published each year (according to Lexis) was obviously out of the question. If our goal was to characterize the *average* article, we might have randomly selected articles to read. If our goal was to characterize the *best* articles, we would have defined "best" and chosen articles that fit that criterion. But our goal, instead, was to see if there existed a sufficiently large group of articles with methodological errors that would justify us writing an article solely about methodology and having a law review publish it. We felt we needed some justification since the publication of our piece would be unusual for a law review: although there is certainly much self-conscious attention to methodology in the context of substantive law review articles (and our paper should not be read as suggesting otherwise), we have seen no law review article (including those cited by our reviewers) that is devoted primarily to methodology, at least of the kind that appear in the methods subfields of most other academic disciplines.

Thus, we wished to see whether we could identify a large group of articles with at least some methodological errors that might have been helped by a discussion of the rules of inference. We therefore selected, read, and evaluated the 231 law review articles with the word "empirical" in their title, supplemented by all articles that had some empirical subject matter in law reviews recognized as among the most prestigious.¹⁶ We found errors in every article. Thus, even if no other law review articles had methodological errors, we felt this result—which required no inference—provided enough justification to write and publish our piece. That is, inference is about using facts you have to learn about facts you do not have, but for these three-hundred-plus articles, we had the facts. Since this result is purely descriptive and requires no inference, it is not subject to the complaints of the reviewers in cleverly but incorrectly applying the rules of inference to our empirical study.

Of course, one need not go too far out on a limb to say also that if we found methodological errors in 100 percent of the more than three hundred articles we read, a good number of other articles would probably have methodological errors too (an inference we discussed

¹⁶ It is true, as Cross, Heise, and Sisk explain, 69 U Chi L Rev at 138 (cited in note 2), that we do not report the degree of overlap between these two categories ("do not disclose" is the phrase they used), but if the precise list of articles is important, we have provided sufficient information in our Article that they or anyone else could easily generate it.

in footnote 41, which our reviewers did not cite). We believe this claim to be correct; our reviewers apparently believe it too, and our evidence is consistent with it, but our conclusion about the merits of writing, publishing, and perhaps reading a methods paper in a law review does not depend on it.

The Responses attributed goals we did not have or need to our qualitative survey of empirical legal literature. For example, Revesz argues:

It may well be the case that empirical legal scholarship is in bad shape. But how do we know that social scientists do not engage in the same, or other, pitfalls when they write about law-related subjects? . . . Any well-designed empirical study would have to make an assessment of practices among social scientists. Indeed, many of the methodological ills that they attribute to legal scholarship, such as explaining in detail how cases are coded or archiving data in generally accessible repositories, are not exclusive to legal scholarship.¹⁷

Revesz is absolutely correct: if you want to know whether law professors are better or worse at empirical analysis than social scientists, you'd need to do a comparative study. But his discussion has little to do with our Article. We were clear: “[W]e do not mean to suggest that empirical research appearing in law reviews is always, or even usually, worse than articles in the journals of other scholarly disciplines. Such comparisons are irrelevant.”¹⁸ Indeed, we still see no productive purpose in such invidious comparisons: political scientists, economists, sociologists, and others make plenty of methodological errors. We have no idea or interest in whether one group does better “on average” (whatever that means in this context). We think versions of the suggestions offered in our concluding section, for example, should also be considered for adoption in the traditional academic fields, although the discussions there have been underway for some time. Revesz’s clever use of rhetoric here, by answering a question we did not ask, is irrelevant to the issue at hand and misleads readers.

Revesz also uses the same rhetorical tactic when he writes:

[G]iven the extent to which joint degree holders and otherwise academically trained individuals have moved into law teaching, it would be surprising if the quality of legal research—particularly empirical research—had not improved dramatically over time. . . . Surprisingly, after engaging in what they take to be an exhaustive

¹⁷ Revesz, 69 U Chi L Rev at 184 (cited in note 10).

¹⁸ Epstein and King, 69 U Chi L Rev at 18 (cited in note 1).

review of the literature, Epstein and King conclude that the results are grim.¹⁹

Indeed, it does sound like what we concluded is surprising, and hence (Revesz implies) probably wrong—until you read more carefully. First, our Article never claims to contain an “exhaustive” review of the literature or anything remotely approaching it. But more importantly, and more to the point, our little empirical study was not designed to address the question of whether research in the law is getting better or changing at all over time, and we drew no conclusions about such time trends.

B. Coding Decisions on Articles We Evaluated

Professors Cross, Heise, and Sisk say they would have preferred if we had assigned numerical codes to every article we read, summarizing the errors committed.²⁰ This is a fine goal, and we tried to do this, but we found that any coding scheme we could devise would not yield quantitative data that accurately represented the articles and their mistakes being coded. In our view, this is precisely when qualitative analysis is appropriate—which we support, analyze, and, in this case, conduct in our Article. Unlike Goldsmith and Vermeule’s misleading caricature of us as wanting “to remake all other disciplines in the image of large-number statistical empiricism,”²¹ we believe that the rules of inference apply equally to quantitative and qualitative research. We believe that the rules of inference apply equally to researchers whose main purpose is explicitly empirical and to researchers who use empirical claims in the service of pushing “interpretive and normative program[s]”;²² to studies that are quantitative and studies that are qualitative. So, for example, we do not advocate, in contrast to Goldsmith and Vermeule’s deceptive claim, any simple method of counting statements in order to ascertain the intent of the Framers or legislators. As we emphasized, many concepts are not easily amenable to quantitative measurement, nor would measurement error necessarily be reduced if quantification were undertaken. Rather, we simply ask researchers to report to readers the information on which they based their conclusions, whether the information is quantified or not. We said this many times in our Article, and even in our Abstract. Whether one should use any particular method of collecting qualitative or quantitative data depends on how best to collect, amass, aggregate, understand, and extract information from the world in any particular

¹⁹ Revesz, 69 U Chi L Rev at 170 (cited in note 10).

²⁰ Cross, Heise, and Sisk, 69 U Chi L Rev at 145 (cited in note 2).

²¹ Goldsmith and Vermeule, 69 U Chi L Rev at 160 (cited in note 3).

²² Id at 157.

situation. The rules of inference are not statistical rules and, although we find statistical language convenient to describe them, other languages could be adopted and would result in identical conclusions. Explaining, as we did, that every article we read contained at least one violation of the rules of inference we described was, for stated purposes, a sufficient description of our observations.

Of the over three hundred individual articles we evaluated for our empirical study, only one of our critiques was questioned by the discussants—Revesz’s article. In his Response, Revesz explains that his article pointed out how selection bias affected prior studies and what he did about it.²³ Revesz may be correct about the consequences of selection bias in previous research but, as we now show, his methodological procedures did not solve the problem and likely made it worse.

Previous studies had included cases decided by different combinations of judges in the same analysis. The bias from this procedure comes from the fact that the probability of success, which varies with the judges who make up the court, is likely correlated with the expected outcome of the case—the definition of selection bias.²⁴ Revesz’s idea to eliminate this correlation was to divide the cases up into thirty-nine “court periods,” such that the probability of success did not vary within each, and so it could not co-vary with the outcome of the case.²⁵ But the court periods were neither mutually exclusive nor exhaustively sampled. Both problems can generate additional selection bias.

That the periods were not mutually exclusive means that Revesz constructed the thirty-nine periods so that they overlapped, and also that the judges who decided each case varied within each of the periods. As a result, the strategic considerations and panel composition effects that caused the probability of success to vary (and co-vary with the likely outcome), and for which Revesz was trying to control, were not controlled and were indeed probably exacerbated. In addition, only ten of the thirty-nine periods actually made it into Revesz’s analyses, and they were not selected from the set of thirty-nine in a manner designed to mitigate selection bias.²⁶ The fact that including only ten of thirty-nine periods is in itself actually a selection decision, and so may result in selection bias if not done correctly, was not addressed, and the additional uncertainty that creates was ignored. Discarding so much information also introduced additional inefficiencies into Revesz’s analysis, which of course can be worse than the bias he was seeking to eliminate. Additionally, the overlapping periods mean

²³ Revesz, 69 U Chi L Rev at 173–75 (cited in note 10).

²⁴ Epstein and King, 69 U Chi L Rev at 112–14 (cited in note 4).

²⁵ Revesz, 83 Va L Rev at 1721–27 (cited in note 11).

²⁶ See *id.*

that, even if there was no selection bias, Revesz overestimated his level of certainty, since the ten periods were not independent, but were treated as independent. The article includes other violations of the rules of inference, in part resulting from Revesz's failure to correct for selection bias, and in part due to his exclusive focus on bias without noticing other effects his procedures had on efficiency (which also results from separately analyzing parts of his data set that would not be independent even if he had constructed each of the court periods correctly). In any event, our only claim was that each article we read violated one or more rules of inference; in the case of Professor Revesz's article, our coding decision appears accurate.²⁷

C. Choosing Examples for Methods Articles

What is the best way to choose an example to illustrate a method? Cross, Heise, and Sisk seem to think that random selection would be optimal and they have found the smoking gun in our hands: "Thus, they effectively admit that the examples adduced in their article were not selected in any fashion even remotely approaching random."²⁸ Yes, we admit it; indeed, as teachers, we are proud of it. Randomness can be a useful selection method when the goal is inference about features of a population, but it is useless when the goal is pedagogy, as was ours. How much would be learned about the key issues in constitutional law in a class where the cases discussed were selected at random from all Supreme Court cases?

²⁷ Our coding decision also appears accurate with respect to Revesz's choice of preference measures. He implies that we are not aware that the District of Columbia lacks representation in the Senate and hence senatorial privilege. See Revesz, 69 U Chi L Rev at 181 (cited in note 10). In fact, the Giles approach to preference measurement we discussed would have reduced the measurement error in Revesz's indicator, even in D.C. See Epstein and King, 69 U Chi L Rev at 95–96 (cited in note 1). This measure, as we discuss in the text, is based on NOMINATE scores of the President and home state senators, not on political party. *Id.* at 95. In the case of D.C., the senator's score obviously drops out, leaving only the appointing president's NOMINATE score—still a measure with considerably less error than Revesz's dichotomous party-based approach. And, in response to Cross, Heise, and Sisk, 69 U Chi L Rev at 136–37 n 9 (cited in note 2), we note that NOMINATE scores, developed in 1982, were publicly available and widely used for at least a decade prior to the publication of Revesz's and Cross's articles. See Keith T. Poole, *NOMINATE: A Short Intellectual History*, 9 *Polit Methodologist* 2 (1999). The forthcoming Giles piece we cited was only one of many articles to use these scores, but even if it was the first and was unavailable, the Sisk, Heise, and Morriss position that readers should accept their empirical conclusions because they tried hard, does not follow. Readers have no professional interest in judging the authors' work habits, only in the likelihood that the article's conclusions are correct. At best, Sisk, Heise, and Morriss should have used a measure with less error. If that were not possible, they should have recognized the consequence of the error in their measurement and reported the likely size and direction of the bias in their results. At worst, if none of this seemed possible, they should have decreased the level of certainty they attached to their conclusions.

²⁸ Cross, Heise, and Sisk, 69 U Chi L Rev at 143 (cited in note 2).

As we explained, we chose examples from articles that were as good as we could find, with the possible exception being the single methodological problem we were trying to illustrate. An optimal example for describing the problem of omitted variable bias is one that has no problems of selection bias, inefficiency, measurement error, replication issues, etc.²⁹ These other issues are obviously important, but trying to explain every problem all at once is hardly an effective teaching tool. This is why carefully selecting an example so that it illustrates only the one methodological principle at issue, gives a clear violation of that principle, and makes it easy to suggest a way to fix the problem, is a standard and time-honored approach to methodological exposition.

By presenting single errors in examples in this way, our goal was not to “indict,”³⁰ “attack,”³¹ “brand,”³² “criticize,”³³ or “trash”³⁴ anyone’s article or person. We recognize that such defensive reactions are the normal consequence of scholarly exchanges in this field and others, but we had a different purpose. We took few positions on whether the substantive conclusions about the legal world would be upheld based on our suggested changes. Our goal was to explicate the rules of inference we adapted for the law, and it seems we were successful since in most cases the discussants did seem to understand the rules as we explicated them—at least enough to agree with them.

D. Justifying Flawed Statistical Advice with Authoritative-Sounding Language

We point out in a footnote in our Article that Sisk, Heise, and Morriss make incorrect statistical decisions in their article due to their misunderstanding of the concept of multicollinearity.³⁵ These authors excluded any explanatory variable from their regression analyses if it

²⁹ See Epstein and King, 69 U Chi L Rev at 15 (cited in note 1):

Indeed, the optimal article for our purposes here is not one that violates each and every rule of inference. That would produce a mess, require long qualifications, and be almost useless from an expository perspective. The best example from our perspective is an article that is perfect in nearly every respect save one—the one that illustrates the rule we are explicating. That way we can demonstrate most cleanly and clearly the advantages of following a particular rule of inference.

³⁰ Goldsmith and Vermeule, 69 U Chi L Rev at 153 (cited in note 3); Revesz, 69 U Chi L Rev at 182 (cited in note 10).

³¹ Cross, Heise, and Sisk, 69 U Chi L Rev at 135 (cited in note 2); Revesz, 69 U Chi L Rev at 169 (cited in note 10).

³² Cross, Heise, and Sisk, 69 U Chi L Rev at 137 (cited in note 2).

³³ Id.

³⁴ Id at 147.

³⁵ Epstein and King, 69 U Chi L Rev at 11 n 28 (cited in note 1), discussing Gregory C. Sisk, Michael Heise, and Andrew P. Morriss, *Charting the Influences on the Statistical Mind: An Empirical Study of Statistical Reasoning*, 73 NYU L Rev 1377 (1998).

correlated at 0.5 or higher with another explanatory variable, and we explained that their understanding of multicollinearity was “incorrect” and their statistical analysis strategy was “flawed.”³⁶

In their response, Cross, Heise, and Sisk write that we “assert error and provide no evidence for [our] claim.”³⁷ This is indeed the case. We did not explain the details of the error the authors made since we did not think it was necessary to teach introductory statistics in this forum, especially since it is so easy to find the correct advice in a wide variety of modern statistical textbooks. Classic errors like this one constitute good examples of the kinds of errors that would likely be caught during peer review, although they obviously should not have been introduced in the first place. In any event, we do not object to this criticism: we could have explained their error in our Article and we will do so here.

However, Cross, Heise, and Sisk go further than asking us to explain. Apparently without reflecting on our criticism, or consulting any of the numerous sources that could have set them straight, they stridently claim that they were correct, and (quite recklessly) imply that other empirical researchers should follow their lead. They write, “We remain confident that our caution in applying the most stringent standard to this common problem remains the better part of scholarly discretion.”³⁸ The persuasive strategy of using language that is far more authoritative, strident, and certain than is accurate to claim is common and sounds impressive, but of course it has nothing to do with whether the advice is correct. In this case, Cross, Heise, and Sisk’s advice is not correct, and it cannot turn bad statistical advice into anything approaching “scholarly discretion.”

What is multicollinearity? Multicollinearity is when one explanatory variable in a regression analysis can be *perfectly* predicted from the other explanatory variables—not approximately predicted, not a 0.5 bivariate correlation, but predicted without error. If the prediction is not perfect, no multicollinearity problem exists. For example, if the other assumptions of the regression model apply, high (but not perfect) correlations among explanatory variables cause no statistical problems whatsoever. Indeed, multicollinearity is the one assumption of regression analysis that can be tested for certain by just running the regression. If the regression program completes, no multicollinearity problem exists; if the program complains about multicollinearity (using the equivalent language: multicollinearity, collinearity, incomplete

³⁶ Epstein and King, 69 U Chi L Rev at 11 n 28 (cited in note 1).

³⁷ See Cross, Heise, and Sisk, 69 U Chi L Rev at 136–37 n 9 (cited in note 2).

³⁸ *Id.*

rank of the explanatory variable matrix, matrix singularity, noninvertible Hessian, etc.), a problem exists.

The reason we require the absence of multicollinearity in regression analysis is that the regression analysis algorithm—known as least squares—yields no unique solution when one or more of the explanatory variables can be perfectly predicted from the others. This makes a lot of intuitive sense. Suppose we wished to explain income with explanatory variables that include sex and participation in an employee training program, but in our data all participants in the program are men and all nonparticipants are women. In this situation, the marginal effect of participation, controlling for sex, is obviously not defined—since in these data the question is equivalent to asking for the effect of participation controlling for participation. If you care about the effect of participation on income controlling for sex, you must turn elsewhere, since no information exists in these data.

What is the consequence of high but not perfect collinearity? The answer is that inferences are not biased at all. The standard errors are larger than they would be if there were lower correlations among the explanatory variables, but the standard errors indeed should be larger and so they are also correct. To see this, imagine that we were able to add to our sample one female participant and one male nonparticipant. With these two additional observations, the correlation between participation and sex is no longer perfect, but it is very high. Now when we try to estimate the effect of participation in the training program on income, controlling for sex, we get unbiased estimates. However, the original set of observations includes very little information and so we can see that multicollinearity is essentially a problem of having a small number of (relevant) observations. This connection is true mathematically, not merely as an analogy. Indeed, Arthur Goldberger asks why, if multicollinearity and a small sample size are equivalent problems, some analyses (incorrectly) focus so much on the former and not on the latter.³⁹ His tongue-in-cheek answer is that we have a big impressive word to describe the former but none for the latter, and so he suggests that we think about introducing tests for “micronumerosity.”⁴⁰ Indeed, if one’s data includes no observations, no inferences can be drawn; a small sample size is not usually a desirable situation, but valid inferences can be drawn from such a sample and, without collecting more data, no problems require correction. The same goes for multicollinearity.

What is the consequence, then, of Cross, Heise, and Morriss’s decision to exclude explanatory variables that correlate at 0.5 or higher

³⁹ Arthur Goldberger, *A Course in Econometrics* 248–49 (Harvard 1991).

⁴⁰ *Id.* at 249.

with each other, and Cross, Heise, and Sisk's defense of this practice? The consequence is almost surely quite serious. As we explain in Part VI of our Article, researchers should control for an omitted variable if it affects the outcome variable, is causally prior to the key causal variable, and is correlated with that key causal variable. If Cross, Heise, and Morriss's analysis had followed all the other rules of inference, their explanatory variables would be identified because they met these conditions. Their decision to exclude variables that were highly correlated with each other thus *guarantees* omitted variable bias in their analysis. In fact, the higher the correlation of the explanatory variables with the one they are excluding (in their terms, the more "stringent" the standard they apply), the worse the bias their decision induces. Suffice it to say that this practice is undesirable and should be discouraged.

E. Timely Research Can Be Scientific without Tradeoffs

Goldsmith and Vermeule write that "[i]f empirical research were costless, [Epstein and King's] prescriptions would be sensible. But given constraints on time, information, expertise, and research funds, academics face inevitable tradeoffs between rigor and accuracy, on the one hand, and timeliness, relevance and utility on the other."⁴¹ The implication, they say, is that "[a] universal insistence on Epstein and King's version of methodological rigor might require making all studies less timely, thereby eliminating as well the studies that are both timely and accurate. Epstein and King have not shown that the costs of methodological perfectionism are worth incurring in the public-policy realm."⁴² Indeed, they are correct that we have not shown this. But this is because we did not claim it and because perfectionism in methods at the expense of other goals is both inappropriate and unnecessary.

As we thought we had made clear in our Article, the fact is that doing serious scientific work well does not require perfection. The legal profession values timeliness, but no more than biologists racing to be the first to locate a disease-causing gene, physicists trying to out-distance each other in the search for new elementary particles, political scientists forecasting and explaining election results or the behavior of policymakers, or economists explaining a stock market crash. Maximizing one goal at the disproportionate expense of another is no more appropriate in methods than in any other area of life.

Applying the rules of inference is not always easy in any particular project, and perfection is normally out of the question. So what do

⁴¹ Goldsmith and Vermeule, 69 U Chi L Rev at 154 (cited in note 3).

⁴² Id at 165.

we ask? We ask that the rules be understood, and that they be applied as intended (in other words, when they do not compromise other goals too much), and that uncorrected methodological problems be flagged for readers and an appropriate amount of additional uncertainty be added to one's conclusions. This point is important enough that it is worth repeating: 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.

If you are worried about bias in a causal effect that may result from an omitted control variable, and you have insufficient time or resources to measure this variable, then you only need to make your best judgment as to the range of possible effects including this variable would have had on your results. This range is precisely the additional uncertainty that should be conveyed to readers when expressing your conclusions. Research with fewer resources, more constraints, or more need for timeliness may turn out to be more uncertain, but hiding this uncertainty and misleading readers is normally considered irresponsible.

Far from requiring perfection and more time than is available, the rules of inference and methods of science make the best use of *any* level of available resources, given *any* deadline. Indeed, in the history of humankind, no method of learning about the world has proven itself to be faster, more productive, or more accurate.

III. CONCLUDING REMARKS

Instead of continuing to detail other misleading statements made in the reviews, we conclude with three general points we hope the legal community will take away from this Exchange. First, and most obviously, serious empirical research cannot take place without attention to the rules of inference. We hope those who do or plan to do empirical research learn these rules and learn how to use them. As this one aspect of legal scholarship integrates into the rest of the academy, we hope scholars are able to take what is useful from the experience others have had in this task, to adapt it for their own purposes, and eventually even to contribute new methods and knowledge of the world back to scholars in other fields pursuing other research questions.

Second, methodological imports alone cannot sustain empirical research for long in an academic discipline as diverse and distinctive as the law. Social and natural scientists have their own research to

work on and their own methodological problems to solve.⁴³ Some from other areas will make their contributions to this field (and we hope our Article is taken in that light), but social scientists are mostly busy doing social science. What other academic disciplines have—but the community of legal scholars lacks—is not a set of specific facts or techniques. It is a well-developed methodological subfield devoted to solving new methodological problems as they arise. We see nothing in legal scholarship comparable to political methodology, econometrics, psychometrics, or sociological methodology. These subfields include methods appropriate to the data and inferential problems that arise in their fields (including the methods of statistics, interviewing, ethnographies, modeling, participant observation, experiments, network analysis, archival work, historical studies, and many other diverse approaches). But these scholars do not have all the methods the legal community needs, and most do not have an interest in developing them for law professors. We think it should be obvious that legal scholarship is sufficiently important and distinct that it too needs its own methodology subfield devoted to solving its own problems.

We close with a plea. When pursuing the goal of learning about the world through empirical research, take the high road: contribute, don't deceive; be an advocate for the community of scholars, not for yourself. You'll learn more; you'll do it faster, better, and less expensively. You'll have more effect on public policy and jurisprudence. And you'll teach more to the rest of us in academia and society at large.

⁴³ In another form of rhetorical flourish, namely exaggeration and misrepresentation combined with authoritative-sounding language, Revesz writes that “Epstein’s and King’s solution [to the methodological problems in the legal community] is to place empirical legal scholarship in a type of intellectual receivership, in which law professors get sent to the functional equivalent of a reeducation camp.” Revesz, 69 *U Chi L Rev* at 169 (cited in note 10). It does sound impressive, but it is wrong. As we made clear in our Article, and repeat below: no, we do not think the methodological problems in the legal community should be solved by others, and we do not think it would work in any event. As this paragraph explains, the point of our Article is the reverse.