Suffolk University Law School Faculty Works   Suffolk University Law School

2-16-2010

Distinguishing Causal and Normative Questions in Empirical Studies of Judging

Patrick S. Shin
Suffolk University Law School, pshin@suffolk.edu

Follow this and additional works at: https://dc.suffolk.edu/suls-faculty

Recommended Citation

DISTINGUISHING CAUSAL AND NORMATIVE QUESTIONS IN EMPIRICAL STUDIES OF JUDGING

Patrick S. Shin†

In this Essay, I raise a metatheoretical question concerning the relationship between what seem to be two distinct categories of projects that might be lumped together under the rubric of empirical study of judicial performance. One kind of empirical project aims broadly at developing a social-scientific theory of judging, or as one legal philosopher recently put it, identifying the "causes" of legal decisions. [1] Another kind of project aims at identifying quantitative, measurable criteria to provide an objective basis for evaluating the quality of judicial performance or, to use a more loaded term, "judicial merit." [2] I

attempt to explain the distinction between these two types of projects and consider whether the very possibility of success in the former undermines the point of the latter. Would a theory that could predict how any given judge would likely decide any given case obviate the usefulness of general criteria for measuring judicial quality? I suggest here that the answer is no, because the two projects address fundamentally different types of questions.

One kind of empirical study of judicial decisionmaking might be regarded as continuous with the broader goal of social science, which I take to be something like understanding human behavior in general. I see the ultimate aim of this sort of empirical inquiry as developing a theory that explains and predicts judicial decisionmaking in roughly the same way a psychological theory might seek to explain and predict other observed phenomena of human behavior, such as the tendency to obey authority. Just as a successful psychological theory of obedience might, among other things, identify the conditions that explain why and predict whether a given subject will obey an order given by an authority figure in a particular context (for example, personal characteristics of the subject, the subject's relation to the authority figure, the nature of the order, and its expected consequences), one might likewise consider an empirical theory of judging in this vein successful if it allows particular conditions to be identified—for example, political ideology, characteristics of the litigants, particular features of a case's history, or the provenance of relevant precedent—that explain and predict judicial outcomes. [3]

Within this broadly defined empirical project of explaining and predicting judicial decisionmaking as human social behavior—one might call it the project of "naturalizing jurisprudence" [4]—there are multiple theoretical perspectives that might be relevant. A single type of observed judicial decisionmaking might be understood simultaneously through the frameworks of sociology, political science, social psychology, cognitive psychology, and perhaps even neuropsychology. Whether or how all these empirical perspectives might be integrated remains unclear. Presumably, a grand unified theory of judicial decisionmaking is no more and no less likely than a grand unified theory of human behavior in general. All of these scientific perspectives may be viewed as having one common, general aim: they seek to provide *causal* explanations of judicial decisions— theories that identify the causal predicates of observed decisions and do so with predictive power. [5]

A second type of empirical study of judicial performance seems quite different in its basic aim from the project of naturalizing jurisprudence. The goal of this second type of study might broadly be described as identifying quantified measures of good judicial performance—for example, citation counts, dissent rates, and productivity—that can be used to assess and even rank the quality of sitting judges, judicial candidates, and courts. This type of undertaking can be seen as a way of compiling otherwise inaccessible

information that would presumably be of significant value to public officials, the citizenry, and judges themselves in evaluating and monitoring judicial performance. Public ratings of judges and courts based on this information might have the added desirable effect of sussing out the opaque criteria that various political actors use to champion particular judges or candidates, insofar as those ratings would exert pressure on such actors to give public explanations supporting any low-rated candidates they seek to promote. [6]

What is the relationship between these two empirical projects of naturalizing jurisprudence and of measuring judicial performance? One possibility is that the need for objective measures of judicial performance is a function of the current infancy of the science of identifying the causes of judicial decisions. That is, objective measures that serve as proxies for judicial quality are only necessary because of the lack robust theories that would predict how a particular sort of judge would likely decide a particular sort of case. If such a theory existed, there would be no point in trying to develop any general measure of judicial quality.

An example may help draw out the intuitive appeal of this conjecture. Suppose one could prove that in any case possessing the set of features F, involving a party with characteristics P, a second party with characteristics D, and given additional specifiable conditions C, a judge with a set of characteristics J will always decide the case in a way that is favorable to the party with characteristics P, whereas a judge lacking J will always decide against that party. One might argue that this postulate, if true, would undermine the relevance of any generic notion of judicial quality—apart from whatever constitutes J characteristics—in the context of cases with features F. If one could predict how any particular judge would decide this type of case, there would be no need for further information about the judge's qualities, at least in that limited context. And if analogous predictions could be made with respect to cases with features F1, F2, and so on, and judges with characteristics J1, J2, and so on, there would be no more reason to care about measuring judicial quality in the context of those cases. Thus, the more progress empiricists make in the project of reducing judicial decisionmaking to its natural causes, the less relevance any project of measuring judicial performance will have.

One might argue, in other words, that judicial quality matters because better judges presumably will make better decisions. But if it were possible to predict judicial decisions in the manner postulated, then no one would have reason to care about generic measures of judicial performance. There would be no reason to fret over proxy measures of good judicial decisionmaking if social science could deliver a theory that directly predicts how a particular judge or candidate would decide particular kinds of cases.

I believe that this argument should be rejected. This argument's fallacy involves its

reliance on the implicit assumption that empirical measures of judicial performance are, at their core, nothing more than an indirect attempt to accomplish one of the goals of the project of naturalizing jurisprudence—namely, developing a theory with the power to predict the outcome of judicial decisions on the basis of specifiable causal predicates. But the project of measuring judicial performance need not be assimilated to that of theorizing the causes of judicial decisions. Rather, the project is fundamentally normative and evaluative in character. The basic question is not about the causes of decisions, but about what makes a good judge, or what constitute the basic virtues of a good judge. [7] Whereas the naturalizing project seeks to provide causal explanations for judicial behavior in a manner continuous with social science, the empirical study of judicial performance seeks to make explicit and then reduce to numbers our value judgments about the relative merits of selected characteristics of judicial performance. The ultimate test of a causal theory is its explanatory and predictive power. The test of a measure of judicial performance is ultimately the normative plausibility of its embedded value judgments about the core virtues of a judge and how well it captures those judgments. [8] The two projects have incommensurable aims.

There are, however, some important caveats. I do not deny the potential *relevance* of the findings of naturalized jurisprudence to the project of measuring judicial performance. For example, some observers might think that any evaluation of judicial performance should incorporate criteria that capture something like political independence or capacity for "nonideological" decisionmaking. [9] But what if there were empirical evidence that indisputably established that, as a matter of fact, political affiliation almost always predicts judicial outcomes in certain types of cases? I suspect that if that were the case, there would be reason to doubt whether criteria aimed at measuring political independence could possibly capture anything meaningful. Standards of judicial quality must be tempered by contemporary knowledge regarding the limits of human psychology. This is a consequence of the basic moral premise that "ought" at least in some sense implies "can." To that extent, the study of the causes of judicial behavior is potentially relevant to the project of measuring judicial quality. My point is not to deny this possible point of congruence, but rather to emphasize that society's concept of a good judge is not simply *given* by how most judges in fact tend to behave. Findings in the science of judicial behavior cannot themselves determine the normative standards by which judges should be measured and evaluated. This is a consequence of another basic moral premise, namely, that "is" does not imply "ought."

The other caveat is that my remarks assume that it is possible to construct a model of a good judge that is at least to some degree independent of considered preferences relating to case outcomes. This assumption means, among other things, that a judgment

about whether a particular individual would make a good judge is not simply reducible to a set of predictions about the outcomes of cases that would come before that individual. But what if one were to reject this assumption? What if the concept of a good judge that best reflected societal and legal norms did in fact turn out to be nothing more than a reflection of collective preferences about case outcomes? [10] In that case, I do think that measures of judicial performance that captured the concept might be collapsible into predictions of judicial behavior. That is, if society's notion of a good judge turns out to be nothing more than a set of predictions about the likelihood of a judge's reaching particular outcomes in particular cases, then measures of judicial performance would be nothing more than proxy predictions about what judges would probably do in such cases. It might follow, then, that my claim—that the project of measuring judicial performance is fundamentally distinct from the project of determining the causes of judicial decisions—depends on the defensibility of a particular kind of concept of a good judge, namely, one that is not tied to the desirability of specific outcomes in particular cases.

I do not think that this conclusion does, in fact, follow. Even if this sort of cynicism about the concept of a good judge were warranted, I believe that the project of naturalizing jurisprudence would still remain fundamentally different from the project of quantifying judicial performance, because the latter is essentially a normative endeavor in a way that the former is not. Even if every theorized measure of judicial performance turned out to be nothing more than an elliptical predictor of a particular set of case outcomes, every theorist's proposed measure of judicial performance would still be answerable to questions about its underlying normative conception of a good judge. Empirical theories about the *causes* of judicial decisions need not answer these questions.

Pretend, for example, that one could show that measures of judicial performance that depend on citation counts tend to highly rank judges who are more likely to invalidate legislation in federal constitutional cases. If that were true, then commentors might observe that these performance measures are empirically reducible to predictors of how judges will decide those kinds of constitutional cases. This finding would not, however, undermine my central argument, which is that the project of empirically measuring judicial performance, in contrast to the project of identifying the causes of judicial decisions, is fundamentally normative. The discovery of a sufficiently tight predictive correlation between citation counts and a disposition to invalidate legislation might support a claim that any interest in the former as a measure of judicial quality is *really* nothing more than an indicator of a preference for the latter outcome. But the question that this discovery would not and could not answer is whether anyone *should* therefore stop using the citation count measure as a benchmark for judicial performance. Arriving at an answer would require normative discussion about principles of judicial evaluation—principles that specify why certain

considerations should count as legitimate reasons for or against appointing someone as a judge. Any given measure of performance may collapse into a prediction about substantive case outcomes. This possibility does not show, however, that the measurement project itself collapses into a predictive one. Whatever an empirical theory of judicial performance might in fact be measuring, it must always answer one normative question that a *purely* predictive one need not answer: *should* the measures in question form the basis for evaluating judges? For this reason, the project of measuring judicial performance is inescapably normative in a way that the goal of naturalizing jurisprudence is not.

Studies of judicial performance that seek to determine judicial quality by quantitative measures (such as citation count) will ultimately stand or fall on the strength of the normative reasons that can be marshaled for valuing as judges the kind of individuals who do well on those measures. Empirical measures of judicial performance ultimately depend on normative claims about what it means for someone to be a good judge, and the strength of any proposed empirical measure is necessarily a direct function of the strength of the justification of those normative claims. Theories about the causes of judicial performance do not depend on these justifications. And no purely empirical project can *supply* those justifications. That is a task for normative, not empirical, inquiry.

† Associate Dean for Academic Affairs and Professor of Law, Suffolk University Law School; J.D., Harvard Law School; Ph.D. (Philosophy), Harvard University. This Article originally appeared on the Legal Workshop as part of the Duke Law Journal's 2010 "Symposium on Evaluating Judging, Judges, and Judicial Institutions."

[1] Leslie Green, *Law and the Causes of Judicial Decisions* (Univ. of Oxford, Oxford Legal Research Paper No. 14/2009, 2009), *available at* http://ssrn.com/abstract=1374608.

[2] *See, e.g.*, Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. Cal. L. Rev. 23 (2004).

[3] The numerous political science studies that examine correlations between judicial behavior and political affiliation within the framework of the "attitudinal" model of judging are perhaps the most familiar examples of this sort of theorizing. *See* Jack Knight, *Are Empiricists Asking the Right Questions About Judicial Decisionmaking?*, 58 Duke L.J. 1531, 1534–38 (2009).

[4] *See* Brian Leiter, Naturalizing Jurisprudence: Essays on American Legal Realism

AND NATURALISM IN LEGAL PHILOSOPHY (2007). I realize my usage here may be somewhat broader than Professor Leiter's intended meaning. For a very helpful explanation of Leiter's notion of naturalized jurisprudence, see Robin Bradley Kar, *On the Prospects of a Naturalized Jurisprudence*, NOTRE DAME PHIL. REV., July 31, 2009, http://ndpr.nd.edu/review.cfm?id=16805 (reviewing LEITER, *supra*).

[5] This "naturalizing" project of identifying the causes of judicial decisions seems entangled with certain substantive jurisprudential claims associated with legal realism. Some empirical studies, for example, seem to start with the assumption that in many or even most cases, the law admits more than one possible outcome and so cannot be the cause of the actual outcome of the case. These studies presuppose that some other set of factors exists and causes the outcome; an empirical question remains regarding what those factors might be. This Essay is not the place to take issue with this argument (which I have admittedly caricatured), but I see no reason why an empirical study of judging should have to accept this argument at the outset. It might turn out, to be sure, that the theories with the most explanatory and predictive power tend to deemphasize the law as a determinant of decisions, but then again, it might not. It seems patently question-begging to assume from the outset that judicial decisions must be explained by extrinsic, nonlegal considerations.

[6] *See* Stephen J. Choi, Mitu Gulati & Eric A. Posner, *Judicial Evaluations and Information Forcing: Ranking State High Courts and Their Judges*, 58 DUKE L.J. 1313 (2009).

[7] *See* Lawrence B. Solum, *Virtue Jurisprudence: A Virtue-Centered Theory of Judging*, 34 METAPHILOSOPHY 178 (2003), available at http://ssrn.com/abstract=369940.

[8] For example, a study of judicial performance that ranks judges by citation count must assume—insofar as it is to be regarded as a study of judicial quality—that the ability to produce opinions that are cited by others is something we reasonably want in a judge. Whether we take the study seriously will depend on whether this assumption can be defended. This defense would require some argument that the ability to produce oft-cited writings captures some trait that is important to our normative ideal of a good judge.

[9] *See, e.g.*, Choi et al., *supra* note 6, at 1323 ("Independence is a hallmark of judicial quality.").

[10] I do not argue for the point here, but I do not believe the concept of a good judge is straightforwardly dependent on preferences regarding case outcomes. I believe, for instance, that two people can agree about the core virtues of judging even if they have different outcome preferences.