

Duke Law Journal Online

VOLUME 67

MAY

2018

TALKING JUDGES

JACK KNIGHT†
MITU GULATI††

The Evaluating Judges conference held in October 2009 was the second in a series of conferences planned at Duke Law School in which judges and academics came together to discuss questions relating to research on judges.¹ As two academics who study judges, we find these opportunities to interact with judges valuable. On occasion, the interactions can be tense—such as when the judges tell us why they think our work is fundamentally flawed, and we try to respond by pointing to data that seem to contradict their arguments, and they assert that the data are biased and so on. This recent conference was not lacking in tension, but it was different in that it did not follow the typical format of involving judges or other practitioners in academic work. Under that format, academics present research papers, and then judges or other practitioners provide commentary. At this conference, by contrast, the judges took the lead role, identifying and discussing the topics they thought were most relevant to research on evaluating judicial behavior. To be sure, part of the conference still involved the judges telling the academics, and particularly the empiricists, that they were headed in the wrong directions and that judging was an “art” and

Copyright © 2018 Jack Knight & Mitu Gulati.

† Frederic Cleaveland Professor of Law and Political Science, Duke University.

†† Professor of Law, Duke University School of Law.

Thanks to our colleague and co-teacher, Dean David Levi (formerly the Chief Judge on the Eastern District of California), for numerous conversations about these issues. He disagrees with us on almost every issue possible, which makes our conversations fun. 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. Duke Law, Evaluating Judging, Judges, and Judicial Institutions, <http://www.law.duke.edu/news/story?id=3871&u=11> (last visited Mar. 9, 2010).

not amenable to measurement.² But the judges also talked about what they thought was important and what kind of information they might consider important in evaluating themselves or their fellow judges. It is the latter set of conversations that we focus on here.

The observations we report on are from two days of conversations between judges and academics, both at the formal sessions and informally during breaks and at meals. We combine those with similar observations from the one-day conference on Measuring Judges held a year ago at Duke. (The earlier conference followed the more typical format, with empiricists presenting research and judges commenting.) What we report here are no more than subjective impressions. For reasons of confidentiality, we provide no identifying information on the sources of our observations. Instead, we paint some general themes with a broad brush. Our reason for reporting on these themes is that there are ideas here about studying judges that would not have occurred to us but for the opportunity to talk to judges themselves. Hopefully, other researchers will find these ideas as interesting as we did.

I. WHAT JUDGES DISLIKE

It was clear that the judges dislike being measured and ranked by academics who do not understand anything about what qualities make for a truly great judge. The first couple of sessions of the conference, therefore, featured a number of the judges talking about how the measures used by academics to evaluate them were bogus. For example, many of the judges found using citation counts to measure the quality of judicial decisions particularly problematic. According to some, citation counts are flawed because they measure judicial expansiveness, rather than careful and narrow fact-driven analysis. Also annoying, according to some of the judges, are attempts to measure independence using the degree of disagreement among judges. Judicial independence is often measured by how often judges dissent, particularly against those from the same party. This is not judicial independence, as these judges saw it: rather, dissenting is more akin to cantankerousness—at best, it is indulgent; at worst, it undermines collegiality.³

2. To which we might ask, “Aren’t there market prices for art?”

3. The contrast between the uses that researchers make of dissents and the distaste that many of the judges appear to have for it is the focus of Joanna Shepherd’s paper for this Workshop. See Joanna Shepherd, *Diversity, Tenure, and Dissent*, LEGAL WORKSHOP (DUKE L.J.,

Fair enough. The point, at first cut, appeared to be that academics like us were drawing the wrong inferences from the data. Instead of concluding that the judges with the most citations were the most influential (and thus the “best”), we should have been saying that they were likely the most expansive (and thus the “worst”). And instead of using the term “independence” for the measure of dissents, perhaps we should have been using the terms “disagreeable” or “uncollegial.”

Indeed, these are questions that can probably be tested against other measures. One could look at the opinions cited more frequently and examine whether they are more expansive than the ones cited less (assuming one could come up with a measure for expansiveness⁴). If it turned out that this were the case, we asked the judges, would they be willing to look at citation rates as a measure of *bad* judicial performance? Their response was that we were still missing the point. The point was not that citation rates showed low-quality judging rather than high-quality judging, but that they showed nothing. It would be as if we had taken data on sunspots and used that to evaluate judges. The bottom line for some of the judges (not all) seemed to be that no data is better than the type of data that many academics were using.⁵ Further, they seemed deeply skeptical of the ability academics to ever come up with objective measures of judicial performance, given the complexity of the job.⁶ From their perspective, such reductionist research is demeaning to the judges and undermines the system as a whole.

Lastly, there was concern expressed that some judges might focus unduly on trying to do well in the academics’ rankings. That is, they might focus their efforts on publishing opinions, dissenting against co-partisans, and obtaining citations instead of on the important aspects of the job—which, at the district level, is mostly about case management. We have to confess that the idea that judges would pay even the slightest bit of attention to a ranking by some academics, let alone that it might affect their behavior, was more than a bit surprising.

Feb. 25, 2010).

4. We probably should have asked the judges to define more precisely what they meant by expansiveness; our guess is that it is the converse of deciding a case narrowly. And narrow decisionmaking, where the judge addresses as few issues as possible, appears to be viewed by many as a virtue. The theoretical basis for the assertion that narrow decisionmaking is necessarily optimal is not, however, clear to us.

5. The words “garbage in, garbage out” came up.

6. This complexity point befuddles us. Are academics really supposed to avoid measuring certain phenomena because they are complex?

After all, a number of the judges also seemed to be of the view that academics these days produce little scholarship of value to the judiciary.

That said, assuming some judges did care about these rankings, it is not clear that this is a bad thing. To do well on the rankings, judges would have to act in a less partisan fashion (by being willing to disagree with co-partisans), write more publishable-quality opinions, and write opinions that others would wish to use to construct their own arguments (to garner more citations). If judges did modify their behavior at the margins to do better in these areas, would that be a bad thing? The point that these may not be characteristics that are valuable at the trial level is valid. But the response to that is simply that the ranking of district judges should focus more on case-management techniques.

There is a bigger point here, though, which is that if judges really do pay attention to academic attempts to rank them, these rankings can be a means to incentivize judges. And incentivizing judges, particularly those with life appointments, has always been a difficult problem. Better rankings will presumably produce better incentives. If so, the goal should be to produce better rankings. At the end of the day, though, we are not persuaded that judges pay much attention to academic rankings. (There are numbers to which they do pay attention: reversal rates. But more on that later.)

Despite their hostility toward academic attempts to measure judge and court performance, the judges were willing to talk about what constituted good judging. Although this was by no means a uniform sentiment, judges appear to value politeness. What is important is showing the appropriate amounts of respect for lawyers, fellow judges, and judges at lower levels of the hierarchy. Assuming that politeness is important, one question is how to measure it. During a coffee break, one judge suggested that empiricists might look at the frequency of the use of certain words—he suggested “frivolous,” among others—that indicated disrespect of the lower courts or the lawyers involved in the case.

The bigger question, though, was whether having strong norms of politeness actually benefited the system of justice.⁷ One answer was that strong norms of politeness promote deliberation and discussion. The term that kept coming up in connection with these ideas was

7. Dissents, we suspect, are not considered particularly polite, especially when a judge dissents a lot.

“collegiality.”⁸ But do we know whether politeness, particularly of the superficial variety, promotes deliberation and discussion? If judges have to constrain their comments to satisfy the norms of politeness, does that not hurt deliberation? Given the evidence on how easily homogeneous groups can descend into groupthink, is it not important to ensure that there are disruptive elements?⁹ We do not have answers to these questions, but they struck us as ones worth pursuing.

As the discussion of the value of deliberation progressed, an issue that we found interesting was the degree to which deliberation among judges, particularly at the appellate level, takes place via highly formalized routines. In some courts, for example, the norm is that judges who are collaborating on an opinion for a case do not simply call each other up and chat about issues with which they might be struggling. Instead, communication is often highly structured and limited to written correspondence. The primary writer of the draft will produce a document and then circulate a copy. The other judges on the panel are not allowed, by norm, to edit that document—not even with track-changes or comment functions. Instead, they respond via memoranda that typically follow a fixed format.

In other courts, there are formal rules about who gets to speak and when they get to speak during judicial conferences (the meetings at which judges discuss the case after hearing it). Often, the rule is that the judges speak in order of seniority, and when they speak they also announce their vote. Furthermore, there may be implicit rules about what kinds of things judges are allowed to say and whether it is acceptable to have debate. For example, is it okay to ask the other judge what his reasons for coming to his decision are? Is it okay to challenge those decisions at the conference? (As best we can tell, the answer is “no” to both of these questions.)

The foregoing raised the question of whether the formal rules that structure judicial conversation and communication (assuming that they

8. Collegiality, what value it brings on the judiciary, and what it means, are also the subject of the recent exchange between Judge Harry Edwards and Judge Richard Posner. See Harry T. Edwards & Michael A. Livermore, *Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decisionmaking*, 58 DUKE L.J. 1895, 1949–52 (2009) (highlighting the benefits of judicial deliberation and collegiality); Richard A. Posner, *Some Realism About Judges: A Reply to Edwards and Livermore*, 59 DUKE L.J. 1177 (2010) (expressing skepticism regarding Judge Edwards’ claims regarding deliberation and collegiality, given the realities of how judges communicate).

9. See, e.g., CASS R. SUNSTEIN, WHY SOCIETIES NEED DISSENT 168–90, 209–13 (2003) (arguing that organizations, including courts, are likely to perform better if they promote dissent).

are widespread) help or harm deliberation on collegial courts. A couple of the female participants pointed out that it was possible that this formality would stifle active deliberation and unduly constrain outsiders to the system (such as women and racial minorities). There is, after all, a growing literature about the need to diversify the judiciary so as to get a broader range of perspectives. But, given that there are relatively small numbers of racial minorities and women on the courts (more of the latter than the former), the only way to have their perspectives heard, we suspect, is to have more deliberation. In other words, at least with respect to diversity, we should want more disagreement and less collegiality.¹⁰

One of the most important features of the justification for deliberation as a way of enhancing collective decisionmaking is the purported benefit of having a wide range of opinions in the decisionmaking process. According to this argument, we need an institutional setting that maximizes the opportunities for these diverse perspectives to participate in the process to realize these beneficial effects. Our discussion raised a basic question: will the purported beneficial effects of diverse perspectives emerge within these conversations' highly constrained structures?

Now, it was not clear to us that the judges themselves were attached to or particularly fond of these constraints on conversation. One cynical judge with whom we discussed this question at a subsequent conference asked whether it really made much of a difference, for example, to say "I respectfully dissent" instead of "I dissent" when the substance of the dissent was to say "Your opinion is completely and utterly wrong." Plus, if everyone understands that "I respectfully dissent" does not signal any respect, who cares?

Finally, neither judges nor academics appeared to have any sense of how much variation there was across courts in these rules of communication, let alone whether they had a meaningful impact on the quality of deliberation (and ultimately, the quality of dispute

10. Along these lines, scholars have examined whether increased levels of gender diversity correlate with higher levels of dissent and slower decisionmaking. Even assuming that that is the case, neither finding should be necessarily dismaying, because one of the values of greater diversity is that there will be a greater range of perspectives and more discussion will be required to reach decisions. *See* Shepherd, *supra* note 3 (finding that higher levels of diversity among state supreme courts is associated with higher levels of dissent); John Szmer, Robert K. Christensen & Elizabeth Wemlinger, Diversity, Conflict and Judicial Efficiency in the U.S. Court of Appeals (Feb. 27, 2009) (unpublished manuscript), *available at* <http://www.ncpsa.net/papers/Szmer.John.pdf> (finding that although diversity among circuit courts leads to losses in judicial efficiency, those losses can be mitigated if circuit courts reach a certain level of diversity).

resolution).

II. REVERSAL RATES

The most interesting parts of the workshop for us were the final two sessions, during which the judges raised the issue of using reversal rates to evaluate judges. The one statistic that a number of the judges appeared to be aware of was their number of reversals. (Further, at least some of the judges had detailed explanations for why the reversals had been unfair or unwarranted.) The academic empirical literature, by contrast, has not done a great deal with reversal rates.¹¹ Part of the reason for this, we suspect, is the dominance of the political science perspective in this area of research. From that perspective, reversals are likely to be a function of political differences between the appeals courts and the lower courts. So, one would expect to see Republican appeals court judges reversing Democrat trial judges more often than they would reverse Republican trial judges. That is, unless Democrat trial judge recognize the preferences of the Republicans on the appellate court and adjust their behavior to please the judges on the court above. Either way, the point is that scholars who think of judges as strategic actors would be unlikely to use reversal rates as a measure of judicial merit.

The fact that judges care about reversals, however, suggests that academics might wish to pay more attention to them. If judges dislike reversals, that suggests that they will take actions to avoid reversals. Further, to the extent some judges dislike reversals more than others, those different levels of reversal aversion should translate into different types of decisionmaking. A couple of the judges suggested that this reversal aversion might manifest itself in decisions at the motion-to-dismiss stage. Grants of motions to dismiss are subject to appellate review, the judges explained, but denials are not. That creates an incentive for reversal-averse judges to be, at the margins, more circumspect about granting dismissals.

Another question that the discussion of reversals raised was what kinds of measures of reversals would be meaningful in assessing the

11. Exceptions include Frank Cross & Stefanie Lindquist, *Judging the Judges*, 58 DUKE L.J. 1383, 1404–05 (2009), which notes that reversal rates are an important metric that may help measure judicial performance, and Corey Rayburn Yung, *Flexing Judicial Muscle: An Empirical Study of Judicial Activism in the Federal Courts* 36–38 (Dec. 6, 2009) (unpublished manuscript), available at <http://ssrn.com/abstract=1434742>, which analyzes reversal rights among circuits and circuit court judges to determine levels of judicial activism.

quality of judicial decisionmaking. Should reversals by same-party appeals courts (Republican appeals panel reversing a Republican trial judge) be a clearer sign of error than reversals by opposite-party panels (Democrat panel reversing Republican trial judges)? The judges did not find our attempt to bring politics back into the discussion of reversal rates appealing, but we suspect that there is something there. One participant also raised the question of whether judges who have more citations (the ones who write expansively) are more likely to reverse lower court judges than those who have fewer citations. We do not know the answer to this question, but it makes sense that reversal rates would be correlated with high citations. If one believes that high citations are a sign of quality—a view with which the judges at the two conferences would disagree—then one would expect more reversals by appeals court judges with more citations. After all, they would be more likely to find errors in the lower court’s analyses. Conversely, oft-cited lower court judges should get reversed less.

We should note here that the judges were *not* saying that academics should be ranking them based on their reversal rates. They would likely be horrified by that suggestion. Rather, our point is that judges pay attention to reversal rates.¹² And even here, we should note that when we brought up the issue of reversal aversion in front of judges on occasions subsequent to the conference (when one of us has been presenting research based on the idea of reversal aversion), we have been told in no uncertain terms by our judge-commentators that judges care not a whit about reversals. (The caveat in that context was with respect to reversals done via summary order; apparently, that is the judicial equivalent of a slap in the face.)

The discussion of reversal rates also led to the question of communication between trial and appeals court judges. Most of the academics at the conference knew little about how much communication occurred between the appeals court judges and the judges they were reversing. For example, was it customary for the appeals judges to call up the trial judges and explain the reason for reversal? Did the appeals judges periodically hold seminars to explain to the trial judges what kinds of errors the trial judges were making systematically? Once the topic of reversals came up, there was unending stream of questions and not enough time for answers.

Suffice it to say, however, that we have embarked on collecting

12. One judge knew not only his reversal rate, but exactly how many times his cases had been appealed over the past few years. He did also say that he didn’t care about being reversed.

reversal rates.

III. SHOWING UP

We finish with our observations on the contrasting behaviors of academics and judges in that most mundane of matters: showing up. Getting the judges to agree to attend was not easy. But once they had agreed to participate, they committed to the enterprise from beginning to end: the judges, at both this and the prior conference, sat patiently through almost all of the sessions. Many of our academic guests, by contrast, viewed attendance as optional. Once they had presented their papers or spoken their piece, the academics (probably including the two of us) were more likely to tune out, go for a walk, check email, text messages, and so on. As best we could tell, the judges also were more likely than the academics to have looked at the background materials for the conferences and done the reading.

Relatedly, the judges seemed cognizant of the hierarchy within their group. And this hierarchy was a function of what level of court they sat on, rather than how knowledgeable or capable they were. Although the ground rules of the conference provided that participants were all at an equal level and were to refer to each other by first name, it would not have taken an external observer long to determine which judges were at which levels in the hierarchy. This is not to say that those at lower levels necessarily deferred to the views of those at higher levels, but there was a pecking order. Needless to say, academics have their own pecking order—few groups are more conscious of social status than law professors. If asked, the judges could probably have discerned a hierarchy among the academics as well. Our point simply is that the judges, at both our conferences, were very aware of relative positions within the *formal* hierarchy. That is, for example in the federal system, magistrate judges are below district court judges, who are below appeals court judges, and so on.

Finally, judges are polite and proper in their interactions (at least, relative to the academics). The judges almost never interrupted the other participants. They raised their hands when they wished to speak and waited to be called on. The problem, though, was that the initial roundtable format of the conversation allowed participants to interject whenever they felt like talking, with the result that the academics tended to talk much more. After a couple of the initial sessions, during which the judges could barely get a word in, the format had to be changed so that a moderator would keep a queue of people who

wanted to speak and make sure that a handful of academics didn't grab all the air time.

These observations will strike some as trivial, but we find the mundane and trivial interesting. What we saw, albeit from the behavior of two small groups, made us wonder whether there was a story here worthy of further investigation. Are the people who choose to become judges also the types of people who respect hierarchy and like following rules, or do judges in the U.S. become socialized into certain patterns of rule-following behavior? Conversely, are people who choose to become academics perhaps those who dislike the constraints of formal rules? To the extent we find that rule-following is more prevalent among judges, that raises additional questions: do outsiders (women and racial minorities) who join the bench tend to violate the rules more or follow them more? More broadly, if the people who become judges are more inclined to be rule followers than those in the general population, does that say something about how laws are likely to evolve in conservative directions?

CONCLUSION

Our observations from these conferences have implications for three different areas of the study of law: explanations of judicial decisionmaking, assessments of the quality of judicial decisionmaking, and analyses of legal and democratic deliberation.

In regard to issues of social scientific explanations of judicial decisionmaking, the discussions with judges suggest that there are two types of factors that may be worthy of greater attention than they presently receive in the literature. The first encompasses psychological factors such as temperament and concern with status. The aversion to reversal on appeal, for example, could significantly restrain the decisions of judges. If this aversion is as widespread as the discussions at these conferences suggested, then we would want to figure out how to incorporate it in a generalizable way into the explanations. The second encompasses group dynamic factors such as the concern with collegiality and the formal structure of deliberation. To the extent that such factors influence the common processes of interaction among judges, our explanations of collegial courts should take better note of them.

In regard to assessments of judicial quality, these discussions challenge us to think harder about what constitutes a good measure of quality. The judges in our conferences think that academics who study

quality are misguided in their basic conceptualization of good judging. And they offer in response some insights into the image of the type of judge with whom they like to interact and the type of judge they would most like to be. In essence, they are offering us a criterion of quality from an internal perspective. Because we can reasonably assume that this self-image motivates, at least in part, their behavior on the bench, it would be a mistake for us to ignore their insights in our consideration of measurement questions. But the internal perspective alone is not enough for an adequate account of judicial quality. Academics who study quality bring a necessary external perspective to such an analysis. When effective, they can develop measures that reflect a criterion of quality based on the effect of judicial decisions on society writ large. Future assessments of judicial quality need to find a way to incorporate both perspectives.

Finally, in regard to analyses of group deliberation, identifying the different rules of deliberation and communication among judges and analyzing those rules' effects on decision quality are matters that both academics and judges should find interesting and relevant. One of the things that struck us about the discussions at the conference was the extent to which academics who study both legal and democratic deliberation may have overestimated the amount of deliberation that actually takes place in the judicial process. Juries and courts are two of the most common examples of group deliberation offered in this literature. But the idealized characterization of deliberation found in the studies is far different from the perspective that emerged from these conferences. Perhaps advocates of deliberation should take a new look at judicial deliberation to see what they might learn about what can and what cannot foster deliberation in group settings. And, at the same time, judges who are persuaded of the benefits of serious deliberation on collegial courts might take a moment to consider if their everyday practices actually foster or hinder such deliberation.