

DO JUDGES BEHAVE AS *HOMO ECONOMICUS*, AND  
IF SO, CAN WE MEASURE THEIR PERFORMANCE?  
AN ANTIPODEAN PERSPECTIVE ON A  
TOURNAMENT OF JUDGES

RUSSELL SMYTH\*

I.	INTRODUCTION.....	1299
II.	WHAT IS <i>HOMO ECONOMICUS</i> ?.....	1301
III.	WHAT MOTIVATES JUDGES TO PERFORM? .....	1302
	A. <i>Financial Incentives</i> .....	1302
	B. <i>Promotion Prospects</i> .....	1304
	C. <i>Respect of Colleagues</i> .....	1306
	D. <i>Influence</i> .....	1307
IV.	HOW SHOULD WE MEASURE AND RANK JUDICIAL PERFORMANCE? .....	1309
	A. <i>Opinions Published</i> .....	1309
	B. <i>Citations</i> .....	1313
	C. <i>Reversals</i> .....	1315
	D. <i>Judicial Gowns or Academic Robes?</i> .....	1316
V.	IS IT APPROPRIATE TO MEASURE AND RANK JUDICIAL PERFORMANCE?.....	1317
VI.	INCENTIVE EFFECTS OF MEASURING AND RANKING JUDICIAL PERFORMANCE .	1325
VII.	CONCLUSION .....	1329

I. INTRODUCTION

The purpose of this Essay is to examine Choi and Gulati’s concept of designing a “tournament of judges,”<sup>1</sup> with particular focus on whether it would be useful in the Antipodes (Australia and New Zealand). Focusing on Australia and New Zealand offers an interesting comparison with the United States and allows one to flesh out some of the issues associated with designing a tournament in countries with seriatim opinion writing and no judicial conference. There are some similarities in terms of promotion arrangements (for example, in New Zealand most members of the Court of Appeal are promoted from the High Court<sup>2</sup>) and some differences (for example, in Australia promotion to the High Court comes from one of several courts or from private practice<sup>3</sup>). There are also differences in the appointment

---

\* Professor, Department of Economics, Monash University. BEc (Hons); LLB (Hons); MEc, Monash University; Ph.D, London University. I thank Mitu Gulati, Ingrid Nielsen, and Richard Posner for helpful comments on an earlier version of this Essay. The discussion of the psychological dimensions of measuring judicial performance was developed with substantial assistance from Ingrid Nielsen.

1. Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CAL. L. REV. 299 (2004).

2. Note that in New Zealand, the New Zealand High Court is for all intents and purposes the second highest court, akin to the courts of appeals in the United States, while the New Zealand Court of Appeal is the final court of appeal, akin to the United States Supreme Court.

3. In Australia, the High Court of Australia is the final court of appeal. Below the High Court of Australia is the Federal Court of Australia, which is broadly similar to the United States courts of appeals and the state supreme courts. If a judge is to be promoted

process. One of Choi and Gulati's main rationales for formulating a tournament of judges is that the process of appointment to the United States Supreme Court is so politicized. This is less so in Australia and New Zealand. The judges of the High Court of Australia and the New Zealand Court of Appeal are appointed by the Governor General, who is the representative of the Queen of the United Kingdom, on the advice of the relevant national government.<sup>4</sup> Thus it is effectively the government that makes the decision on who to appoint, but without the series of Senate hearings into the background of the candidate, which characterizes the appointment process in the United States.

Having said this, the appointment process in Australia and New Zealand still lacks transparency. Governments in both countries are under no obligation to consult with state governments (relevant in the Australian case), the judiciary, or the profession more generally, and with the exception of a few spasmodic anecdotal accounts,<sup>5</sup> exactly how appointments are made is largely shrouded in mystery. The basic notion that some of that nontransparency could be cleared up if the government would provide hard data about judicial performance to back its claim that a particular judge is the best candidate for a position still holds.

There is ongoing debate in Australia and New Zealand on the issue of how judges should be selected, with some arguing for public scrutiny of the opinions and attitudes of judges similar to what occurs in the United States.<sup>6</sup> There is also continuing controversy over whether the High Court of Australia should be more representative of the population (for example, in terms of gender, there has only ever been one female judge on the High Court of Australia) and the implications this would have for judicial performance.<sup>7</sup> Former Chief Justice of the High Court of Australia, Sir Anthony Mason, has stated that while a more representative judiciary "may assist in

---

to the High Court from a lower court, it will typically be from the Federal Court or one of the state supreme courts.

4. For a discussion of the judicial selection process in New Zealand, see Sir Geoffrey Palmer, *Judicial Selection and Accountability: Can the New Zealand System Survive?*, in *COURTS AND POLICY: CHECKING THE BALANCE* 11 (B.D. Gray & R.B. McClintock eds., 1995). For a discussion of the judicial selection process in Australia, see George Winterton, *Appointment of Federal Judges in Australia*, 16 MELB. U. L. REV. 185 *passim* (1987).

5. See, e.g., Palmer, *supra* note 4, at 40-52; Winterton, *supra* note 4, at 185-88.

6. Catriona MacLennan, *Cast the Net Wider for Effective Judges*, N.Z. HERALD, May 10, 2004, at 13; David Solomon, *Judge Backs Scrutiny of High Court*, COURIER-MAIL (Brisbane, Austl.), Mar. 21, 1997, at 14.

7. A similar debate on "representativeness," or "diversity," in the judiciary is occurring in the United Kingdom, particularly since the passage of the Human Rights Act, 1998, c. 42. The Act made the European Convention on Human Rights a part of United Kingdom law, giving the English what is, in effect, a bill of rights. For a discussion of these issues in the United Kingdom context, see Sir Sydney Kentridge, *The Highest Court: Selecting the Judges*, 62 CAMBRIDGE L.J. 55, 56-65 (2003).

maintaining public confidence in the administration of justice, . . . it is essential that [it] be achieved without any diminution in the quality of judicial performance. The insistent demand for enhanced judicial performance requires the appointment of those who are best qualified.”<sup>8</sup> If Australia was to go down that track, the notion of a tournament of judges would become more pressing. In Australia, there is also continuous discussion of the merits (or lack thereof) of introducing performance standards raised by law societies,<sup>9</sup> to which quantitative indicators of the sort suggested by Choi and Gulati would be a useful contribution. One difference, however, between what Australia would use the tournament for and the manner in which Choi and Gulati set up their tournament is that the debate in Australia tends to center on prescribing *minimum* standards to get the below-average judges up to scratch rather than setting indicators designed to identify the highfliers.

## II. WHAT IS *HOMO ECONOMICUS*?

The issue whether judges behave as *homo economicus* (economic man) is a useful starting point, because the premise underpinning Choi and Gulati’s tournament of judges is that judges respond to incentives.<sup>10</sup> The view was long held that judges were immune to self-interest.<sup>11</sup> From the late 1990s, however, sparked mainly by Posner’s seminal work on judicial decisionmaking,<sup>12</sup> a few articles started to appear that argued that judges, like the rest of us, are self-interested and respond to incentives.<sup>13</sup> While the notion that judges are self-interested welfare maximizers has not been unanimously accepted,<sup>14</sup> it has taken root in law and economics’ treatment of judicial behav-

8. David Hodgkinson, *Why Judges Should Not Be Seen as Representatives*, AUSTRALIAN FIN. REV., Oct. 18, 1995, at 16 (internal quotation marks omitted).

9. Janet Fife-Yeomans, *Tougher Guidelines Needed for Judges: QC*, AUSTRALIAN, Nov. 1, 1995, at 6.

10. Choi & Gulati, *supra* note 1, at 305.

11. Richard A. Epstein, *The Independence of Judges: The Uses and Limitations of Public Choice Theory*, 1990 BYU L. REV. 827, 827-32.

12. Richard A. Posner, *What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does)*, 3 SUP. CT. ECON. REV. 1 (1993).

13. See, e.g., Ronald A. Cass, *Judging: Norms and Incentives of Retrospective Decision-Making*, 75 B.U. L. REV. 941, 946-48 (1995); Michael Heise, *The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism*, 2002 U. ILL. L. REV. 819, 839-43; Jonathan R. Macey, *Judicial Preferences, Public Choice, and the Rules of Procedure*, 23 J. LEGAL STUD. 627, 627-32 (1994); Frederick Schauer, *Incentives, Reputation, and the Inglorious Determinants of Judicial Behavior*, 68 U. CIN. L. REV. 615, 619-21 (2000).

14. For criticisms, see Harry T. Edwards, *The Effects of Collegiality on Judicial Decision Making*, 151 U. PA. L. REV. 1639, 1653 n.44 (2003) (“[I do not] give much credence to the theory that judges are self-interested pursuers of prestige, reputation, careerist ambition, or influence.”); and Lynn A. Stout, *Judges as Altruistic Hierarchs*, 43 WM. & MARY L. REV. 1605, 1609 (2002) (“[T]he modern judiciary rests on the expectation that judges will behave in an altruistic fashion.”).

ior. The notion of self-interest has always been at the center of neo-classical economics. Bowles and Gintis note that generations of students have been told that “[t]he strength of the neoclassical paradigm . . . lies in its hardheaded grounding in a general model of self-interested action.”<sup>15</sup> The role of self-interest in the economics literature, though, has evolved over time. Bowles and Gintis further suggest that “self-interested behavior underlying neoclassical theory is artificially truncated: it depicts a charmingly Victorian but utopian world in which conflicts abound but a handshake is a handshake.”<sup>16</sup> The reality, of course, is more complex. In the 1970s, a new *homo economicus* started to emerge in the principal-agent, the economics of information, and the radical political economy literature that reflected these complexities in human behavior. “The new economic man is not a Victorian gentleman: he is uncompromisingly thorough in pursuing objectives, and often he is less benign.”<sup>17</sup>

Williamson suggests that the new *homo economicus* will engage in a “full set of *ex ante* and *ex post* efforts to lie, cheat, steal, mislead, disguise, obfuscate, feign, distort, and confuse.”<sup>18</sup> If *homo economicus* behaves like this, one’s initial reaction might be that it is ridiculous to suggest that judges behave in this fashion. To take an extreme example, most would agree that it is absurd to postulate that judges (or most judges) are likely to expend effort on stealing. But, at the same time, this is also true for most members of the broader community. Judges, along with members of the broader community, might, however, engage in shirking of work effort, which represents a form of theft of time. And judges might engage in behavior designed to mislead or disguise their real output. The point is that *homo economicus* will behave in an opportunistic manner to achieve his self-interested objectives. It is important to take this into account when designing incentives to improve performance and mechanisms to monitor their implementation.

### III. WHAT MOTIVATES JUDGES TO PERFORM?

#### A. *Financial Incentives*

Before we can say anything about how to measure judicial performance and how such measurements will affect judicial behavior, it is important to say something about what motivates judges. If it is accepted that judges do indeed respond to incentives, an examination

---

15. Samuel Bowles & Herbert Gintis, *The Revenge of Homo Economicus: Contested Exchange and the Revival of Political Economy*, J. ECON. PERSP., Winter 1993, at 83, 83.

16. *Id.*

17. *Id.* at 84.

18. OLIVER E. WILLIAMSON, *THE ECONOMIC INSTITUTIONS OF CAPITALISM* 51 n.8 (1985).

of the motives of judges will shed light on the most appropriate incentive structure. In Australia and New Zealand, as in the United States, associate judges of the same court get paid the same amount. Given there are often sizeable differences in output across judges, wages do not reflect marginal product. The compensation of judges does not depend on their rulings, and the quality of the precedents set by the judge does not result in any direct pecuniary benefit.<sup>19</sup> From time to time, judges have expressed dissatisfaction with their compensation in the face of mounting workloads in both Australia and the United States.<sup>20</sup> In Australia, the issue of judicial compensation was a major issue at the last election in 2004, with the Federal opposition stating that if elected, it intended to abolish the present noncontributory judicial pension scheme and replace it with community-standard superannuation arrangements (the opposition was defeated). In part, this was a response to the blowout in the unfunded liability for judicial pensions. The law societies responded by arguing that this would have an adverse effect on the ability to attract the best people to the bench.<sup>21</sup> There is some evidence that pension entitlements influence the retirement decisions of judges. Squire<sup>22</sup> and Zorn and Van Winkle<sup>23</sup> find that pension eligibility has a statistically significant positive effect on the propensity to retire from the United States Supreme Court. Spriggs and Wahlbeck reach the same conclusion for the United States courts of appeals,<sup>24</sup> although Barrow and Zuk<sup>25</sup> and Hall<sup>26</sup> find that financial incentives are less important in the lower federal courts and state supreme courts, respectively. Maitra and Smyth find that pension eligibility has a significant positive effect on the retirement decision in the High Court of Australia.<sup>27</sup>

---

19. Thomas J. Miceli & Metin M. Coşgel, *Reputation and Judicial Decision-Making*, 23 J. ECON. BEHAV. & ORG. 31, 34 (1994); cf. Richard S. Higgins & Paul H. Rubin, *Judicial Discretion*, 9 J. LEGAL STUD. 129, 137-38 (1980) (suggesting that there is little economic payoff from “good” judging).

20. See Albert Yoon, *Love’s Labor’s Lost? Judicial Tenure Among Federal Court Judges: 1945-2000*, 91 CAL. L. REV. 1029 (2003).

21. Chris Merritt, *Latham Puts Judges’ Pay on Agenda*, AUSTRALIAN FIN. REV., Feb. 19, 2004, at 3; Chris Merritt, *Parity Call for Judges’ Pension Plan*, AUSTRALIAN FIN. REV., Feb. 14, 2004, at 3.

22. Peverill Squire, *Politics and Personal Factors in Retirement from the United States Supreme Court*, 10 POL. BEHAV. 180, 185-86 (1988).

23. Christopher J.W. Zorn & Steven R. Van Winkle, *A Competing Risks Model of Supreme Court Vacancies, 1789-1992*, 22 POL. BEHAV. 145, 154-55 (2000).

24. James F. Spriggs, II & Paul J. Wahlbeck, *Calling It Quits: Strategic Retirement on the Federal Courts of Appeals, 1893-1991*, 48 POL. RES. Q. 573, 590-91 (1995).

25. Deborah J. Barrow & Gary Zuk, *An Institutional Analysis of Turnover in the Lower Federal Courts, 1900-1987*, 52 J. POL. 457, 466-69 (1990).

26. Melinda Gann Hall, *Voluntary Retirements from State Supreme Courts: Assessing Democratic Pressures to Relinquish the Bench*, 63 J. POL. 1112, 1130 (2001).

27. Pushkar Maitra & Russell Smyth, *Determinants of Retirement on the High Court of Australia*, 81 ECON. REC. (forthcoming Dec. 2005) (manuscript at 20-22, on file with author).

Nevertheless, it is highly questionable whether there is much scope to improve judicial performance through improved financial rewards. The reason is that it is unlikely that many judges are motivated by financial returns. Monetary compensation only represents a small portion of the returns to judging, and it is a diminishing proportion the higher the court, because on higher courts the nonmonetary returns are higher.<sup>28</sup> While there are no studies on this point, the decision to take a judicial appointment is reflected in the facts. The opportunity cost of judging, purely in terms of monetary income, is high. Most judges would have been earning considerably more in private practice and, as such, give up substantial monetary incomes to join the bench.<sup>29</sup> Similarly, most judges could earn more by resigning from the bench and returning to private practice or even acting as a consultant to a major law firm, because of the premium that the prestige of being a former judge would attract.<sup>30</sup> While inadequate pension entitlements are sometimes given as a reason for judges retiring early and returning to private practice in Australia,<sup>31</sup> on the High Court of Australia at least, the examples of judges having to linger in office because of inadequate funds or having to return to private practice to replenish their coffers all occurred long ago.<sup>32</sup> Anecdotal evidence suggests that now, by the time they accept judicial appointment, most senior barristers in Australia have substantial sums invested in private pension plans and have accrued large numbers of assets, so they are not reliant on their judicial pensions to survive in retirement.

### B. Promotion Prospects

Judges rarely, if ever, would admit that desire for promotion was a motivating factor.<sup>33</sup> However, as Schauer notes in the U.S. context, while judges may be circumspect in admitting it, "it is hardly implausible to suspect that many trial judges desire to become appellate judges, and that most judges of intermediate appellate courts (including the United States Courts of Appeals) desire to become

---

28. Paul E. Greenberg & James A. Haley, *The Role of the Compensation Structure in Enhancing Judicial Quality*, 15 J. LEGAL STUD. 417, 423-25 (1986).

29. Cass, *supra* note 13, at 970-71.

30. Macey, *supra* note 13, at 630 n.9.

31. P.W. Young, *Judges' Retirement*, 71 AUSTRALIAN L.J. 733 (1997).

32. In the early years of the High Court of Australia, there is anecdotal evidence of judges remaining in office in the face of failing health for financial reasons. O'Conner (1903-1912) was unable to retire despite suffering from chronic illness from 1907 onward because he was pensionless; he eventually died in office. Griffith (1903-1919) suffered a stroke in 1917 and sat on few cases in his last two years on the Court, but he refused to retire because he had insufficient funds; he eventually retired when the government passed legislation granting him a pension.

33. Schauer, *supra* note 13, at 623.

judges of courts of last resort.”<sup>34</sup> There is some evidence that judges in Japan and the United States are motivated by prospects of promotion. Cohen presents empirical evidence that desire for promotion influences the behavior of district judges in the United States.<sup>35</sup> Taha shows that judges with higher prospects of promotion are more likely to publish their opinions.<sup>36</sup> In a series of articles, Ramseyer and Rasmusen find that promotion prospects influence the behavior of lower court judges in Japan.<sup>37</sup>

Desire to be promoted is also likely to be a motivating factor in New Zealand and the United Kingdom. Both nations have career-based judiciaries, with most judges of the English House of Lords being promoted from the English Court of Appeal and most judges of the New Zealand Court of Appeal coming from the New Zealand High Court. It is interesting to note that, in the case of both New Zealand and the United Kingdom, existing studies suggest that governments take into account both political factors (independence exercised by the judge) and indicators of the quality of judicial decision-making (such as the proportion of cases in which the judge has been reversed) in deciding whether and when the judge should be promoted.<sup>38</sup>

In Australia, promotion is less likely to be a motivating factor because the prospects of promotion are fewer, at least to the High Court of Australia. There have been only forty-four members of the High Court, including the current justices; and they have been appointed from a variety of places, including lower judicial office, politics, and private practice. Less than half (twenty-one) have held prior judicial office. Of those who have held prior judicial office, most (seventeen) have been appointed from state supreme courts, primarily from either Victoria or New South Wales, with the others being pro-

---

34. *Id.* at 631-32.

35. Mark A. Cohen, *Explaining Judicial Behavior or What's "Unconstitutional" About the Sentencing Commission?*, 7 J.L. ECON. & ORG. 183, 192-95 (1991) [hereinafter Cohen, *Explaining Judicial Behavior*]; Mark A. Cohen, *The Motives of Judges: Empirical Evidence from Antitrust Sentencing*, 12 INT'L REV. L. & ECON. 13, 17-27 (1992).

36. Ahmed E. Taha, *Publish or Perish? Evidence of How Judges Allocate Their Time*, 6 AM. L. & ECON. REV. 1, 21-23 (2004).

37. J. Mark Ramseyer & Eric B. Rasmusen, *Judicial Independence in a Civil Law Regime: The Evidence from Japan*, 13 J.L. ECON. & ORG. 259 (1997); J. Mark Ramseyer and Eric B. Rasmusen, *Why Are Japanese Judges So Conservative in Politically Charged Cases?*, 95 AM. POL. SCI. REV. 331 (2001); J. Mark Ramseyer and Eric B. Rasmusen, *Why Is the Japanese Conviction Rate So High?*, 30 J. LEGAL STUD. 53 (2001).

38. For a study of the determinants of promotion from the New Zealand High Court to the New Zealand Court of Appeal, see Pushkar Maitra & Russell Smyth, *Judicial Independence, Judicial Promotion and the Enforcement of Legislative Wealth Transfers—An Empirical Study of the New Zealand High Court*, 17 EUR. J.L. & ECON. 209 (2004). For a study of the determinants of promotion from the English Court of Appeal to the House of Lords, see Eli Salzberger & Paul Fenn, *Judicial Independence: Some Evidence from the English Court of Appeal*, 42 J.L. & ECON. 831 (1999).

moted from the Federal Court. Because there is a large pool of potential candidates across all the state supreme courts—at least compared with the smaller, more concentrated pools of candidates for judicial promotion in New Zealand and the United Kingdom—a judge’s probability of being promoted to the High Court of Australia is very low.

### C. *Respect of Colleagues*

There is a large law and economics literature that suggests that reputation among one’s peers is a powerful nonpecuniary motivating factor for many people.<sup>39</sup> In this respect, judges are likely to be the same as everyone else, although there is room to debate how wide a judge’s reference group will be. Miceli and Coşgel suggest that “how a judge is viewed by his or her colleagues, by law professors, law students, and the general public” is a motivating factor to perform well.<sup>40</sup> In Australia and New Zealand, where judges are less visible to the public than Justices of the United States Supreme Court, few members of the general public would know of any members of even the highest courts, so public esteem is unlikely to be a motivating factor. For judges of the High Court of Australia and the New Zealand Court of Appeal, whose opinions appear in casebooks and are debated in journal articles, they may well care how their output is viewed by legal academics and law students. In Australia, this might also be true for judges of state supreme courts, who are fairly visible to legal academics, at least those located in the same state.

For most “run of the mill” judges of lower courts, whose opinions are rarely discussed in law lectures or extracted in casebooks, respect of the academic community is unlikely to be a major motivating factor. These judges are likely to have a more narrow reference group, which is restricted to their colleagues in the same court and the lawyers who appear before them. The breadth of the reference group is important when thinking about what motivates judges. For a lower court judge, the main indicator of judicial performance might be how well he or she manages his or her caseload, and esteem among the relevant reference group will be given or withheld on this basis.<sup>41</sup> The

---

39. Richard H. McAdams, *Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination*, 108 HARV. L. REV. 1003 (1995); Eric A. Posner, *Symbols, Signals, and Social Norms in Politics and the Law*, 27 J. LEGAL STUD. 765 (1998); Schauer, *supra* note 13, at 633.

40. Miceli & Coşgel, *supra* note 19, at 32; *see also* Robert D. Cooter, *The Objectives of Private and Public Judges*, 41 PUB. CHOICE 107, 129 (1983) (arguing that judges seek prestige among the lawyers and litigants who bring cases before them); Posner, *supra* note 12, at 13 (arguing that judges care about how they are viewed by lawyers, but not litigants, who appear before them).

41. *See* Hillary A. Sale, *Judging Heuristics*, 35 U.C. DAVIS L. REV. 903, 910-11 (2002); Schauer, *supra* note 13, at 633.

art of writing good opinions, which are cited frequently or published in casebooks or legal commentary, is likely to be less important for these judges. Schauer even suggests that if lower court judges spend too much time writing opinions (or worse still journal articles), then this will have a negative effect on how their colleagues perceive them—they will be branded as too ambitious for spending too much time on self-advertising or as free riders because time spent on writing elaborate opinions detracts from time that could be spent on moving the docket along.<sup>42</sup> In contrast, judges of higher (appellate) courts, while also seeking the respect of their colleagues on the bench and practicing profession, have more opportunities to be cited in other opinions and in academic references. These judges will have greater incentive to spend more time writing better opinions.

#### D. Influence

Posner argues that other factors motivating judges are the power to influence outcomes through voting and the evolution of the law through creating precedent.<sup>43</sup> Landes and Posner posit that judges derive utility from imposing their policy preferences on the community.<sup>44</sup> Schauer suggests that “judges could plausibly select outcomes, or select substantive or methodological ‘trademarks,’ for the purpose of maximizing their own influence.”<sup>45</sup> In this respect, the time final appellate court judges spend on writing opinions to increase influence might be a function of the state of the law. It might be argued that, generally, the evolution of the law follows a steady-state growth path. There will be periods, however, when the law is in a state of historical flux. This was the case, for instance, in the High Court of Australia in the 1990s, which is often attributed to the activist legacy of the Mason Court (1987-1995). One commentator on the High Court in the 1990s noted that the law is contestably “more evident and accepted than ever” and that “there are seen to be many more legitimate approaches to argumentation than ever before.”<sup>46</sup>

There are analogies between the argument being made here and the external environment facing the firm. In the steady-state growth path, the firm will be operating in a technologically stable environment. In most circumstances, decisions will be routine, repetitive, and follow well-defined patterns; opportunities for technological breakthrough and, consequently, for making sizeable economic rents

---

42. Schauer, *supra* note 13, at 633.

43. Posner, *supra* note 12, at 15-23.

44. William M. Landes & Richard A. Posner, *Legal Precedent: A Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249, 274-75 (1976).

45. Schauer, *supra* note 13, at 633.

46. Graeme Orr, *Verbosity and Richness: Current Trends in the Craft of the High Court*, 6 TORTS L.J. 291, 292 (1998).

will be relatively small. In contrast, in periods of flux, there are more opportunities for endogenous innovation—which suggest several possible multiple trajectories where focus is on the creative, rather than allocative, functions of the market—and there is potential for high returns in a high-risk environment.<sup>47</sup> In these periods of instability in the evolution of the common law’s long-run growth path, as occurred in Australia in the 1990s, there is greater potential for judges to earn economic rents in the form of influence over the evolution of precedent. In these circumstances, one would expect the judge, as *homo economicus*, to write longer and more complex opinions in an attempt to win the battle of ideas. This is what happened in the High Court in the 1990s; there was a sizeable increase in the number and length of concurring opinions.<sup>48</sup> Leaving a legacy in the form of a body of precedent could be an important motivating factor for many final appellate judges to write complex and lengthy opinions. This is more pressing in Australia and New Zealand relative to most jurisdictions in the United States, because in Australia and New Zealand judges face mandatory retirement ages.

The objective of Choi and Gulati’s tournament is to identify highfliers for promotion.<sup>49</sup> While the highfliers might have a relatively low marginal time preference for leisure, there is considerable evidence that “run of the mill” judges attempt to minimize workloads. This is most true of lower courts. Cohen shows that the district judges in the United States behave so as to minimize workloads.<sup>50</sup> Taha finds that district judges with lighter workloads have more time to publish their decisions.<sup>51</sup> However, it is also the case for some judges on appellate courts. Clearly, not all judges are motivated by the prospect of leaving a legacy. Throughout the history of the United States Supreme Court, there have been Justices who appear to have had a preference for minimizing their workload by writing few opinions and free riding on the efforts of others.<sup>52</sup> This is also the case for the High Court of Australia, with some Justices well known for hav-

---

47. See NEIL M. KAY, *THE EVOLVING FIRM: STRATEGY AND STRUCTURE IN INDUSTRIAL ORGANIZATION* 58-61 (1982); Nicholas Kaldor, *The Irrelevance of Equilibrium Economics*, 82 *ECON. J.* 1237, 1244-46 (1972).

48. For jurimetric evidence on opinion-writing trends over the history of the High Court of Australia, which documents an increase in the length of opinions and number of concurring opinions in the 1990s, see Matthew Groves & Russell Smyth, *A Century of Judicial Style: Changing Patterns in Judgment Writing on the High Court 1903-2001*, 32 *FED. L. REV.* 255 (2004).

49. See Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 *S. CAL. L. REV.* 23, 29-30 (2004).

50. Cohen, *Explaining Judicial Behavior*, *supra* note 35, at 193.

51. Taha, *supra* note 36, at 20.

52. See David P. Currie, *The Most Insignificant Justice: A Preliminary Inquiry*, 50 *U. CHI. L. REV.* 466 *passim* (1983); Frank H. Easterbrook, *The Most Insignificant Justice: Further Evidence*, 50 *U. CHI. L. REV.* 481, 492-96 (1983).

ing a proclivity for joining in joint opinions or writing short concurring opinions.

#### IV. HOW SHOULD WE MEASURE AND RANK JUDICIAL PERFORMANCE?

##### A. *Opinions Published*

In their proposed tournament of judges, Choi and Gulati treat the number of published opinions as an important indicator of judicial productivity.<sup>53</sup> The rationale is that

it takes greater effort and skill to write more published opinions. Measuring effort exerted by a judge is an important element of the tournament because comparing the past effort levels of the various judges (a) helps predict future effort levels (we want Justices who will exert high levels of effort) and (b) helps determine who among the lower court judges should be rewarded for their efforts . . . .<sup>54</sup>

On the face of it, this reasoning is sound in the United States context. There is much evidence that writing published opinions involves expending more effort than writing unpublished opinions. In their analysis of published and unpublished opinions by the United States courts of appeals, Reynolds and Richman conclude that “anyone who reads even a small number of unpublished opinions must conclude, given their brevity and informality, that considerable effort has been spared in their preparation.”<sup>55</sup> This reasoning, however, does not extend to Australian and New Zealand courts, which follow the English practice of seriatim opinion writing and use a different reporting system.

In Australia and New Zealand, there is no single signed opinion of the court where the author is readily identifiable. Instead, each judge writes his or her own opinion. Sometimes two or more judges, or more rarely the whole court, give a joint opinion, but who actually wrote the opinion in such situations is not disclosed. Those who endorse the opinion are listed in order of seniority at the beginning. The actual opinion might have been written by one or more of the judges who “sign on.” Attempts to measure judicial output over time are further complicated by changing opinion-writing practices over time. For example, on the High Court of Australia, particularly in its early years, it was common for Justice X to write a short concurring opinion of the form “I agree” with Justice Y, while at other points in time, Justices X and Y would write a joint opinion.<sup>56</sup> In the former

---

53. See Choi & Gulati, *supra* note 49, at 42-47.

54. *Id.* at 42 (footnote omitted).

55. William L. Reynolds & William M. Richman, *An Evaluation of Limited Publication in the United States Courts of Appeals: The Price of Reform*, 48 U. CHI. L. REV. 573, 594-95 (1981).

56. Groves & Smyth, *supra* note 48, at 267.

case, the contributions of Justices *X* and *Y* are separable products, but in the latter case, it is impossible to know who wrote the opinion.

The system of law-reporting is different in Australia and New Zealand in the sense that in the past, the decision to publish a case was based on its perceived precedent value, rather than the inherent reasoning of the judge(s). Because only a small proportion of decisions were reported in the official and unofficial law reports, published opinions were not a good indicator of effort. In Australia and New Zealand, unpublished opinions have precedent value in the same way as published opinions. Before the widespread use of the Internet, relatively few unpublished opinions were cited because of the difficulty of locating them. However, since the development of the Internet, unpublished decisions are readily available on the Web, free of charge, for a range of courts, and these are widely cited *as if* they were in the law reports.<sup>57</sup> Indeed, publishing lags, coupled with high purchase costs, are increasingly making law reports redundant. At least at the appellate level, there is no difference in effort needed to produce published opinions that appear in the law reports and unpublished opinions that are available on the Internet.

The practice of seriatim opinion writing, where the author of joint opinions is not disclosed, raises the well-known free-rider problem with team production. The problem with team production, as Alchian and Demsetz note, is that the individual marginal products of team members cannot be verified.<sup>58</sup> The implication is that determining who shirks and by how much is not possible with such a joint product. This comes back to the motivations of judges discussed above. The “run of the mill” judges will have an incentive to shirk by participating in joint opinions to minimize their workload in the absence of effective monitoring.<sup>59</sup> This will be particularly true in the face of increasing workloads. There are instances throughout the history of the High Court of Australia where this is apparent. For example, even casual inspection of the *Commonwealth Law Reports* suggests that Rich (1913-1950) and McTiernan (1930-1976), the two longest-serving members of the High Court of Australia, frequently free rode on the work ethic and intellect of Dixon (1929-1964), who is Australia’s most celebrated judge, by participating in joint opinions or writ-

---

57. Decisions of a number of Australian courts, including the High Court of Australia, are available free of charge on the AustLII databases. *AustLII Databases*, Australian Legal Information Institute, <http://www.austlii.edu.au/databases.html> (last updated Apr. 7, 2005).

58. Armen A. Alchian & Harold Demsetz, *Production, Information Costs, and Economic Organization*, 62 AM. ECON. REV. 777, 779 (1972).

59. See RICHARD A. POSNER, *OVERCOMING LAW* 123-26 (1995).

ing short concurring opinions.<sup>60</sup> Starke (1920-1950), who was a contemporary of all three judges on the Court, caricatured McTiernan as a “parrot” who always copied Dixon.<sup>61</sup> While this characterization was motivated by personal differences, it is supported by studies of voting behavior, which show that McTiernan exhibited a high level of agreement with Dixon.<sup>62</sup> Diaries that Dixon kept provide some insights into the free-riding problem:

Numerous entries show [Dixon] regularly helping Rich with his judgments (even when Dixon had not sat on the case), and occasionally he helped McTiernan with his. Later entries state that he wrote some of Rich’s judgments for him, though it is possible that Dixon meant he was writing *sections* of Rich’s judgments.<sup>63</sup>

Under these circumstances, it will be difficult to detect the true level of effort judges expend using opinions as a measure of performance. The only real monitoring device will be norms on the court as to whether a judge’s contribution to the workload of the court is acceptable, and this is difficult for those outside the court to effectively observe, let alone quantify in a meaningful manner.

One approach to detecting effort levels, which distinguishes between full written opinions and short concurring opinions, would be to examine the length of the opinion. The argument is that longer opinions address the relevant issues more fully and require more effort. Such a measure attempts to tap into the motives of judges. Judges who want to leave their mark on the law will write longer opinions. Currie uses “pages per year” in constitutional cases to explore the question of who is the most insignificant Justice in the history of the United States Supreme Court.<sup>64</sup> In their study of the determinants of promotion from the English Court of Appeal to the House of Lords, Salzberger and Fenn attempt to confront the free-rider issue by giving short concurring opinions a lower weighting.<sup>65</sup>

There are, however, problems with focusing on the length of opinion. One problem is that it gives a biased indicator of effort expended over time. As discussed earlier, if judges behave as *homo economicus*, opinions will be longer when there are more opportunities to leave a legacy, which will occur at points in time when the evolution of the

60. Russell Smyth, *Judicial Interaction on the Latham Court: A Quantitative Study of Voting Patterns on the High Court 1935-1950*, 47 AUSTRALIAN J. POL. & HIST. 330, 340-42 (2001).

61. See Clem Lloyd, *Not Peace But a Sword!—The High Court Under J.G. Latham*, 11 ADEL. L. REV. 175, 181 (1987).

62. See Russell Smyth, *Explaining Voting Patterns on the Latham High Court 1935-50*, 26 MELB. U. L. REV. 88, 100 tbl.1 & tbl.2 (2002); Smyth, *supra* note 60, at 338-40.

63. Philip Ayres, *Dixon Diaries*, in THE OXFORD COMPANION TO THE HIGH COURT OF AUSTRALIA 222, 223 (Tony Blackshield et al. eds., 2001).

64. Currie, *supra* note 52, at 469.

65. Salzberger & Fenn, *supra* note 38, at 837.

law is not on its long-run growth path. Measures such as “pages per year” have to be examined relative to opportunities and in the context of changing norms governing opinion writing over time.<sup>66</sup> Another problem is that, even when examining performance at a given point in time, there is an argument that verbosity should not be rewarded.<sup>67</sup> The increase in the length of opinions on the High Court of Australia in the 1990s, interspersed with increasing numbers of footnotes and citations to academic references, is a trend which has been criticized rather than applauded in some quarters.<sup>68</sup> It has been criticized for increasing the cost to the “consumers” of the opinion, who are defined as those who need to take time to read and digest the implications of the opinion.<sup>69</sup>

Another problem in counting opinions, particularly in the Australian and New Zealand context where there is no effective distinction between published and unpublished opinions, is that there is no product differentiation between publishing outlets—that is, there are no “first-tier,” “second-tier,” and “third-tier” law reports. Studies that attempt to measure the productivity of academic departments or individual academics by counting publications will invariably use some sort of quality-adjusted output measure, so that a publication in a first-tier journal will count for more than a publication in a third-tier journal. An interesting feature of publishing judicial opinions is that at the time of publication, there is either no or little peer assessment of quality. This, of course, is also true of student-edited law reviews in the United States, although there are first-tier, second-tier, and third-tier law reviews; and through the multiple submission process, it is arguable that a manuscript will end up at about its right level.<sup>70</sup> In a Coaseian world, free of transaction costs, one solution might be to have a series of electronic law reports (tier one, tier two, tier three, tier four, and rejected) similar to the Berkeley Electronic Press’s “families” of journals.<sup>71</sup> In practice, though, the transaction costs in-

---

66. See Easterbrook, *supra* note 52, at 491-92 (arguing that “pages per year” has to be examined relative to opportunities and viewed along with other indicators as a measure of judicial performance).

67. Cf. Brian Leiter, *Measuring the Academic Distinction of Law Faculties*, 29 J. LEGAL STUD. 451, 467-68 (2000) (arguing that productivity is an important measure for determining the academic distinction of law faculties but that there must be quality checks in place to avoid giving credit for “junk”).

68. E.g., Enid Campbell, *Reasons for Judgment: Some Consumer Perspectives*, 77 AUSTRALIAN L.J. 62, 63 (2003).

69. *Id.* at 64.

70. Note that this differs from Australia, where student-edited law reviews are peer reviewed.

71. See Berkeley Electronic Press, at <http://www.bepress.com/index.html> (last visited Apr. 12, 2005). For more information about the quality-rating system, among other things, utilized by the Berkeley Electronic Press, see *bepress: A Revolution in Scholarly Publishing*, Berkeley Electronic Press, at <http://www.bepress.com/revolution.html#qr> (last visited Apr. 12, 2005).

volved in such a proposal would be high, and it would be complicated by the perceived precedent value of the decision, which may have nothing to do with the effort expended or quality of the writing. There is peer assessment of opinion quality at the time of consumption—if the consumer does not think the publication is worthwhile, it will not be cited. However, this raises the issue of how to determine quality. When citing journals, academics can use the ranking of the journal as a signalling device of quality because highly cited journals tend to be highly read journals. No such signalling device exists for judges. Typically consumers will base their decision on the reputation capital of the individual judge or, perhaps, the circuit.

### B. Citations

Choi and Gulati suggest that citations to published opinions are a market test for the quality of opinions.<sup>72</sup> Adjusted citation counts to judicial opinions have been widely used in recent years to measure what has variously been described as a judge's influence, prestige, or reputation.<sup>73</sup> These studies generally emphasize that citations are at best a rough proxy for quality, although in the subsequent empirical implementation of their tournament, Choi and Gulati treat citation counts as a measure of the quality of the judicial opinion.<sup>74</sup> Citations have also been used in studies of the relationship between aging and judicial productivity, that is, as a measure of the quality of judicial output as a judge gets older.<sup>75</sup> Some have criticized the use of citation counts as a measure of quality, focusing on the limitations of their

---

72. Choi & Gulati, *supra* note 1, at 305-06.

73. See RICHARD A. POSNER, *CARDOZO: A STUDY IN REPUTATION* 74-91 (1990); Mita Bhattacharya & Russell Smyth, *The Determinants of Judicial Prestige and Influence: Some Empirical Evidence from the High Court of Australia*, 30 J. LEGAL STUD. 223 (2001) [hereinafter Bhattacharya & Smyth, *Judicial Prestige and the High Court*]; David Klein & Darby Morrisroe, *The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals*, 28 J. LEGAL STUD. 371 (1999); Montgomery N. Kosma, *Measuring the Influence of Supreme Court Justices*, 27 J. LEGAL STUD. 333 (1998); William M. Landes et al., *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271 (1998); Richard A. Posner, *The Learned Hand Biography and the Question of Judicial Greatness*, 104 YALE L.J. 511, 534-40 (1994) (book review); Russell Smyth & Mita Bhattacharya, *What Determines Judicial Prestige? An Empirical Analysis for Judges of the Federal Court of Australia*, 5 AM. L. & ECON. REV. 233 (2003) [hereinafter Smyth & Bhattacharya, *Judicial Prestige and the Federal Court*].

74. Choi & Gulati, *supra* note 49, at 48-61.

75. See Mita Bhattacharya & Russell Smyth, *Aging and Productivity Among Judges: Some Empirical Evidence from the High Court of Australia*, 40 AUSTRALIAN ECON. PAPERS 199 (2001) [hereinafter Bhattacharya & Smyth, *Aging and Productivity and the High Court*]; Russell Smyth & Mita Bhattacharya, *How Fast Do Old Judges Slow Down? A Life Cycle Study of Aging and Productivity in the Federal Court of Australia*, 23 INT'L REV. L. & ECON. 141 (2003) [hereinafter Smyth & Bhattacharya, *Aging and Productivity and the Federal Court*].

use.<sup>76</sup> One such limitation is evidence of like-minded citations where “conservative” or “liberal” judges tend to cite members of their respective group to the exclusion of the other camp.<sup>77</sup> This is problematic where the court is dominated by appointments on one side of the political fence. This has traditionally been the case in Australia and New Zealand, where most appointments to the High Court of Australia and New Zealand Court of Appeal have been made by the “conservative” side of politics. The problem in Australia, at least, has been exacerbated by the unorthodox approach to opinion writing by some Labor appointments, which have been shunned by their more “conservative” colleagues. While this issue and problems with citation counts are well known, it is generally recognized in a wide range of disciplines that in spite of these limitations, there is a correlation between citations received and the quality of output.<sup>78</sup>

One problem with using citation counts to judicial opinions in the United States is the problem of ghostwriting, where it is the law clerk, rather than the judge, who writes the opinion.<sup>79</sup> As Landes, Lessig, and Solimine note, “[c]itations may reflect the influence or quality of law clerks as ‘ghost writers’ rather than that of judges . . . .”<sup>80</sup> However, as Choi and Gulati argue, the problem is tempered because judges who rely heavily on clerks will produce opinions of more varying quality, because it will be difficult even for a high-quality judge to consistently retain high-quality clerks.<sup>81</sup> Moreover, even with judges who rely heavily on law clerks, citations might still reflect performance as a manager.<sup>82</sup> The reasoning is that there will be a signalling device where the best clerks (who write the best opinions and get cited the most) will be attracted to the best judges, although the market for clerks is imperfect.<sup>83</sup> The issue of ghostwriting does not impede using citation counts in Australia. The limited anecdotal evidence that exists from both judges<sup>84</sup> and former clerks<sup>85</sup> suggests

76. Arthur Austin, *The Reliability of Citation Counts in Judgments on Promotion, Tenure, and Status*, 35 ARIZ. L. REV. 829 (1993). For counterarguments in the context of counting citations to judicial opinions, see Landes et al., *supra* note 73, at 271-76; Richard A. Posner, *An Economic Analysis of the Use of Citations in the Law*, 2 AM. L. & ECON. REV. 381 (2000).

77. Austin, *supra* note 76, at 833-34.

78. See Choi & Gulati, *supra* note 49, at 48 n.38 (citing a multitude of studies “that either suggest or assume a relationship between citation counts and quality”).

79. Nadine J. Wichern, Comment, *A Court of Clerks, Not of Men: Serving Justice in the Media Age*, 49 DEPAUL L. REV. 621, 649-50 (1999).

80. Landes et al., *supra* note 73, at 274.

81. Choi & Gulati, *supra* note 49, at 52.

82. Cass, *supra* note 13, at 989-91.

83. Epstein, *supra* note 11, at 841-44.

84. Michael Kirby, *What Is It Really Like to Be a Justice of the High Court of Australia? A Conversation Between Law Students and Justice Kirby*, 19 SYDNEY L. REV. 514, 519-20 (1997); P.W. Young, *Judgment Writing*, 70 AUSTRALIAN L.J. 513, 513-16 (1996).

that judges write their own opinions in Australia. There is, however, another practical problem with using citations to measure the quality of judicial opinions in Australia and New Zealand. This stems from the earlier observation that there is no signed opinion of the court. Thus, when a case is cited, it is impossible in most instances to attribute authorship. Therefore, existing studies of judicial influence and productivity for Australian courts have relied on invocation rates.<sup>86</sup> Choi and Gulati suggest that invocation rates are more an indication of reputation, while citation counts are a measure of quality.<sup>87</sup> It can be argued that invocation rates are preferable to citation counts as a measure of the influence of the opinion on the citing judge's thinking, even in the United States, because there is no requirement that the citing judge go the extra step and refer to the author of the cited opinion by name.<sup>88</sup>

### C. Reversals

Posner suggests that while judges do not like to be reversed, aversion to reversal does not figure prominently in the judicial utility function. He reasons that most reversals reflect differences in judicial philosophy or legal policy, rather than error by the judge below.<sup>89</sup> However, while some reversals will undoubtedly reflect policy issues, it is reasonable to assume that a judge gets disutility when his or her decision is reversed by judges who are in higher courts or by future judges.<sup>90</sup> Elsewhere Posner accepts that since judges might seek to be promoted, they may display extreme "sensitivity to being reversed."<sup>91</sup> Cass agrees: "Judicial reversal reflects professional criticism by other professionals. It will likely be a negative for almost all judges in almost all circumstances."<sup>92</sup> For this reason, judges will expend effort on ensuring that their opinions are well-reasoned to reduce the occurrence of reversal. The proportion of cases in which a judge has been reversed is something that is readily quantifiable as one component of a tournament of judges.

---

85. Andrew Leigh, *Behind the Bench: Associates in the High Court of Australia*, 25 ALTERNATIVE L.J. 295, 296-97 (2000).

86. Bhattacharya & Smyth, *Aging and Productivity and the High Court*, *supra* note 75, at 203; Bhattacharya & Smyth, *Judicial Prestige and the High Court*, *supra* note 73, at 233-34; Smyth & Bhattacharya, *Aging and Productivity and the Federal Court*, *supra* note 75, at 149-50; Smyth & Bhattacharya, *Judicial Prestige and the Federal Court*, *supra* note 73, at 238-39.

87. Choi & Gulati, *supra* note 49, at 58-59.

88. Klein & Morrisroe, *supra* note 73, at 375-76.

89. Posner, *supra* note 12, at 14-15.

90. Youngsik Lim, *An Empirical Analysis of Supreme Court Justices' Decision Making*, 29 J. LEGAL STUD. 721, 723 n.8 (2000).

91. RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* § 19.6, at 543 (6th ed. 2003).

92. Cass, *supra* note 13, at 984.

Previous studies for a range of courts both inside and outside the United States have used reversals as a measure of judicial performance.<sup>93</sup> Posner uses reversals by the Supreme Court to assess the quality of the United States Court of Appeals for the Ninth Circuit relative to other circuits.<sup>94</sup> Reversals have also been employed in studies of the English Court of Appeal and the New Zealand High Court as an indicator of the quality of decisionmaking.<sup>95</sup> While Maitra and Smyth found that reversals had a statistically insignificant effect on promotion prospects of judges on the New Zealand High Court,<sup>96</sup> Salzberger and Fenn found that those judges on the English Court of Appeal who had a higher proportion of their decisions reversed by the House of Lords were less likely to be promoted.<sup>97</sup>

#### D. *Judicial Gowns or Academic Robes?*

Judges' contributions to academic scholarship in the form of law review articles has been described as a form of judicial leadership.<sup>98</sup> If so, it is arguable that extrajudicial contributions such as academic writing should be a part of the tournament. Currently, if anything, judges in the United States vying for appointment to the Supreme Court have a disincentive to publish in law reviews, because it contributes to a Bork-like paper trail of their views. There is empirical support for this view. Gaille examined the academic publication patterns of judges on the United States courts of appeals prior and subsequent to the Bork confirmation hearings and found that judges published less in academic journals after the hearings.<sup>99</sup>

The main argument supporting the view that academic publications should be viewed favorably is that there are complementarities between good judging and good writing abilities reflected in law review articles. Thus judges who are novel thinkers are more likely to publish in law reviews, and this proclivity is correlated with citations received.<sup>100</sup> On the United States courts of appeals, academics turned judges, such as Easterbrook and Posner, are cases in point.<sup>101</sup> How-

93. See, e.g., Richard A. Posner, *Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality*, 29 J. LEGAL STUD. 711 *passim* (2000).

94. *Id.*

95. E.g., Maitra & Smyth, *supra* note 38, at 221; Salzberger & Fenn, *supra* note 38, at 837.

96. Maitra & Smyth, *supra* note 38, at 226-27.

97. Salzberger & Fenn, *supra* note 38, at 842.

98. Kenneth F. Ripple, *Judges on Judging: The Judge and the Academic Community*, 50 OHIO ST. L.J. 1237, 1241-42 (1989).

99. S. Scott Gaille, *Publishing by United States Court of Appeals Judges: Before and After the Bork Hearings*, 26 J. LEGAL STUD. 371, 372-73 (1997).

100. Klein & Morrisroe, *supra* note 73, at 383.

101. See Choi & Gulati, *supra* note 49, at 74 (finding that Easterbrook and Posner come out at the top in the tournament of judges).

ever, in these cases, it is their citation counts, rather than scholarly publication records, that make them good judicial performers.<sup>102</sup> There is an alternative argument that outsiders on the court are more likely to publish in law reviews because their views are not gaining acceptance on the court.<sup>103</sup> There is casual evidence to support this on the current High Court of Australia: Justice Kirby has the highest dissent rate in the history of the Court and is also the most prolific publisher in law reviews, publishing three times more than any other Justice.<sup>104</sup> If judges behave as *homo economicus*, then including outside activities, such as writing journal articles, in the tournament of judges also potentially creates perverse incentives. As Choi and Gulati note, "Giving a judge credit for doing other things will diminish the incentive to spend time on opinion writing."<sup>105</sup>

#### V. IS IT APPROPRIATE TO MEASURE AND RANK JUDICIAL PERFORMANCE?

Choi and Gulati's whole rationale for a tournament of judges premised on objective quantitative indicators is that it would make the judicial appointment process more transparent and less politicized.<sup>106</sup> Some, however, have expressed reservations about the extent to which judicial performance can be quantified. The outcome of a tournament of judges has to be viewed subject to these reservations, which fall into two main categories. The first concerns the merits of applying empirical techniques to the study of judicial behavior because of inherent data problems. The second is that quantitative studies cannot capture important qualitative dimensions associated with good judging, and in particular, quantitative studies are unable to measure the positive effects of collegiality. In this Part I, provide a brief review of these concerns.

The law and economics movement, together with allied researchers in political science, has been at the forefront of explaining judicial behavior using quantitative tools. One argument against the expansion of economics beyond its traditional domain of markets is that the data is too thin to reliably test hypotheses formulated in the social sciences.<sup>107</sup> While markets generate a range of data on variables

---

102. *Id.* at 68 tbl.11.

103. Peter McCormick, *Judges, Journals and Exegesis: Judicial Leadership and Academic Scholarship*, 45 U.N.B. L.J. 139, 145-46 (1996).

104. On dissent rates of Australian Justices, see Groves & Smyth, *supra* note 48, at 268-73. On academic publication patterns of Australian judges, see Russell Smyth, *Judges and Academic Scholarship: An Empirical Study of the Academic Publication Patterns of Federal Court and High Court Judges*, 2 QUEENSL. U. TECH. L. & JUST. J. 198 (2002).

105. Choi & Gulati, *supra* note 49, at 43.

106. *Id.* at 29-30.

107. See Richard A. Posner, *The Law and Economics Movement*, 77 AM. ECON. REV. PAPERS & PROC. 1, 3 (1987).

of interest—such as output, prices, and employment—many studies of courts and judges are restricted to a universe of published, written decisions. The focus on such decisions, along with an emphasis on appellate courts, reflects greater data availability, reduces the extent to which findings can be generalized, and raises the prospect of selection bias.<sup>108</sup> In some studies of judicial behavior, sample size has been problematic, with far-reaching conclusions drawn on the basis of a small number of observations.

In studies of judicial decisionmaking, there are also problems with comparability and experimental design.<sup>109</sup> Prior to the 1980s, most empirical work on judicial decisionmaking used discrete datasets bearing on a limited set of research issues, which made it difficult to compare and replicate results because different studies often made different assumptions or collected data on variables in different ways, or both. Longitudinal studies, employing time series data over time, have often had to make the heroic assumption that cases from different eras are sufficiently similar that the results are comparable and generalizable. One approach to addressing the limitations on the assumption of comparability has been to explore alternative research designs such as experimental methods.<sup>110</sup> The problem, though, with experimental studies that employ simulated cases to gauge judicial performance is the associated loss of authenticity.<sup>111</sup>

Having said this, empirical researchers aware of the limitations have recently made considerable progress dealing with many of these data issues. One development has been a trend in the United States toward the use of datasets containing both published and unpublished decisions in recognition of the biases introduced with only using published decisions. A second occurrence in the United States has been the emergence of multiuser datasets designed for use in a range of research problems.<sup>112</sup> These datasets are not restricted to the Supreme Court but have also been developed for the courts of appeals and state supreme courts.<sup>113</sup> A third development, which is particularly apt in the context of this Essay's Antipodean focus, is that studies looking at different aspects of judicial performance are starting to emerge for courts outside the United States. It has to be said,

---

108. Heise, *supra* note 13, at 843-44; R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. 1105, 1128-29 (2004).

109. The rest of this paragraph draws on Heise, *supra* note 13, at 845-48.

110. See, e.g., Peter J. van Koppen & Jan ten Kate, *Individual Differences in Judicial Behavior: Personal Characteristics and Private Law Decision-Making*, 18 LAW & SOC'Y REV. 225 (1984) (administering psychological tests to Dutch judges participating in a court simulation).

111. Heise, *supra* note 13, at 846.

112. *Id.* at 830.

113. See *id.* at 848-49.

though, that progress on this front has been patchy. Atkins' observations, made at the beginning of the 1990s, are still true now: "Despite much progress during the past few decades in the study of courts and judicial behavior, most theory and data developed by social scientists for understanding legal systems still remain very much the product of, and thus bound to, the inevitable peculiarities of the U.S. context."<sup>114</sup> Atkins goes on to suggest, "This must represent a curious inconsistency in the logic of our research agenda, since the scientific method to which we adhere is supposed to encourage theory building and description at the broadest level possible."<sup>115</sup> These comments underpin the need, emphasized in the previous Part, to be flexible in the design of a tournament of judges intended to measure judicial performance and to cater to different institutional contexts.

Another objection to attempting to quantify judicial performance is the argument that quantitative tools are not apposite to capture the multifaceted aspects of being a good judge. Hensler, for instance, argues, "Researchers simply do not have available very good quantitative approaches to studying large social organizations [such as courts] or interaction processes [within courts]."<sup>116</sup> The argument is that there are important qualitative aspects to being a good judge which are difficult to quantify in a tournament. In addition to diligence and ability, which can be measured using opinion and citation counts, other generally agreed useful qualities include integrity, fairness, temperament, and collegiality.<sup>117</sup> Judge Edwards has been one of the strongest critics of attempts to quantify judicial performance on the basis that collegiality cannot be measured:

Collegiality is a *qualitative* variable in appellate decision making, because it involves mostly private personal interactions that are not readily susceptible to empirical study. Regression analysis does not do well in capturing the nuances of human personalities and relationships, so empirical studies on judicial decision making that rely solely on this tool are inherently flawed.<sup>118</sup>

Some of these desirable qualities, though, have been captured in surveys such as that of the American Bar Association.<sup>119</sup> Moreover, contrary to what the quote from Judge Edwards implies, there is much

---

114. Burton M. Atkins, *Party Capability Theory as an Explanation for Intervention Behavior in the English Court of Appeal*, 35 AM. J. POL. SCI. 881, 881 (1991).

115. *Id.*

116. Deborah R. Hensler, *Researching Civil Justice: Problems and Pitfalls*, 51 LAW & CONTEMP. PROBS. 55, 63 (1988).

117. See Susanne Di Pietro et al., *Judicial Qualifications and Judicial Performance: Is There a Relationship?*, 83 JUDICATURE 196, 197-98 (2000).

118. Edwards, *supra* note 14, at 1656 (footnote omitted); see also Harry T. Edwards, *Collegiality and Decision Making on the D.C. Circuit*, 84 VA. L. REV. 1335 (1998).

119. Laura E. Little, *The ABA's Role in Prescreening Federal Judicial Candidates: Are We Ready to Give Up on the Lawyers?*, 10 WM. & MARY BILL RTS. J. 37, 41 (2001).

evidence of the success of regression analysis in capturing these nuances of human behavior. For instance, there have been hundreds of published articles in which regression has been used to analyze scores from personality inventories, such as the Minnesota Multi-phasic Personality Inventory, the 16PF, and the Myers-Briggs Type Indicator, where regression analysis has been used precisely for this purpose.<sup>120</sup>

Having said this, Edwards' concerns have important implications for how the efficiency aspects of measuring judicial performance are interpreted. The Alchian and Demsetz team production model, with its emphasis on monitoring the individual, is concerned with allocative efficiency.<sup>121</sup> In the Alchian and Demsetz model, the firm and the market are opposite sides of the same coin.<sup>122</sup> In the context of this discussion, "the firm" can be interpreted as the court. In the Alchian and Demsetz conception, there is no role for the court as a social organization. The efficiency gains stem from specialization and the detailed division of labor. But the emphasis Judge Edwards and others place on the importance of collegiality in the court suggests that in the court's team production, there are potential benefits from organizational learning. These stem from continual interaction between members of the court that cannot be captured with a more narrow focus on allocative efficiency. There are positive externalities in team production from collegiality that are manifest in the form of collective learning on the court. The relevant efficiency criteria is productive efficiency rather than allocative efficiency, and the efficiency gains accrue from economies of scope rather than economies of scale.<sup>123</sup> Edwards' concerns suggest that attempts to measure judicial performance using similar tournament designs to Choi and Gulati's focus heavily on the allocative-efficiency dimension. The productive-efficiency dimension tends to be downplayed. Thus, it is important to recognize that the hardheaded figures which come out of a tournament based on citation counts, while insightful, tell only part of the story.

A different issue is whether the acceptance of broader aspects of human behavior—such as collegiality, integrity, and fairness as desirable attributes in a judge—is commensurate with a conceptualization of the one-dimensional *homo economicus*. One might argue that

---

120. For a meta-analysis, see James M. Schuerger et al., *Factors That Influence the Temporal Stability of Personality by Questionnaire*, 56 J. PERSONALITY & SOC. PSYCHOL. 777 (1989).

121. See Alchian & Demsetz, *supra* note 58, at 779-81.

122. *Id.* at 793-94.

123. For an extended theoretical discussion of these ideas in a different context, see Russell Smyth & Dic Lo, *Theories of the Firm and the Relationship Between Different Perspectives on the Division of Labour*, 12 REV. POL. ECON. 333 (2000).

the model on which he is conceived is at best myopic and at worst incommensurable with regard to the component of perceived moral obligation that arguably influences much of human social behavior, including behavioral intentions and, importantly, attitudes.<sup>124</sup> If this is the case, “Justice *Homo Economicus*” might be a distorted lens through which to view judicial behavior. Instead, perhaps we should be asking what motivates “Justice *Homo Sociologicus*,” the (irrational?) mirror image of economic man. The teleological underpinnings of the two judges differ in that *homo economicus* has essentially an ego-goal orientation, whereas *homo sociologicus* is a role player acting under the instructions of his environment. While *homo economicus* is assumed to be motivated by self-interest, *homo sociologicus* is conceived in the constructivist tradition, which acknowledges that all actors are resident within some social context. The social context, or more precisely the social norms that emerge from within the social context (including norms governing judges’ interaction on the court), influence behavior and, important in the context of measuring judicial performance, the pursuit of goals. This is because they impose upon judges social roles, inherent in which are prescriptions about behavior, group interaction, and moral responsibilities.

Another challenging aspect of measuring and ranking judicial performance centers on how to choose the most appropriate performance standard referents. Perhaps due to the fact that we have well-developed quantitative methods to assess individual performance relative to group performance, performance measurement has traditionally (and it seems stubbornly) relied upon extrinsic normative referents. Reference to extrinsic norms lends itself nicely to performance ranking. This is useful, because a ranking is a clear and concise manner in which to understand performance. Ranks are conceptually simple, and they are not readily misinterpretable; so in these terms, adopting a ranking approach maximizes the golden rules of measurement: validity and reliability. Having said that, it seems strange to suggest that these normatively referenced performance measurement “habits” that we have gotten into might indeed not be ideal when it comes to a tournament of judges. But consider this: While one’s performance ranking within a group is undoubtedly informative, it is limited, because assessing performance relative to a group norm and to observed group performance bounds assumes a commonality between the *observed* upper performance bound of the group and the *potential* upper performance bound of the highest flier. But what if the highest flier is in fact underperforming in terms of

---

124. See Paul Sparks & Richard Shepherd, *The Role of Moral Judgments Within Expectancy-Value-Based Attitude-Behavior Models*, 12 ETHICS & BEHAV. 299, 299-302 (2002).

his or her own potential performance bound? What if he or she could do better, but simply does not need to in order to maintain a number one ranking?

This brings us fundamentally to the *raison d'être* of a tournament of judges. One fundamental purpose of a tournament of judges indeed should be to identify those ripe for promotion—and these are likely the normatively referenced highfliers. Another purpose, of course, should be to identify those who are dragging their heels. But the assumption that the highfliers never drag their heels is a dangerous one. Certainly in competitive sports, it is entirely conceivable that some elite athletes could perform well below their personal best yet claim a gold medal. It should be no different within other domains. Hence, in terms of the performance referent “sensitivity,” it seems imperative to look beyond extrinsic referents and to also consider the assessment of performance in terms of self-referenced potential performance parameters.

While such traditional nomothetic methods as the use of Z-score-type analysis tell part of the performance story—and probably tell it well for “run of the mill” judges—a tournament of judges, in which such methods are to be exclusively adopted as *the* tool to identify the shining and not-so-shining stars, seems to require a set of tools not to be found in the typical bag of tricks of the nomothetic research analyst. One way of measuring performance within a tournament of judges that attempts to tackle these issues may be to adopt the “idiothetic” approach developed by Lamiell.<sup>125</sup> This approach, developed and utilized for the most part in the psychology literature pertaining to personality, represents a blending of the idiographic and nomothetic research paradigms. Unlike either pure idiography, with its case study methods and focus on the qualitative understanding of development and performance, or pure nomothesis, with its aggregate statistical methods and focus on the group, Lamiell’s *idiothesis* provides a vehicle for the marriage of these seemingly incommensurable research positions.<sup>126</sup> Lamiell has presented a statistical model for the study of individual differences in performance that shifts the study referent from some measure of central tendency to the self.<sup>127</sup> In simple terms, the model assesses performance not in terms of what others do, but in terms of what a person *tends not to do but*

---

125. James T. Lamiell, *Toward an Idiothetic Psychology of Personality*, 36 AM. PSYCHOLOGIST 276 (1981) [hereinafter Lamiell, *Toward an Idiothetic Psychology*]; see also James T. Lamiell & Steven J. Trierweiler, *Personality Measurement and Intuitive Personality Judgments from an Idiothetic Point of View*, 6 CLINICAL PSYCHOL. REV. 471 (1986).

126. See James W. Grice, *Bridging the Idiographic-Nomothetic Divide in Ratings of Self and Others on the Big Five*, 72 J. PERSONALITY 203, 204-05 (2004); Albert Silverstein, *An Aristotelian Resolution of the Idiographic Versus Nomothetic Tension*, 43 AM. PSYCHOLOGIST 425 (1988).

127. See Lamiell, *Toward an Idiothetic Psychology*, *supra* note 125, at 281-86.

*could do*.<sup>128</sup> In presenting this model, two main advances are made: first, the individual differences “error” variation that is suppressed in traditional inferential statistical models becomes friend, and not foe; second, the positivist scientific rigor that is “compromised” using idiography is retained.<sup>129</sup> In terms of a tool for assessing performance within a tournament of judges, this approach has several advantages. First, it can be applied to populations as small as  $N = 1$ , and hence the naturally occurring small population of judges is no longer a methodological stumbling block. Second, it does not place unrealistic performance expectations or impose unfair value judgments upon those judges who will never be highfliers. Third, it does not allow the highfliers to stop too often to smell the roses, live off their reputation, or do any of the things that past successes might tempt us all to do every once in a while.

The challenge, of course, is how we determine what one *could* do; or how we construct an index to estimate the bounds of individual performance expectations. This is an area that has received scant scholarly attention and, to my knowledge, none within the area of judicial performance.<sup>130</sup> Certainly the magic answer is not to be found in this Essay; but my purpose will be served if it stirs the reader in any way to consider the possibilities that such a model may present for better measurement of judicial performance. Any index of what one *could* do needs necessarily to account for both individual differences and accepted norms of what the peer group can do and has done. Yes, there is certainly a place for extrinsically referenced norms within idiotesis. After all, one only gets to be a judge because many years before one scored high enough relative to one’s peers to attend law school, and so on. Thus current measures of judicial performance discussed earlier in this Essay probably have a place in the model, as does past performance, since generally there is a high positive correlation between one’s past and future performance.<sup>131</sup> But, these should be assessed relative to individual differences in the cognitive, motivational, and social peculiarities of each judge that affect day-to-day functioning, as well as relative to peer norms. Again, while research is scant in terms of applications to judges, there is a

---

128. *Id.* at 281.

129. *See id.* at 281-86.

130. For an application of the idiotetic approach to vocational choices, see Terence J.G. Tracey & Maria Darcy, *An Idiotesic Examination of Vocational Interests and Their Relation to Career Decidedness*, 49 J. COUNSELING PSYCHOL. 420 (2002). For an application of the idiotetic approach to exploring the competence of counselors in multicultural settings, see Maria Darcy et al., *Complementary Approaches to Individual Differences Using Paired Comparisons and Multidimensional Scaling: Applications to Multicultural Counseling Competence*, 51 J. COUNSELING PSYCHOL. 139 (2004).

131. *See* Icek Ajzen, *Residual Effects of Past on Later Behavior: Habituation and Reasoned Action Perspectives*, 6 PERSONALITY & SOC. PSYCHOL. REV. 107, 107-08 (2002) (stating that past behavior is a good predictor of future behavior).

large body of literature that demonstrates that occupational performance is affected by such things as individual differences in motivation and personality,<sup>132</sup> as well as external life events.<sup>133</sup> Each of these contextual factors will bear upon what a judge *could* do. And each of them will be different for different judges.

Finally, if one were able to construct such an index, how would one use it as a decisionmaking tool within a tournament of judges? In terms of identifying those who appear to be underperforming, the index would be useful in determining whether intervention would be likely to address that underperformance. If underperformance is identified relative to a self-referenced performance bound, then it may be that manipulation of a contextual factor—say, motivation—would improve performance. Once improved, this judge may be an ideal candidate for promotion. But if underperformance is identified relative to a group-referenced performance bound, and if contextual factors cannot explain the underperformance, then it simply may be that current performance levels are being maximized. This person is clearly not one bound for promotion. The high flier who excels relative to both the group and self-referents may not be the ideal candidate for promotion in the long term. While there will always be those who will excel normatively without maximum effort, they are few and far between. It may well be that the judge who performs consistently at a self-referenced best will be the one whose career is cut short by occupational burnout or one of the many stress-related diseases well demonstrated to be associated with the Type A “workaholic” personality.<sup>134</sup> Maybe the judge most deserved of promotional attention is the one who is a good performer relative to the group, but slightly less so in relation to the self. Clearly this judge has the intellectual wherewithal necessary for promotion. The key to understanding this person is to understand what is holding him or her

132. See, e.g., Robert P. Tett & Dawn D. Burnett, *A Personality Trait-Based Interactionist Model of Job Performance*, 88 J. APPLIED PSYCHOL. 500 (2003).

133. For a widely used measure of life events, see Thomas H. Holmes & Richard H. Rahe, *The Social Readjustment Rating Scale*, 11 J. PSYCHOSOMATIC RES. 213 (1967). For measures of motivation, see Inge Houkes et al., *Specific Determinants of Intrinsic Work Motivation, Emotional Exhaustion and Turnover Intention: A Multisample Longitudinal Study*, 76 J. OCCUPATIONAL & ORGANIZATIONAL PSYCHOL. 427 (2003); Inge Houkes et al., *Work and Individual Determinants of Intrinsic Work Motivation, Emotional Exhaustion, and Turnover Intention: A Multi-Sample Analysis*, 8 INT'L J. STRESS MGMT. 257 (2001); and Nico W. van Yperen & Mariët Hagedoorn, *Do High Job Demands Increase Intrinsic Motivation or Fatigue or Both? The Role of Job Control and Job Social Support*, 46 ACAD. MGMT. J. 339 (2003).

134. See Bruce D. Kirkcaldy et al., *The Influence of Type A Behaviour and Locus of Control upon Job Satisfaction and Occupational Health*, 33 PERSONALITY & INDIVIDUAL DIFFERENCES 1361, 1367-70 (2002). For a Singapore study of the causes of work stress among lawyers and five other professional groups, with a particular focus on the effects of personality on work stress, see Kwok Bun Chan et al., *Work Stress Among Six Professional Groups: The Singapore Experience*, 50 SOC. SCI. & MED. 1415 (2000).

back. Perhaps he or she is bored and a new and more fulfilling role will spark his or her occupational enthusiasm. Perhaps he or she is experiencing non-work-related problems—the impact of which may well dissipate with intervention. There are many “perhapses,” and the discussion of the many possibilities is beyond the scope of this Essay. The challenge is to concentrate research efforts on identifying the most salient “perhapses” for judges. Once we have a better understanding of these factors, our task in constructing the most parsimonious idiothetic performance model will be all the easier.

#### VI. INCENTIVE EFFECTS OF MEASURING AND RANKING JUDICIAL PERFORMANCE

Choi and Gulati emphasise the positive incentive effects of their proposed tournament:

We . . . believe that a judicial tournament would provide appellate judges with the otherwise absent external incentive to exert greater effort than they currently do. If high effort as a circuit judge (for example, publishing more opinions or hearing oral argument in more cases) is a criterion for promotion, rewarding effort with a higher ranking would induce circuit judges to work harder at their jobs.<sup>135</sup>

If, in fact, judges do behave as *homo economicus*, it is reasonable to expect that a tournament with the elements proposed by Choi and Gulati should generate positive incentive effects, at least for appellate judges whose opinions are read and cited. At the very least, a high objective ranking in such an exercise brings professional respect and increased visibility, which in turn contribute to the prospects of promotion. Choi and Gulati make the good point that while judges sitting on the same court probably already have a good sense of each other’s abilities, objective rankings can have positive incentive effects even on mid-level judges with little chance of promotion, because they can use the rankings to market themselves to those outside the court, such as potential law clerks.<sup>136</sup>

There is, however, a problem. Choi and Gulati’s application of their tournament<sup>137</sup> produces such interesting results, which are useful in providing objective criteria in relation to who would make a good Supreme Court Justice, precisely because the rules of the tournament were not preannounced. If the President were to announce that from now on Supreme Court Justices would be appointed based

---

135. Choi & Gulati, *supra* note 1, at 304.

136. *Id.* at 314; see also Patricia M. Wald, *The Rhetoric of Results and the Results of Rhetoric: Judicial Writings*, 62 U. CHI. L. REV. 1371, 1372 (1995) (arguing that one reason judges publish their opinions is to attract the best law clerks).

137. Choi & Gulati, *supra* note 49.

on the outcome of a Choi and Gulati-type tournament of judges, because of the very fact that judges do behave as *homo economicus*, it would have both positive and negative effects on judicial behavior. Williamson describes the new *homo economicus* as someone who engages in “self-interest seeking with guile.”<sup>138</sup> This brings us back to the point made in Part II that *homo economicus* will behave in an opportunistic manner to achieve his self-interested objectives.

One potential adverse implication concerns the negative effect of increased competition. Choi and Gulati suggest that one of the positive spin-offs of their tournament is that it will provide incentives to compete and that this will result in higher levels of effort.<sup>139</sup> In the spirit of the Alchian and Demsetz model of team production,<sup>140</sup> the focus here is clearly on reducing shirking and improving allocative efficiency on the court. This, however, might come at the cost of reduced productive efficiency if a more competitive environment has a negative effect on the level of collaboration and potential for collective learning on the court. Chen shows that participants in tournaments have incentives to sabotage others in a bid to improve their own relative position.<sup>141</sup> Choi and Gulati seek to downplay this possibility, suggesting that judges are unlikely to undermine their colleagues.<sup>142</sup> But, if one accepts that judges behave as *homo economicus*, there is no reason to think that judges would behave differently from others interacting in small social groups. There is plenty of anecdotal evidence to suggest that courts, like other social organizations, have been affected by personality differences, which a tournament that promoted competition may exacerbate. The High Court of Australia for much of its history has been impaired by personal tensions and petty jealousies between the judges that have adversely affected its work.<sup>143</sup> One (albeit extreme) example was Isaacs, who was a Justice of the Court from 1906-1931 and Chief Justice from 1930-1931.<sup>144</sup> His biographer notes that he would hide cases and play down the significance of issues in argument so as to give himself a head start relative to the other judges when it came to writing opinions.<sup>145</sup>

Making the number of published opinions the criteria for measuring good performance can also have negative side effects. If the focus

138. See Oliver E. Williamson, *The Economics of Governance: Framework and Implications*, 140 J. INSTITUTIONAL & THEORETICAL ECON. 195, 198 (1984).

139. Choi & Gulati, *supra* note 1, at 313-15.

140. Alchian & Demsetz, *supra* note 58, at 779-81.

141. Kong-Pin Chen, *Sabotage in Promotion Tournaments*, 19 J.L. ECON. & ORG. 119, 121-34 (2003).

142. Choi & Gulati, *supra* note 1, at 309.

143. Amelia Simpson & Troy Simpson, *Personal Relations*, in THE OXFORD COMPANION TO THE HIGH COURT OF AUSTRALIA 528, 528-31 (Tony Blackshield et al. eds., 2001).

144. ZELMAN COWEN, ISAAC ISAACS 113 (1967)

145. *Id.* at 124-25.

is on counting opinion numbers, there is always the potential, in the United States setting, that it will promote a publish-or-perish mentality, where the number of published opinions increases but the overall quality of those opinions decreases. One might argue that a judge will not submit “substandard” opinions for publication, because publishing poor-quality opinions would hurt his or her reputation. But this assumes a certain degree of farsightedness, because the reputation effect is likely to occur only with a lag, while in practice judges confronted with the prospect of improving their finishing position in the next tournament may well be myopic. Even if the suggestion made earlier of having multitiered law reports containing opinions of different quality was considered fanciful,<sup>146</sup> a more realistic solution to this problem would be to set the initial hurdle for publication higher.

A related problem with focusing on opinion counts is the potential adverse incentives this has for managing caseloads. Even in higher courts, where the opinions of judges are read, opinions published might not be the best measure of productivity if time spent on crafting elaborate opinions is at the expense of case management. One of the main reasons for criticism of longer opinions in the High Court of Australia since the 1990s has been the view that the time taken to write longer opinions affects the capacity of judges to deal with their overall caseloads within a reasonable period of time.<sup>147</sup> The argument is that longer, multiple opinions which contain excessive citation are responsible for an increase in the average time elapsed “between the conclusion of [the] hearing and pronouncement of a judgment which has been reserved.”<sup>148</sup> This is a sensitive issue in Australia, as in the United States, where increasing backlogs have intensified community pressure for speedier justice.<sup>149</sup> In 1998, court delays came under increased media scrutiny in Australia, when New South Wales Supreme Court Justice Vince Bruce faced a State Parliament vote to remove him from office for taking up to three years to deliver a judgment.<sup>150</sup> This has forced courts to monitor delays. In an attempt to address this issue, judges of the New South Wales Supreme Court are now required to make monthly progress reports on delivery of their judgments and time limits have been placed on the time period

---

146. See *supra* notes 70-71 and accompanying text.

147. John Doyle, *Judgment Writing: Are There Needs for Change?*, 73 AUSTRALIAN L.J. 737, 738 (1999).

148. Campbell, *supra* note 68, at 63; see also Doyle, *supra* note 147, at 738.

149. For an Australian perspective on the issue, see Michael Kirby, *On the Writing of Judgments*, 64 AUSTRALIAN L.J. 691 (1990). On caseload pressures in the United States, see RICHARD A. POSNER, *THE FEDERAL COURTS: CHALLENGE AND REFORM* 53-86 (1996).

150. Damien Murphy, *Long Wait for Day in Court to Be Shortened*, SYDNEY MORNING HERALD, Oct. 19, 1998, at 5.

between committal hearings and the trial in criminal cases.<sup>151</sup> Similar procedures operate in the United States. For example, on the United States Court of Appeals for the District of Columbia Circuit, a judge who has three or more assigned opinions pending that are not in circulation is not allowed to sit on any new cases until the backlog is cleared.<sup>152</sup>

One might argue that the approach to avoid the problem of judges publishing large numbers of poor-quality decisions is to put more emphasis on citation counts. The reasoning is that poor-quality decisions will not get cited, so the marginal return from expending effort on publishing poor-quality decisions will be low. But if the focus is solely on citation counts, this also presents problems. Kobayashi and Lott suggest that judges who want to maximize citation counts to their own decisions will have an incentive to replace existing precedents with their own inefficient decisions, which will generate more litigation and hence more opportunities to be cited.<sup>153</sup> While Posner argues that most judges would not behave in this fashion,<sup>154</sup> it is reasonable to expect that judges confronted with a tournament that explicitly rewards citation counts and “self-interest seeking with guile” would implement strategies to maximize their citation count. One possible strategy is the formation of citation clubs, where members cite each other’s opinions to the exclusion of others on the court. To some extent this exists already. Studies for the United States courts of appeals<sup>155</sup> and Federal Court of Australia<sup>156</sup> have found that judges who attended elite law schools receive more citations. This might be because the judges from elite schools are more able, or it might be because like-minded judges cite each other. Landes, Lessig, and Solimine prefer the latter explanation, suggesting that on the United States courts of appeals, graduates from Harvard and Yale form “part of [a] large network of ‘like-thinking’ judges who will tend to cite each other more often than they cite other judges.”<sup>157</sup> Choi and Gulati, while acknowledging that judges might cite similar-minded colleagues who agree to reciprocate, argue that this might in fact hurt a judge’s prospect of winning the tournament, because the judge will not be citing the best opinions available; therefore, the quality of the output will fall.<sup>158</sup> But, this is only true if the probability of being

151. *Id.*

152. Edwards, *supra* note 14, at 1665.

153. Posner, *supra* note 12, at 6 n.9 (citing Bruce H. Kobayashi & John R. Lott, Jr., *Judicial Reputation and the Efficiency of the Common Law* (1993) (unpublished manuscript)).

154. *See id.*

155. Landes et al., *supra* note 73, at 324.

156. Smyth & Bhattacharya, *Aging and Productivity and the Federal Court*, *supra* note 75, at 156-58.

157. Landes et al., *supra* note 73, at 324.

158. Choi & Gulati, *supra* note 1, at 311 & n.31.

cited is purely a function of the quality of the opinion, which need not be the case in the presence of citation clubs.

There are parallels here with the “insider”-“outsider” distinction in legal scholarship; that is, the argument that while “outsiders”—critical legal theorists, feminists, and critical race theorists—are producing good scholarship, it is not being cited by “insiders.”<sup>159</sup> Levit suggests that, as a result, “[r]eliance on quantitative assessments of legal scholarship may tend to subtly perpetuate existing hierarchies of race, gender, and theory prominence, while telling little about the substantive or foundational qualities of a theory.”<sup>160</sup> There is no evidence of gender or racial discrimination in citation counts on the courts,<sup>161</sup> but the point about theory building is arguably true for citations to judicial opinions, with “outsiders” holding nonmainstream views or employing unorthodox opinion-writing techniques being overlooked. An illustration of an “outsider” on the High Court of Australia was Lionel Murphy, who was a Justice from 1975-1986. Murphy largely ignored the doctrine of precedent, arguing from first principles.<sup>162</sup> At the time he was often in dissent, but many of his ideas have since gained acceptance in the Court’s jurisprudence, albeit in more complex forms.<sup>163</sup> Murphy, however, is rarely cited as being the originator of those ideas.<sup>164</sup>

## VII. CONCLUSION

The concept of a tournament of judges is a useful one. In the United States, where the judicial appointment process has become so politicized, and even in the Antipodes, where judicial appointments are not as politicized but still nontransparent, the use of objective indicators of performance would make a positive contribution to the selection process. However, as Choi and Gulati freely admit, a tournament of judges is not the be-all and end-all when it comes to deciding who are the best performers.<sup>165</sup> A potential problem with explicit reliance on a tournament, which this Essay has emphasized, is that it

---

159. See, e.g., Frances Olsen, *Affirmative Action: Necessary but Not Sufficient*, 71 CHI.-KENT L. REV. 937 (1996).

160. Nancy Levit, *Defining Cutting Edge Scholarship: Feminism and Criteria of Rationality*, 71 CHI.-KENT L. REV. 947, 949 (1996).

161. See, e.g., Landes et al., *supra* note 73, at 324 (finding no evidence of racial or sexual discrimination in citation practices on the United States courts of appeals).

162. Michael Kirby, *Lionel Murphy and the Power of Ideas*, 18 ALTERNATIVE L.J. 253, 254-57 (1993).

163. *Id.* at 254-58.

164. See JUSTICE LIONEL MURPHY: INFLUENTIAL OR MERELY PRESCIENT? *passim* (Michael Coper & George Williams eds., 1997); Kirby, *supra* note 162, at 254-55.

165. Choi & Gulati, *supra* note 1, at 322.

can result in game playing.<sup>166</sup> But, as Choi and Gulati state, “Our primary goal is not to produce winners, but to enable transparency in the nomination process.”<sup>167</sup> Given that each of the quantitative indicators have positive and negative points, it is important to use a range of quantitative indicators, as Choi and Gulati do in their “composite measures,”<sup>168</sup> in order to obviate the limitations of any one measure.

While the tools exist to evaluate qualitative dimensions of judicial performance, research in this area is less advanced due, at least in part, to more restricted data availability relative to counting citations and opinions. One approach discussed in this Essay to tackle the qualitative dimension of judging, which has not been utilized to this point, is to borrow idiopathic methods from psychology. It is an important step forward, both conceptually and methodologically, to further refine idiopathic methods that may enable us to move beyond the typically accepted tradition of “extrinsic normative comparison” that persists in performance evaluation. Such comparison, which lends itself to conclusions based upon ranking procedures, seems myopic with respect to identification of negative change among high-level performers and positive change among low-level performers as well as over acceptance of the status quo. In the end, subject to these reservations, the most appropriate role for the tournament proposed by Choi and Gulati might be to help generate a pool of qualified candidates, from which candidates for promotion can be selected. Finally, this Essay has discussed the tournament in the specific context of the Australian and New Zealand legal setting, where several institutional arrangements differ from the United States. These different institutional arrangements do not mean that a tournament of judges has no relevance to the Antipodes. On the contrary, I have argued quite the opposite, but what they do highlight is that the design of a tournament, in whichever court it is played, needs to be tailored to meet the specific institutional context in which the court operates.

---

166. For further discussion of the game-playing issue in the context of designing optimal incentives, see George Baker et al., *Subjective Performance Measures in Optimal Incentive Contracts*, 109 Q.J. ECON. 1125 (1994).

167. Choi & Gulati, *supra* note 49, at 81.

168. *Id.* at 70-75.