

GHOSTS FROM THE HIGH COURT'S PAST: EVIDENCE FROM COMPUTATIONAL LINGUISTICS FOR DIXON GHOSTING FOR MCTIERNAN AND RICH

YANIR SEROUSSI,* RUSSELL SMYTH** AND INGRID ZUKERMAN***▲

I INTRODUCTION

Sir Edward McTiernan was a member of the High Court of Australia (hereafter High Court) from 1930 to 1976, making him the Court's longest serving Justice. Sir George Rich was a member of the High Court from 1913 to 1950. The High Court terms of both McTiernan and Rich overlapped to a considerable extent with that of Sir Owen Dixon, who was a Justice of the High Court from 1929 to 1952 and Chief Justice from 1952 to 1964. It is well known that McTiernan and Rich had a tendency to join in the judgments written by Dixon, or to write a short concurring judgment agreeing with him.¹ Sir Hayden Starke considered McTiernan a 'parrot' because he often agreed with Dixon.² Legg writes: 'Sir Edward McTiernan was not regarded with as much awe as his brethren on the Bench. On one occasion, much to everyone's amazement, Sir Owen Dixon concurred in one of his judgments. About 99 per cent of the time it was the other way around'.³

Similarly, with respect to Rich, Merralls states: 'With Sir Owen Dixon, Rich found the perfect exemplar from whom he seldom differed'.⁴ Ash states, Rich 'decided matters on the facts before him and explained, through himself *or*

* PhD. candidate, Faculty of Information Technology, Monash University.

** Professor, Department of Economics, Monash University.

*** Professor, Faculty of Information Technology, Monash University.

▲ We thank Jessica O'Leary for research assistance on this project, Minh Duc Cao for discussions throughout the project and Matthew Groves, Wendy Zukerman and the referees for insightful comments on the paper. The original text files, together with tokenised text, sufficient to reproduce the results reported in this paper, as well as a full breakdown of the results are available at <<http://www.csse.monash.edu.au/research/umnl/data>>.

1 See Matthew Groves and Russell Smyth, 'A Century of Judicial Style: Changing Patterns in Judgment Writing on the High Court 1903–2001' (2004) 32 *Federal Law Review* 255.

2 See Clem Lloyd, 'Not Peace but a Sword! – The High Court under J G Latham' (1987) 11 *Adelaide Law Review* 175, 181.

3 David Ash, 'Sir George Rich' [2010] (Winter) *Bar News (Journal of the New South Wales Bar Association)* 46, 68 (emphasis added).

4 J D Merralls, 'Rich, Sir George Edward (1863–1956)' in *Australian Dictionary of Biography* (Melbourne University Press, 1988) vol 11, 371, 372.

others, how he got there'.⁵ Ash is fairly sympathetic to Rich. As far as the tendency for Rich to agree with Dixon, Ash asks: 'And would the nation really be served by one Bench full of Dixons? Even Dixon would demur'.⁶ However, others are less sympathetic to revelations that Rich rarely contributed much to joint judgments to which he added his name. For example, Hughes refers to a meeting he once had with Rich, stating, 'in the course of conversation [Rich] referred to a case in which, as he put it, "I and Evatt" had written a judgment – an assertion of active authorship which subsequent historical revelations may call for the application of a grain of salt'.⁷

The proclivity of McTiernan and Rich to concur with the judgments of Dixon has earned them a reputation for indolence. Cope writes of 'Rich's lack of energy and propensity to attach his name to the judgments of others'.⁸ In his autobiography, Sir Robert Menzies states of Rich,

truth requires me to say that he was inclined to be indolent. ... He certainly wrote a few individual judgments that were a joy to read; but on the whole he preferred to attach his name to a joint judgment, the labour of writing which he left to his judicial partners'.⁹

Similarly, Merralls states: 'Rich's standing as a judge suffered from his reputation for indolence'.¹⁰ Dixon himself considered both McTiernan and Rich indolent. Ayres states that Dixon thought McTiernan 'lazy and unqualified'.¹¹ At Rich's ninetieth birthday party, Rich made a speech in which he argued 'Australians should work'. Dixon wrote to his daughter, Anne: 'The idea of Sir George preaching the doctrine of work struck Mum as particularly amusing'.¹² It seems that Sir Frank Gavan Duffy, to whose judgments Rich had frequently attached his name until Dixon joined the Court, considered Rich idle. Sir Robert Menzies records Rich exclaiming: 'Duffy, the problem with you is that you talk too much from the Bench' and of Gavan Duffy replying: 'Small wonder since I have to talk for two'.¹³

Sir Owen Dixon kept a diary for 1911, the first two months of 1929 and for each year between 1935 and 1965. In the Dixon Diaries, Dixon records that not only did McTiernan and Rich regularly attach their names to Dixon's judgments, but that Dixon actually wrote sole-authored judgments which have been attributed to McTiernan and Rich.¹⁴ Ayres draws the link between Rich's propensity to concur with Dixon and having Dixon write his judgments:

5 Ash, above n 3, 68 (emphasis added).

6 Ibid.

7 T E F Hughes, 'High Court Recollections' (2003) 77 *Australian Law Journal* 661, 662.

8 Malcom Cope, 'Owen Dixon' (2005) 8 *Canberra Law Review* 125, 126.

9 Robert G Menzies, *The Measure of the Years* (Cassell Australia, 1970) 265.

10 Merralls, 'Rich, Sir George Edwards', above n 4.

11 See Philip Ayres, *Owen Dixon* (Miegunyah Press, 2003) 59.

12 Ibid 240.

13 Menzies, above n 9, 266.

14 See Philip Ayres, 'Dixon Diaries' in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 222.

The urbane George Rich ... had ability but lacked energy, tending to attach his name to the judgments of others. He would find Dixon so persuasive, he would take the logical course and often ask Dixon to write judgments for him.¹⁵

In the Dixon Diaries, Dixon makes reference to writing Rich's judgments in several cases, including cases in which Dixon had not sat.¹⁶ Dixon also records helping Rich compose a non-dissenting judgment in a case where Dixon himself dissented, 'like a debater momentarily changing sides'.¹⁷ In another case, Dixon records writing Rich's judgment at first instance, then sitting as a member of the Court that heard the appeal from Rich's/Dixon's judgment.¹⁸ In the Dixon Diaries, Dixon also refers to helping McTiernan compose his judgments, although this appears to have occurred less frequently.¹⁹

There has been widespread reaction to these revelations in the Dixon Diaries, particularly in the aftermath of the publication of Philip Ayres' biography of Dixon, in which they feature prominently.²⁰ Revelations that Dixon ghost-wrote judgments for McTiernan and Rich have been variously described as 'very revealing'²¹ and 'shocking'.²² However, while we know from the Dixon Diaries that Dixon ghost-wrote judgments for Rich and, to a lesser extent, McTiernan, there is no concrete evidence about the extent of this practice. Thus, the actual extent to which this occurred has been a matter of anecdote and rumour.

The purpose of this paper is to use methods from computational linguistics to examine the extent to which Dixon wrote McTiernan's and Rich's judgments. Our main, somewhat provocative, finding is that Dixon wrote about 18 per cent of Rich's judgments and about four per cent of McTiernan's judgments for the periods Dixon was on the High Court with Rich and McTiernan respectively. Our confidence in these attributions is more than 99 per cent for judgments written by

15 Ayres, *Owen Dixon*, above n 11, 56.

16 According to Ayres, *Owen Dixon*, above n 11, 320 there are several examples in Dixon Diaries 'of Dixon writing a judgment for Rich in a case in which Dixon had not sat'. For instance, Ayres records that Dixon wrote Rich's judgment in *Federated State School Teachers' Association of Australia v Victoria* (1929) 41 CLR 569, 590: at 57–8, 317.

17 Ayres, *Owen Dixon*, above n 11, 73. The case was *R v Brislan; Ex parte Williams* (1935) 54 CLR 262.

18 Ayres, *Owen Dixon*, above n 11, 93–4. The case was *Sun Newspapers Ltd and Associated Newspapers Ltd v Federal Commissioner of Taxation* (1938) 61 CLR 337.

19 Ayres, *Owen Dixon*, above n 11, 68, 320 reports that in the Dixon Diaries for 10 July 1935 Dixon records: 'McTiernan dined with me and then we composed his judgment in Southport case'. According to the Dixon Diaries, Dixon wrote McTiernan's addendum to the majority judgment, also authored by Dixon, in *Hughes and Vale Pty Ltd v New South Wales (No. 2)* (1955) 93 CLR 127. In the addendum, McTiernan/Dixon expressed lingering discontent with the Dixon view of s 92 of the *Constitution*! See Michael Coper, 'Concurring Judgments' in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 129, 130; Justice Michael Kirby, 'McTiernan, Edward Aloysius' in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 466, 468.

20 Ayres, *Owen Dixon*, above n 11.

21 Justice P W Young, 'Melbourne Gentleman of the 30s: Owen Dixon Book Review' (2003) 77 *Australian Law Journal* 685, 686.

22 Sean Huber, 'Sir Owen Dixon: The Life and Times of Australia's Finest Jurist' (2004) 23 *University of Queensland Law Journal* 277, 278.

Dixon for McTiernan and more than 97 per cent for judgments written by Dixon for Rich.

II WHY DOES IT MATTER IF DIXON WROTE JUDGMENTS FOR MCTIERNAN AND RICH?

While these instances of ghost-writing occurred a long time ago, the extent to which they occurred is more than an historical curiosity for several reasons. One reason is that attribution matters. As Fisk puts it: 'The reputation we develop for the work we do proves to the world the nature of our human capital. Credit is instrumentally beneficial in establishing a reputation and intrinsically valuable simply for the pleasure of being acknowledged'.²³ The judgments of justices of the High Court establish binding precedents of legal authority that determine the evolution of the common law in Australia. These judgments provide guidance to future judges, legal practitioners or other repeat players toward the proper resolution of like cases.²⁴ Landes and his colleagues argue that prestigious judges, such as Dixon, develop a brand name or trademark that signifies quality, and that such a brand name reduces the search costs for users associated with finding high quality judgments.²⁵ Such an argument rests on the assumption that judges build reputation capital through their judgments which appear in the law reports. If judgments are not being properly attributed, this assumption will no longer hold. More generally, when judgments are cited by judges in future courts and other users of law reports, it is implicitly assumed that they are being properly attributed. If some judgments are in fact being ghosted by other judges, this will not be the case.

Recognition that attribution is important raises the more general issue of breaches of judicial propriety on the Court. If a first year law student were to pass off an in-course assignment as his or her own and the lecturer were to find that it was in fact written by someone else, the penalties for misrepresentation would be severe.²⁶ It is reasonable to expect that judgments written by High Court justices, who are at the pinnacle of the legal profession in Australia, should be held to the same standards as beginning law students. Dixon is widely regarded as one of the greatest common law lawyers ever.²⁷ He has achieved almost deity status in the

23 Catherine Fisk, 'Credit Where It's Due: The Law and Norms of Attribution' (2006) 95 *Georgetown Law Journal* 49, 50.

24 See Justice Michael Kirby, 'On the Writing of Judgments' (1990) 64 *Australian Law Journal* 691, 692.

25 William M Landes, Lawrence Lessig and Michael E Solimine, 'Judicial Influence: A Citation Analysis of Federal Courts of Appeal Judges' (1998) 27 *Journal of Legal Studies* 271, 272.

26 See Stephen J Choi and G Mitu Gulati, 'Which Judges Write Their Opinions (and Should We Care)?' (2005) 32 *Florida State University Law Review* 1077, 1077.

27 See J D Merralls, 'The Rt Hon Sir Owen Dixon, OM, GCMG, 1886–1972' (1972) 46 *Australian Law Journal* 429; Justice Kenneth Hayne, 'Dixon, Owen' in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 218; Sir Ninian Stephen, *Sir Owen Dixon: A Celebration* (Melbourne University Press, 1986).

Australian legal profession, being likened to Bradman²⁸ and Mozart.²⁹ Hulme quotes Dame Pattie Menzies as saying to her husband: ‘Robert, you must remember that Owen Dixon is not God’. He replied: ‘No, my dear, but only just’.³⁰ The period when Dixon was Chief Justice is regarded as the Court’s ‘Golden Age’ of jurisprudence.³¹ However, it is arguable that there has been a tendency to mythologise Dixon.³² Dixon was a man, not a god, and he had human failings like the rest of us. Laurence Maher and others have given examples of Dixon breaching judicial propriety, including repeated breaches of constitutional convention.³³

Dixon held other judges to high standards. For example, when it emerged that Dixon’s friend, ‘Sammy’ Clyne, a judge of the Federal Court of Bankruptcy, had written Rich’s judgment in *Isaacs v McKinnon*³⁴ in an appeal against a judgment of Clyne himself, Dixon considered it scandalous and expressed disapproval to Latham CJ in the strongest of terms.³⁵ As Ash notes, ‘this hardly sits well with Dixon’s earlier – and presumably continuing – practice’ at the time this occurred of ghosting for Rich.³⁶ Leaser describes Dixon ghosting for Rich in stronger terms, stating:

For someone who was always complaining about the ethics of other justices, this sort of conduct is evidence of at least a modicum of hypocrisy. Perhaps his physical pain and the volume of extra work [from ghosting] explain his dislike of judicial work, regular contemplations of resignation and discouragement of others from judicial office.³⁷

Additionally, proper attribution matters for empirical scholarship of the law. In the United States, Choi and Gulati have introduced the notion of a tournament of judges.³⁸ Choi and Gulati argue that the selection of justices for the United States Supreme Court ought to be based on a tournament, where judges who possess the most merit as measured empirically, would be selected over their

28 S E K Hulme, ‘Recollections Mainly to Do with the Dixon Court’ (2003) 77 *Australian Law Journal* 653, 654.

29 Justice Peter Heerey, ‘A Towering Figure in the Law of This Country: Owen Dixon – Book Review’ (2003) 77 *Australian Law Journal* 682, 684.

30 Hulme, above n 28, 655.

31 Leslie Zines, ‘Dixon Court’ in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 220, 220.

32 See David Ritter, ‘The Myth of Sir Owen Dixon’ (2005) 9 *Australian Journal of Legal History* 249.

33 See Laurence Maher, ‘Tales of the Overt and Covert, Judges and Politics in Early Cold War Australia’ (1993) 21 *Federal Law Review* 151, 168–75; Geoff Lindsay, ‘Owen Dixon: A Biography’ (2003) 23 *Australian Bar Review* 198; David Ritter, ‘Greatness as Measure? Recent Writings on the High Court of Australia’ (2004) 50 *Australian Journal of Politics and History* 434; Dennis Rose, ‘Sir Owen Dixon’ (2003) 6 *Constitutional Law and Policy Review* 18.

34 (1949) 80 CLR 502.

35 Ayres, *Owen Dixon*, above n 11, 191.

36 Ash, above n 3, 63.

37 Julian Leaser, ‘Book Review: Owen Dixon’ (2003) 26 *University of New South Wales Law Journal* 335, 340.

38 Stephen Choi and Mitu Gulati, ‘A Tournament of Judges?’ (2004) 92 *California Law Review* 299.

lower ranked peers.³⁹ The notion of a tournament of judges is taken very seriously in the United States, where it has been well-debated among scholars of empirical legal studies⁴⁰ as well as in the popular media.⁴¹ While there has not been a similar debate in Australia,⁴² citation analysis has been used to rank judges of the High Court, based on their influence over the development of the common law.⁴³ This analysis suggested that, excluding self-citations and adjusting for length of time on the Court and depreciation of legal precedent, Dixon is the most cited judge in the history of the High Court, while McTiernan and Rich are toward the bottom of the rankings.⁴⁴ Other studies for the High Court have used judicial output in the form of judgments from the period Dixon, McTiernan and Rich were on the Court to examine issues such as judicial productivity over the period of a judge's appointment⁴⁵ and factors that explain the decision to retire from the High Court.⁴⁶ The integrity of such findings depends on the proper attribution of judgments. If Dixon were found to have written substantial numbers of Rich's and/or McTiernan's judgments, empirical studies like these would turn out to be based on distorted data.

Other empirical studies have analysed voting patterns on the High Court, including those for the period when Latham was Chief Justice and Dixon,

-
- 39 Stephen Choi and Mitu Gulati, 'Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance' (2004) 78 *Southern California Law Review* 23; Stephen J Choi and G Mitu Gulati, 'Mr. Justice Posner? Unpacking the Statistics' (2005) 61 *New York University Annual Survey of American Law* 19.
- 40 For example, there is an entire special issue of the *Florida State University Law Review*, (2005) 32(4), titled 'Empirical Measures of Judicial Performance', devoted to debating the merits of a tournament of judges. For an overview, see Steven G Gey and Jim Rossi, 'Empirical Measures of Judicial Performance: An Introduction to the Symposium' (2005) 32 *Florida State University Law Review* 1001.
- 41 For example, in May 2009, the *New York Times* published an online piece entitled, 'Assessing Sotomayor's Influence,' which purported to reveal how Justice Sonia Sotomayor, who was not part of Choi and Gulati's tournament, would have fared in a tournament of judges: 'Assessing Sotomayor's Influence', *The New York Times* (online), 28 May 2009 <<http://www.nytimes.com/interactive/2009/05/28/us/politics/0529-judge-graphic.html>>. This multimedia presentation was linked to a print article, Jo Becker and Adam Liptak, 'Assertive Style Raises Questions on Demeanor', *The New York Times* (New York) 29 May 2009, A14. For a discussion of media treatment of Choi and Gulati's tournament in the United States see Marin K Levy, Kate Stith and Jose A Cabranes, 'The Costs of Judging Judges by the Numbers' (2010) 28 *Yale Law and Policy Review* 313.
- 42 For a discussion of the arguments for and against adopting a tournament of judges in selecting judges for higher office in Australia see Russell Smyth, 'Do Judges Behave as *Homo Economicus*, and, if so, Can We Measure Their Performance? An Antipodean Perspective on a Tournament of Judges' (2005) 32 *Florida State University Law Review* 1299.
- 43 See Russell Smyth, 'Who Gets Cited? An Empirical Study of Judicial Prestige in the High Court' (2000) 21 *University of Queensland Law Journal* 7; Mita Bhattacharya and Russell Smyth, 'The Determinants of Judicial Influence and Prestige: Some Empirical Evidence from the High Court of Australia' (2001) 30 *Journal of Legal Studies* 223.
- 44 Smyth, 'Who Gets Cited?', above n 43, 18 (table 4).
- 45 Mita Bhattacharya and Russell Smyth, 'Aging and Productivity Among Judges: Some Empirical Evidence from the High Court of Australia' (2001) 40 *Australian Economic Papers* 199.
- 46 Pushkar Maitra and Russell Smyth, 'Determinants of Retirement on the High Court of Australia' (2005) 81 *Economic Record* 193.

McTiernan and Rich were members of the Court.⁴⁷ It seems to be generally accepted that while Dixon wrote judgments for McTiernan and Rich, the actual voting decision was made on each occasion by McTiernan or Rich (and not Dixon).⁴⁸ Nevertheless, the fact that Dixon appears, to differing degrees, to have written judgments for McTiernan or Rich or, on occasion, assisted them in writing their judgments, lends credence to the concerns of Sir Hayden Starke, evident from the Latham papers, that Dixon was exercising subtle undue influence over both.⁴⁹ This raises concerns not only for empirical studies of voting patterns, but more gravely for the independence of judges and the proper administration of justice.⁵⁰ Justice Michael Kirby has suggested that ‘ultimately a judgment is written for the judge who writes it. It must have integrity and carry with its words the evidence of the manifest impartiality and intellectual honesty of the writer’.⁵¹ This leads to another reason why this study is important. Having an opinion rendered by several different people gives some indication of how widely the view thus expressed is held. If one person is ghosting for another, the point of view that emerges may in fact be less widely held. Yet another reason is that a judgment may be viewed in a different light when one knows the identity of the person who wrote it. For example, a judgment in which the defendant is known to be a communist might be viewed in a different light if its author is known to dislike communism.

Finally, the extent to which Dixon wrote judgments for McTiernan and Rich is potentially relevant for Dixon’s reputation for legalism. For Dixon, the basal assumption of legalism, espoused in ‘Concerning Judicial Method’ is ‘that the decision of the court will be “correct” or “incorrect”, “right” or “wrong” as it conforms with ascertained legal principles and applies them to a standard of reasoning which is not personal to the judges themselves’.⁵² Dixon’s importance

47 For the seminal jurimetrics studies of the High Court that preceded, and inspired, the later more conventional studies of voting patterns, see Glendon Schubert, ‘Political Ideology and the High Court’ (1968) 3 *Politics* 21; Glendon Schubert, ‘Judicial Attitudes and Policy-Making on the Dixon Court’ (1969) 7 *Osgoode Hall Law Journal* 1; Anthony Blackshield, ‘Quantitative Analysis: The High Court of Australia 1964– 1969’ (1972) 3 *Lawasia* 1; Anthony Blackshield, ‘Judges and the Court System’ in Gareth Evans (ed) *Labor and the Constitution 1972– 1975* (Heinemann, 1977); Anthony Blackshield, ‘X/Y/Z/N Scales: The High Court of Australia 1972– 1976’ in Roman Tomasic (ed) *Understanding Lawyers* (Allen & Unwin, 1978). For a study of voting patterns on the Mason Court see Russell Smyth, ‘Some Are More Equal than Others: An Empirical Study of Voting Behaviour on the Mason Court’ (1997) 6 *Canberra Law Review* 193. For studies of voting patterns on the Latham Court see Roger Douglas, ‘Judges and Policy on the Latham Court’ (1969) 4 *Politics* 20; Russell Smyth, ‘Explaining Voting Patterns on the Latham High Court 1935– 50’ (2002) 26 *Melbourne University Law Review* 88; Russell Smyth, ‘Judicial Interaction on the Latham Court: A Quantitative Study of Voting Patterns on the High Court 1935– 1950’ (2001) 47 *Australian Journal of Politics and History* 330.

48 See Ash above n 3; Leeser, above n 37.

49 See Lloyd, above n 2, 181. See also Amelia Simpson and Troy Simpson, ‘Personal Relations’ in Blackshield, Coper and Williams (eds), above n 14, 530.

50 Starke was concerned about the implications of Dixon’s influence on the other judges for judicial independence: Simpson and Simpson, above n 49, 530.

51 Kirby, above n 24, 695.

52 Sir Owen Dixon, ‘Concerning Judicial Method’ in Judge Severin Howard Zichy Woinarski (ed), *Jesting Pilate and Other Papers and Addresses* (W S Hein, 2nd ed, 1997) 152, 155.

today, apart from that which attaches to any particular judgment or area of law, is perhaps his association with 'Concerning Judicial Method' and legalism in judging.⁵³ If Dixon were writing judgments for McTiernan and Rich, it might be argued, by applying his judicial method in judgments he wrote for McTiernan and Rich, that he was surreptitiously entrenching legalism in the Court's reasoning. Dixon was potentially exercising unprecedented influence not only over the evolution of the jurisprudence of the Court, but also its approach to reasoning in the 1930s and 1940s. This influence was solidified in the 'long' 1950s when the length of Menzies' tenure as prime minister, and his relationship with Dixon, meant that the Court was packed with appointments approved of by Dixon, ensuring a commonality of judicial approach.⁵⁴

Today, legalism, in many quarters, has been bestowed the role of a 'high technique' that befits a great judge. This role extends beyond a narrow definition of judicial approach. As Ritter puts it, the reasoning is as follows: 'Dixon was the greatest judge, the greatest judge must, ergo, have had the greatest technique, while correspondingly it must have been strict adherence to legalist principles which rendered the maestro supreme'.⁵⁵ Seen in this broader context, legalism encompasses general legal virtues, such as continuity, objectivity and public confidence in the administration of justice.⁵⁶ This broader conception of Dixon's legalism underscores the fact that there is a lot of debate about what Dixon's judicial method actually represents. In this respect, Dixon's legalism has been used by both critics and supporters of the Mason and Brennan Courts,⁵⁷ and invoking Dixon's legalism is to invoke the technique of a great judge in support of one's own approach. If it were discovered that Dixon wrote large numbers of judgments for McTiernan and Rich, this may impinge on Dixon's reputation for legalism, while potentially undermining his reputation more generally as a great judge.

53 See Sir Daryl Dawson and Mark Nicholls, 'Sir Owen Dixon and Judicial Method' (1986) 15 *Melbourne University Law Review* 543; Michael Coper, 'Concern about Judicial Method' (2006) 30 *Melbourne University Law Review* 554; Justice Kenneth M Hayne, 'Concerning Judicial Method – Fifty Years On' (2006) 32 *Monash University Law Review* 223; John Gava, 'Sir Owen Dixon, Strict Legalism and *McRae v Commonwealth Disposals Commission*' (2009) 9 *Oxford University Commonwealth Law Journal* 141; John Gava, 'Dixonian Strict Legalism, *Wilson v Darling Island Stevedoring* and Contracting in the Real World' (2010) 30 *Oxford Journal of Legal Studies* 519; John Gava, 'A Study in Judging: Sir Owen Dixon and *McDonald v Dennys Lascelles*' (2009) 32 *Australian Bar Review* 77; John Gava, 'Another Study in Judging: Sir Owen Dixon and *Yerkey v Jones*' (2010) 26 *Journal of Contract Law* 248.

54 Brian Galligan, *Politics of the High Court: A Study of the Judicial Branch of Government in Australia* (University of Queensland Press, 1987) 202.

55 Ritter, above n 32, 261.

56 Sir Anthony Mason, 'Future Directions in Australian Law' (1987) 13 *Monash University Law Review* 149.

57 See Ritter, above n 32, 259–60 and references cited therein.

III EXISTING LITERATURE

There is a large body of literature employing computational linguistics to trace authorship in fields outside the law.⁵⁸ For example, computational linguistics has been used extensively to examine the authenticity of the authorship of Shakespearean texts.⁵⁹ It was used to unmask Joe Klein as the anonymous author of *Primary Colors*, the novel inspired by Bill Clinton's first presidential campaign.⁶⁰ Computational linguistics has been employed in a wide range of criminal investigations,⁶¹ including high profile murder investigations in the United States, such as the Jon-Benet Ramsey murder investigation and the Unabomber prosecution.⁶² Another legal application is its use to verify confessions in criminal trials.⁶³

Computational methods are starting to be regarded as useful tools for analysing different aspects of judicial opinions in the United States.⁶⁴ Specifically, there have been a couple of attempts to use computational linguistics to ascertain authorship of judgments in the United States, where many judges delegate opinion-writing tasks to their clerks.⁶⁵ These studies seek to examine which judges write their own opinions and which judges delegate opinion-writing tasks to their clerks. Wahlbeck and his colleagues used techniques from computational linguistics to ascertain the relative levels of

-
- 58 See Choi and Gulati, 'Which Judges Write Their Opinions (and Should We Care)?', above n 26, 1096–9 for a discussion of attempts to test authorship in fields outside the law. An important example is the pioneering work of Mosteller and Wallace, who attempted to determine the authorship of the unattributed essays in *The Federalist Papers* (as between Hamilton and Madison): Frederick Mosteller and David Wallace, *Inference and Disputed Authorship: The Federalist Papers* (Addison Wesley, 1964). See also C B Williams, *Style and Vocabulary: Numerical Studies* (Griffin, 1970) 132–3.
- 59 See, eg, C B Williams, 'Mendenhall's Studies of Word-Length Distribution in the Works of Shakespeare and Bacon' (1975) 62 *Biometrika* 207 (using word-length frequency distributions to examine whether Bacon or Marlowe wrote poems attributed to Shakespeare).
- 60 See Don Foster, *Author Unknown: On the Trail of Anonymity* (Henry Holt, 2000). Foster compared the text of *Primary Colors* with Joe Klein's *Newsweek* columns and found similarities in terms of use of particular adverbs and adjectives ending in the letter 'y', use of unusual adjectives, use of compound adjectives, high-frequency use of particular words, similar sentence structure and similarities in punctuation use.
- 61 Lawrence Solan and Peter Tiersma, *Speaking of Crime: The Language of Criminal Justice* (University of Chicago Press, 2005) 149–80.
- 62 See Foster, above n 60. In the Unabomber trial, Foster and others compared the known writings of Theodore J Kaczynski with the known writings of the Unabomber and found similarities in terms of sentence structure, spelling of particular words and use of grammar/punctuation. In the Jon-Benet Ramsey murder investigation, Foster used word frequency to attempt to identify the murderer.
- 63 Bryan Niblett and Jillian Boreham, 'Cluster Analysis in Court' [1976] *Criminal Law Review* 175.
- 64 See, eg, Michael Evans et al, 'Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research' (2007) 4 *Journal of Empirical Legal Studies* 1007; Chad Oldfather, Joseph Bockhorst and Brian Dimmer, 'Judicial Inaction in Action? Toward a Measure of Judicial Responsiveness' (Legal Studies Paper No. 10-32, Marquette University Law School, 15 July 2010).
- 65 See, eg, David Crump, 'Law Clerks, Their Roles and Their Relationships with Their Judges' (1986) 69 *Judicature* 236; Sally Kenney, 'Puppeteers or Agents? What Lazarus's Closed Chambers Adds to Our Understanding of Law Clerks at the US Supreme Court' (2000) 25 *Law and Social Inquiry* 185; Todd C Peppers, Michael W Giles and Bridget Tainer-Parkins, 'How Federal District Court Judges Select and Use Their Law Clerks' (2008) 71 *Albany Law Review* 623.

delegation by Justices Lewis Powell and Thurgood Marshall on the United States Supreme Court.⁶⁶ Choi and Gulati used computational linguistics techniques to examine the extent to which Federal Circuit Court judges in the United States wrote their own published opinions or delegated the opinion-writing task to clerks in the late 1990s.⁶⁷ This is not an issue in Australia, where evidence suggests judges' associates are far less important in drafting judgments.⁶⁸ In addition, none of these American studies examined the question of whether the judgments of Justice X were in fact written by Justice Y, which is the purpose of the current study.

IV DATA AND METHODOLOGY

A Data

Table 1 shows the number of single-authored High Court judgments attributed to Dixon, McTiernan and Rich that are available on the Australian Legal Information Institute ('AustLII') database.⁶⁹ We use judgments reported on AustLII because the methods we employ require us to automatically process the text files obtained from the judgments. The single-authored judgments attributed to Rich and McTiernan are divided into two parts: (1) judgments given at the time when Dixon served on the High Court (between 1929 and 1964), and (2) judgments given before or after Dixon was on the High Court (1913–1928 for Rich and 1965–1975 for McTiernan). The last column in Table 1 contains

66 Paul J Wahlbeck, James F Spriggs and Lee Sigelman, 'Ghostwriters on the Court? A Stylistic Analysis of US Supreme Court Opinion Drafts' (2002) 30 *American Politics Research* 166. These authors used eight measures; namely, type-token ratios (number of different words as a percentage of total number of words), once words (relative frequency of words that appear exactly once); average word length, word-length diversity, average sentence length, sentence length diversity, footnote frequency and footnote length.

67 See Choi and Gulati, 'Which Judges Write Their Opinions (and Should We Care)?', above n 26. These authors looked for repeated phrases in judgments using the GZip algorithm as well as examining self-citation rates and opinion length.

68 See Andrew Leigh, 'Behind the Bench: Associates in the High Court of Australia' (2000) 25 *Alternative Law Journal* 295, 296–7; Andrew Leigh, 'Associates' in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 34, 35; Rebecca Craske and Richard Haigh, 'Judgment Production' in Tony Blackshield, Michael Coper and George Williams (eds) *The Oxford Companion to the High Court of Australia* (Oxford University Press, 2001) 369, 369.

69 <<http://www.austlii.edu.au>>.

judgments flagged by Dixon in his diaries as being ghosted for Rich or McTiernan, as recorded by Ayres in his biography of Dixon.⁷⁰

Table 1: Judgments by Dixon, McTiernan and Rich, and flagged judgments

Justice	Rich		McTiernan		Dixon	Flagged
	1913-1928	1929-1964	1929-1964	1965-1975	1929-1964	1929-1964
Years	1913-1928	1929-1964	1929-1964	1965-1975	1929-1964	1929-1964
Number of Judgments	187	394	571	253	902	6

B Method

Our approach consists of applying techniques from computational linguistics to assign each judgment to one of the three judges in question. This is known as a *classification* task, where in our case the classes correspond to Dixon, McTiernan and Rich, and the task is to discriminate between Dixon's and McTiernan's judgments and between Dixon's and Rich's. The tools used to perform the classification task are known as *classifiers*. Classifiers are often built on the basis of samples for which the classes are known. In our case, we built classifiers for each pair of judges (McTiernan/Dixon and Rich/Dixon) based on all of Dixon's judgments and judgments written by McTiernan and Rich in the periods where they did not overlap with Dixon. We then used these classifiers to assign the disputed judgments (those written by McTiernan and Rich while Dixon served on the Court) to the most likely author. This section explains this process in detail.

1 From Texts to Features

The first step in every text classification task is to convert the texts into *features* that can be analysed by the classifiers. Commonly used features are: token unigrams (the actual words and punctuation marks in the text), bigrams (pairs of tokens), trigrams (triples), and so on; and part-of-speech unigrams, bigrams, trigrams, etc (a part-of-speech represents the syntactic category of a token, for example, noun, verb, adjective). More sophisticated features pertain to the structure of the text.⁷¹ We experimented with the following feature

70 Ayres, *Owen Dixon*, above n 11, refers explicitly to five judgments attributed to Rich and one judgment attributed to McTiernan that Dixon claims to have written, according to his diaries. The five judgments attributed to Rich are as follows (with the page reference to Ayres in parenthesis): *Federated State School Teachers' Association of Australia v Victoria* (1929) 41 CLR 569 (57); *R v Brislan: Ex parte Williams* 54 CLR 262 (73); *Sun Newspapers Ltd v Federal Commissioner of Taxation* (1938) 61 CLR 337 (93); *James v Commonwealth* (1935) 52 CLR 570 (320); *Peterson v Coleman* [1938] WCR 62 (320). The one judgment attributed to McTiernan that Dixon claims to have written is *Council of the Town of Southport v Corporation of the Trustees of the Order of the Sisters of Mercy in Queensland* [1935] HCA 53 (11 July 1925) (320). There are several other instances where Ayres refers to Dixon writing in his diaries that he had written judgments for McTiernan or Rich, but there is not enough information to ascertain the cases to which these refer.

71 Ol'ga Feiguina and Graeme Hirst, 'Authorship Attribution for Small Texts: Literary and Forensic Experiments' (2007) 22 *Literary and Linguistic Computing* 405.

combinations: (1) token *n*-grams;⁷² (2) token *n*-grams plus part-of-speech *n*-grams; and (3) token *n*-grams plus structural features of the text.⁷³ For each feature combination, we experimented with *n*-grams of different lengths, viz $n=1$; $n=1, 2$; $n=1, 2, 3$; and $n=1, 2, 3, 4$ (the same n was used for tokens and for parts-of-speech). The best results, which are reported in Part V, were obtained with the second feature combination for $n=1, 2, 3, 4$ (ie, token *n*-grams up to length 4, and part-of-speech *n*-grams up to length 4).

For example, one of Dixon's judgments starts with the sentence '[i]n my opinion these applications should be refused'.⁷⁴ This sentence is transformed into the following token feature list (the features are italicised):

- Unigrams: [in, my, opinion, these, applications, should, be, refused]
- Bigrams: [in my, my opinion, opinion these, these applications, applications should, should be, be refused]
- Trigrams: [in my opinion, my opinion these, opinion these applications, these applications should, applications should be, should be refused]
- 4-grams: [in my opinion these, my opinion these applications, opinion these applications should, these applications should be, applications should be refused].

Similarly, parts-of-speech are automatically extracted to obtain the following feature list:⁷⁵

- Unigrams: [IN, PRP\$, NN, DT, NNS, MD, VB, VBN]
- Bigrams: [IN PRP\$, PRP\$ NN, NN DT, DT NNS, NNS MD, MD VB, VB VBN]
- Trigrams: [IN PRP\$ NN, PRP\$ NN DT, NN DT NNS, DT NNS MD, NNS MD VB, MD VB VBN]
- 4-grams: [IN PRP\$ NN DT, PRP\$ NN DT NNS, NN DT NNS MD, DT NNS MD VB, NNS MD VB VBN]

This process is repeated for every sentence in each judgment. Then, the occurrences of each token and part-of-speech *n*-gram are counted to yield a representation of the judgment that is based on *n*-gram frequency⁷⁶ (in other words, the order of the *n*-grams is discarded to yield a simpler representation that can be analysed by the classifiers). For example, in the above judgment, the total number of token unigrams is 553 and the unigram *in* appears 18 times. Thus, the

72 An *n*-gram is a generic designation for a unigram, bigram, trigram, etc, where $n=1, 2, 3$, etc respectively.

73 Structural features and part-of-speech tags were automatically obtained using the tools in *OpenNLP* 1.4.3 <<http://opennlp.sourceforge.net>>. In addition, to ensure that judgements were classified according to language use, we automatically removed quotes and numbers in a pre-processing step.

74 *Wendo v The Queen* (1963) 109 CLR 559.

75 The part-of-speech tags are specified in the Penn Treebank Tag Set, *The Penn Treebank Project* <<http://www.cis.upenn.edu/~treebank>>, and are replicated in the appendix. For example, 'IN' stands for preposition or subordinating conjunction, and 'PRP\$' stands for possessive pronoun.

76 *N*-gram frequency is the number of times an *n*-gram appears in a document divided by the total number of *n*-grams of the same length in the document.

final representation of the judgment includes the frequency of *in* as $18 / 553 = 0.033$.

As this simple example illustrates, even though the n-gram order is disregarded, we may end up with a very large number of features (about 150 000). Therefore, we employ *feature selection* prior to building the classifiers in order to reduce the number of features and improve performance.⁷⁷ Specifically, the best performance was obtained with 30 000 features in the Rich/Dixon case, and 10 000 features in the McTiernan/Dixon case.

2 *Building and Using Classifiers*

In the last five decades, many techniques for building classifiers have been suggested. We chose to employ Support Vector Machines ('SVMs'),⁷⁸ one of the statistical classification methods developed in recent times, which exhibit state-of-the-art performance with respect to the authorship attribution task.⁷⁹ SVMs view each judgment as a point in space that is represented by a list of features. Given two sets of judgments with known authorship,⁸⁰ the SVM algorithm finds the best separation of these judgments into two areas in space based on their feature values, where each area corresponds to a different judge.⁸¹ Then, given a judgment with disputed authorship, its features are used to determine the area in which it falls. The judgment is assigned to the judge that corresponds to this area.

Figure 1 illustrates a simplified version of the operation of the SVM algorithm for the Rich/Dixon case, where we assume that there are only two features (as seen above, the typical number of features is much larger): each judgment with known authorship is represented as a point in the plane. In this case, the dashed line is chosen as a separator for judgments by the two judges, since there is a large margin between the line and each justice's judgments. Given a judgment with disputed authorship, its features are extracted and it is converted to a point in space. If this point falls above the dashed line, it is determined to be a Dixon judgment. Otherwise, it is attributed to Rich.

77 For a survey of feature selection methods see George Forman, 'An Extensive Empirical Study for Feature Selection Metrics for Text Classification' (2003) 3 *Journal of Machine Learning Research* 1289. To select the most informative features for the classification task, we employed a method called *Information Gain*, which is based on heuristics from information theory.

78 Vladimir N Vapnik, *Statistical Learning Theory* (Wiley-Interscience, 1998).

79 SVMs were found to yield the best performance in a competition of authorship attribution methods: Patrick Juola, 'Authorship Attribution' (2006) 1 *Foundations and Trends in Information Retrieval* 233.

80 In our case, these are (1) all of Dixon's judgments versus Rich's 1913–1928 judgments; or (2) all of Dixon's judgments versus McTiernan's 1965–1975 judgments.

81 In technical terms, SVMs find the hyperplane that best separates the samples from the two classes.

Figure 1: A simple SVM example for the Rich/Dixon classification problem

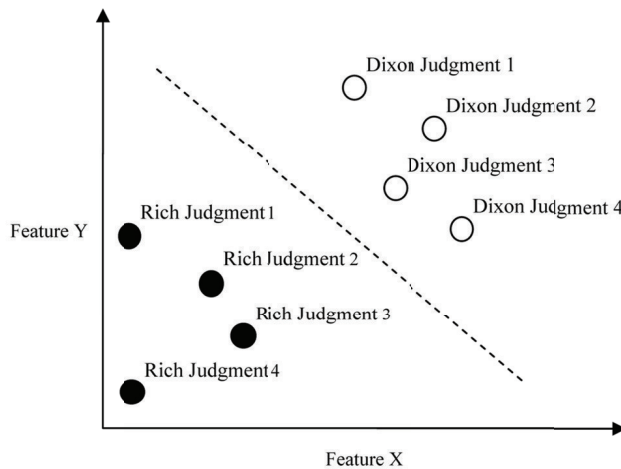


Table 2 presents the most informative token and part-of-speech n-grams for the Rich/Dixon classifier (columns 1 and 2) and the McTiernan/Dixon classifier (columns 3 and 4). *FREQ* means that the frequency of a token or part-of-speech n-gram is important, and *PRES* means that its mere presence is significant. Specifically, for Rich and Dixon, the entries in the upper part of the table present, for Rich's prose when compared with Dixon's, the 10 most informative token n-grams (column 1) and the 10 most informative part-of-speech n-grams (column 2), while the entries in the lower part of the table give similar information for Dixon's prose when compared to Rich's. Columns 3 and 4 present similar information for McTiernan versus Dixon.

Table 2: Top 10 token and part-of-speech n-grams that discriminate between Richard Dixon, and between McTiernan and Dixon

Rich/Dixon Classifier		McTiernan/Dixon Classifier	
Tokens	Parts of Speech	Tokens	Parts of Speech
Rich		McTiernan	
FREQ <i>being</i>	FREQ IN TO	PRES <i>I would</i>	FREQ DT NN IN DT
FREQ <i>this</i>	FREQ COMMA CC DET	PRES <i>necessary</i>	FREQ NN IN DT
FREQ <i>, and</i>	FREQ CC RB	PRES <i>. In my</i>	FREQ IN DT
FREQ <i>seems</i>	FREQ VBN NN	PRES <i>in my</i>	FREQ DT
FREQ <i>as to</i>	FREQ COMMA CC IN	FREQ <i>the</i>	FREQ NN IN DT NN
FREQ <i>. this</i>	FREQ EX VBZ	PRES <i>question</i>	FREQ IN DT NN
FREQ <i>only</i>	FREQ EX	PRES <i>."</i>	FREQ NN IN
FREQ <i>these</i>	FREQ VBP JJ	PRES <i>. I would</i>	FREQ DT NN
PRES <i>view</i>	FREQ COMMA JJ	PRES <i>opinion the</i>	FREQ VBP IN DT NN
FREQ <i>there is</i>	FREQ NN COMMA CC	PRES <i>".</i>	FREQ DT NN IN

Dixon		Dixon	
FREQ <i>I think that</i>	FREQ NN WDT MD	PRES <i>me to be</i>	FREQ NN COMMA RB
FREQ <i>when</i>	FREQ VBG NN IN	PRES <i>to me to be</i>	FREQ VBN COMMA
FREQ <i>the present</i>	FREQ VBN IN VBG	PRES <i>over</i>	FREQ MD VB VBN IN
PRES <i>is enough</i>	FREQ MD	PRES <i>true</i>	FREQ IN RB
FREQ <i>. in the</i>	FREQ VBD PERIOD	PRES <i>upon a</i>	FREQ VBN RB
FREQ <i>the provision</i>	FREQ NNS VBG DT	PRES <i>us</i>	FREQ MD VB
PRES <i>course</i>	FREQ RBR	PRES <i>the expression</i>	FREQ CC
FREQ <i>must be</i>	FREQ DT CD NNS	PRES <i>what</i>	FREQ NN PERIOD CC
FREQ <i>must</i>	FREQ TO PRP TO	FREQ <i>upon the</i>	FREQ WRB
FREQ <i>means</i>	FREQ DT NN WDT	PRES <i>the view</i>	FREQ COMMA

3 Dealing with Dataset Imbalances

As Table 1 shows, Dixon rendered significantly more judgments than Rich or McTiernan. The performance of statistical classifiers, including SVMs, is affected by such an imbalance in the number of judgments by each judge. To address this problem, we employ a *classifier ensemble*, where each classifier is built based on a subset of Dixon's judgments and all of the judgments by Rich or McTiernan.⁸² Specifically, in order to discriminate between Rich's and Dixon's judgments on the basis of the 187 judgments rendered by Rich prior to 1929 and all the 902 judgments rendered by Dixon (Table 1), we divide Dixon's judgments into five subsets, where each subset contains about 180 judgments. We then employ an ensemble comprising five classifiers, such that each classifier is built on the basis of one subset of Dixon's judgments and all the 187 judgments by Rich. Similarly, in order to build an ensemble to discriminate between McTiernan's and Dixon's judgments on the basis of the 253 judgments rendered by McTiernan after 1965 and all of Dixon's judgments, we divide Dixon's judgments into three subsets, each containing about 300 judgments. In this case, only three classifiers are sufficient to overcome the imbalance in the number of judgments rendered by McTiernan and Dixon.

Owing to the differences between the Dixon subsets used for building the classifiers in the ensembles, the classifiers may not always return the same justice for a particular judgment. For example, in the Rich/Dixon case, two classifiers

82 Two types of solutions for the class imbalance problem have been proposed: data-based and algorithmic (see Nitesh V Chawla, Nathalie Japkowicz, and Aleksander Kolcz, 'Editorial: Special Issue on Learning from Imbalanced Data Sets' (2004) 6 *ACM Special Interest Group on Knowledge Discovery and Data Mining Explorations* 1). The solutions at the data level comprise different forms of re-sampling, such as *oversampling* and *undersampling*. *Oversampling* involves replicating judgments by the judge with *less* judgments until the number of judgments (approximately) matches the number of judgments by the judge with *more* judgments. A classifier is then built using the judgments by the more prolific judge and the oversampled set of judgments by the less prolific judge. *Undersampling* involves drawing a subset from the judgments by the judge with *more* judgments that is (approximately) equal in number to the number of judgments by the judge with *less* judgments. A classifier is then built using the judgments by the less prolific judge and the undersampled set of judgments by the more prolific judge. We have adopted an undersampling approach, as it tends to outperform oversampling. In order to take into account all of Dixon's judgments, we employed the ensemble approach: see Yang Liu, Aijun An and Xiangji Huang, 'Boosting Prediction Accuracy on Imbalanced Datasets with SVM Ensembles' (Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, 2006) 107–18.

may return Rich as the author of a given judgment, while the other three classifiers may assign the same judgment to Dixon. One could envisage two main ways of reconciling these differences, by *majority vote* or by *unanimous vote*. According to the majority vote scheme, if a judgment is assigned to different judges by different classifiers, then the judge with most votes wins (for example, if a particular judgment is assigned to Dixon by two classifiers and to Rich by three classifiers, then it is deemed to be a Rich judgment). In contrast, the unanimous vote scheme requires all the classifiers to agree in order to determine the author of a judgment. If the classifiers do not produce a unanimous vote, then the judgment in question is not classified.

4 *Measuring Classification Performance*

Any solution to a classification problem involves selecting the most suitable methods from many existing options. For example, in our case we had to choose feature types (Part IV(B)(1)), a classification algorithm (Part IV(B)(2)), and an approach to dealing with the class imbalance problem (Part IV(B)(3)). An important question that arises is: how can we ensure that we made the right choices? That is, how can we evaluate the performance of our system?

Stratified ten-fold cross-validation is a procedure that is commonly used to evaluate classification performance based on samples with known classes (judgments with known authors, in our case). Under this procedure, the judgments are split into ten distinct folds (or subsets), such that nine folds are used for building the classifier ensemble, and one fold is used for testing. The folds are sampled in a *stratified* way that ensures that the class balance is the same across all folds. For instance, about 812 of the 902 judgments by Dixon and 168 of the 187 judgments by Rich (Table 1) are used to build the classifier ensemble.⁸³ The performance of this ensemble is evaluated by classifying the judgments in the testing fold, which in this case comprises the remaining 90 judgments by Dixon and 19 judgments by Rich (the number of judgments by each justice retains their proportion in the entire dataset). We then calculate the *accuracy* of this classification, which is defined as the percentage of judgments that were correctly assigned to the judge who wrote them.⁸⁴ This procedure is repeated ten times, each time with a different testing fold (with the remaining folds used for building the ensemble). The accuracies obtained with each testing fold are averaged to obtain the *overall accuracy* of the ensemble. This overall accuracy can be seen as our level of confidence in the classifications produced by our system. The stratified ten-fold cross-validation procedure ensures that the performance exhibited by the classifier ensemble is representative of all the data, and that results with a high or low accuracy are not due to a particularly felicitous

83 As indicated in Part IV(B)(3), each of the five classifiers in the ensemble is constructed using a balanced set of judgments comprising 162 of Dixon's judgments (one fifth of 812) and all 168 of Rich's judgments.

84 For a survey of performance measures see Marina Sokolova and Guy La Palme, 'A Systematic Analysis of Performance Measures for Classification Tasks' (2009) 45 *Information Processing & Management* 427. We chose the accuracy measure due to its simplicity and ease of interpretation.

or infelicitous testing fold. Once the accuracy of our ensemble is deemed satisfactory, it can be deployed on judgements with disputed authorship.

When we employ the unanimous voting scheme, which does not return a judge unless the vote is unanimous, we must also consider a measure of *coverage*, which indicates how often the classifier ensemble was able to classify a judgement (whether correctly or incorrectly). Typically, there is a trade-off between accuracy and coverage, that is, higher coverage normally yields lower accuracy. For instance, a system based on majority voting has full coverage, but makes decisions on cases about which it is less certain, thus having a lower accuracy than if it had not covered these cases.

V RESULTS

In this section, we report the results obtained with our system. We first evaluate the performance of our system using the judgments with known authorship, as described in Part IV(B)(4). We show that the classifications yielded by our system are highly accurate, meaning that our confidence in the system's performance on the disputed judgments is very high. We then use our system to classify the disputed judgments and show that it is quite likely that Dixon has authored about 18 per cent of Rich's judgments and four per cent of McTiernan judgments.

A Evaluating the Performance of our System

In order to measure the performance of our system, we conducted an experiment where we used the judgments known to have been written by only one of the three judges (in the case of Rich and McTiernan, the judgments that were certainly not written by Dixon). That is, we took the 187 judgments rendered by Rich prior to 1929, the 253 judgments rendered by McTiernan after 1964, and all of Dixon's judgments, and built two classifier ensembles: one to discriminate between the judgments of Rich and Dixon, and one to discriminate between McTiernan and Dixon. To obtain the accuracy and coverage of each ensemble, we performed stratified ten-fold cross validation, as described in Part IV(B)(4).

Table 3 shows the accuracy and coverage obtained by the classifier ensembles that yielded the best overall performance (token and part-of-speech n-grams up to length 4, as outlined in Part IV(B)(1)). For instance, the ensemble employed to discriminate between Rich's judgments and Dixon's yielded a unanimous vote for 92.96 per cent of the judgments (ie, 92.96 per cent coverage), of which 97.31 per cent were classified correctly (ie, 97.31 per cent accuracy).

Table 3: Accuracy and coverage of the classifiers on judgments with known authorship

Justice pairs	Rich/Dixon	McTiernan/Dixon
Accuracy	97.31%	99.11%
Coverage	92.96%	97.22%

As Table 3 shows, the classifier ensembles did not yield a unanimous vote for 7.04 per cent of the judgments by Rich and Dixon, and 2.78 per cent of the judgments by McTiernan and Dixon. When we employed majority voting to determine the authors for these few judgments, the accuracy of the Rich/Dixon classifier ensemble was only 60.53 per cent and that of the McTiernan/Dixon classifier was only 50 per cent. This indicates that majority voting is not suitable in this case.

Thus, owing to the reliability of the results obtained with unanimous voting (and the lack of reliability of the non-unanimous results), we propose to consider only unanimous voting for the remaining two experiments, where we classify judgments whose authorship is disputed. Based on the results of the first experiment, our confidence in the accuracy of classifications obtained by unanimous voting is more than 97 per cent for Rich/Dixon and more than 99 per cent for McTiernan/Dixon.

B Classifying Judgments with Disputed Authorship

In principle, with the exception of brief periods when Dixon was away from the Court,⁸⁵ all the judgments given by Rich and McTiernan during the time they served with Dixon on the Court could have been written by Dixon. Additionally, there are six judgments mentioned explicitly in Dixon's diaries as being ghosted for Rich or McTiernan.

We first proceed to classify the 394 judgments rendered by Rich and the 571 judgments written by McTiernan in the years 1929–1964. Table 4 shows the results obtained by the ensemble of five Rich/Dixon classifiers described in Part IV(B)(3). The first column indicates the percentage of classifiers that voted for Rich as the author of a judgment; the second column specifies the number of judgments that had the voting pattern stated in the first column; and the third column shows the percentage of these judgments (out of the total number of judgments by Rich). For instance, 100 per cent of the classifiers (that is, all five of them) voted for Rich as the author of 59.90 per cent of his judgments, 80 per cent (that is, four of the five) voted for Rich as the author of another 3.81 per cent of his judgments, and so on. There were 72 judgments (18.27 per cent of the total) for which none of the classifiers identified Rich as the author: ie, they all voted for Dixon as the author of these judgments. Table 5 presents similar information for the ensemble of three McTiernan/Dixon classifiers (as mentioned in Part IV(B)(1)), only three classifiers were required to overcome the imbalance in a number of judgments between McTiernan and Dixon).

As seen in Tables 4 and 5, the Rich/Dixon classifier ensemble returns a unanimous vote in 78.17 per cent of the cases (59.90 per cent plus 18.27 per cent), while this occurs in 86.34 per cent of the cases for the McTiernan/Dixon ensemble (81.96 per cent plus 4.38 per cent). These unanimous votes attribute 18.27 per cent of Rich's judgments and 4.38 per cent of McTiernan's judgments

85 For example, when Dixon served as Australian Minister to Washington for part of World War II and as a United Nations mediator in the Indo-Pakistani dispute over Kashmir in 1950.

to Dixon. This result suggests that a conservative estimate regarding the authorship of the judgments in question is that Dixon wrote approximately 18 per cent of Rich's judgments and four percent of McTiernan's judgments for the period Dixon was on the High Court.

Table 4: Attribution of Rich's judgments on the basis of a 5-classifier ensemble

Percentage of classifiers that attributed Rich's judgments to Rich	Number of judgments	Percentage of judgments
100%	236	59.90%
80%	15	3.81%
60%	27	6.85%
40%	23	5.84%
20%	21	5.33%
0%	72	18.27%
Total	394	100%

Table 5: Attribution of McTiernan's judgments on the basis of a 3-classifier ensemble

Percentage of classifiers that attributed McTiernan's judgments to McTiernan	Number of judgments	Percentage of judgments
100%	468	81.96%
66.7%	44	7.71%
33.3%	34	5.95%
0%	25	4.38%
Total	571	100%

We now repeat this exercise for the six judgments that were flagged by Dixon in his diaries as having been ghosted for Rich or McTiernan. Indeed, the three McTiernan/Dixon classifiers agree that the one judgment flagged by Dixon was written by him, rather than McTiernan. However, this does not happen for all the five judgments by Rich: three out of these judgments were unanimously attributed to Dixon, one judgment was attributed to Dixon by four out of five classifiers, and one judgment was unanimously attributed to Rich. It may be the case that this judgment was indeed authored by Rich, or it could have been

misattributed, as the expected accuracy of our method is not perfect (97 per cent in the Rich/Dixon case).⁸⁶

VI CONCLUSION

We have offered a computational mechanism to determine authorship of judgments on the basis of low-level linguistic features of discourse. There has been much speculation in legal circles about the extent to which Dixon ghosted for Rich and, to a lesser extent, for McTiernan. However, to date, there has been no quantitative evidence on the issue. Our approach shows a high level of coverage and accuracy for the unanimous voting scheme, enabling us to make our determinations with a high degree of confidence. Specifically, we find that about four per cent of McTiernan's judgments and 18 per cent of Rich's judgments were very likely authored by Dixon. These results are important because they suggest that breach of judicial propriety was more widespread on the Court in the middle decades of the twentieth century than previously thought, and that attribution of authorship on the Court cannot be taken at face value.

Our finding that Dixon wrote almost one in every five of Rich's judgments for the period both were on the Court may surprise some readers. This figure is at the upper-end of what we were expecting to find. It suggests that the extent to which Dixon ghosted for Rich is more pervasive than, perhaps, what was ever thought before, based on anecdotal evidence from the Dixon Diaries and other sources. However, we emphasise that our estimates are based on unanimous voting, which, in the world of computational linguistics, can be regarded as being fairly conservative. Our confidence in these attributions is more than 99 per cent for McTiernan's judgments and more than 97 per cent for Rich's judgments.

While our confidence in these attributions is very high, it is important to be aware of some of the main limitations of the study. First, we only considered each judgment as a single unit of text, while it may be possible that only certain parts of some of McTiernan's or Rich's judgments were written by Dixon. Indeed, Ayres suggests that Dixon may have written sections of McTiernan's and Rich's judgments.⁸⁷ Considering all the possible subparts of each judgment is an intractable problem, but a limited study on the attribution of smaller subparts, such as paragraphs, may improve performance in terms of accuracy and coverage. Such a study is beyond the scope of this paper, but it is an interesting avenue for future work.

Another possible limitation is that it might be argued that the whole premise of the study is subject to hindsight bias.⁸⁸ There is much research suggesting that people who are informed about how an event ended will view all information

86 The judgment attributed to Rich is *Federated State School Teachers' Association of Australia v Victoria* (1929) 41 CLR 569.

87 Ayres, 'Dixon Diaries', above n 14, 223.

88 See Hal R Arkes et al, 'Eliminating the Hindsight Bias' (1988) 73 *Journal of Applied Psychology* 305; Ralph Hertwig, Carola Fanselow and Ulrich Hoffrage, 'Hindsight Bias: How Knowledge and Heuristics Affect Our Reconstruction of the Past' (2003) 11 *Memory* 357.

about that event through that prism of knowledge. The classic example of such bias is the *res ipsa loquitur* fallacy – that is, the tendency to construe a defendant’s behaviour as negligent simply because we know what damage it caused.⁸⁹ The implicit assumption in this study, and in the controversy over Dixon ghosting for McTiernan and Rich, is that everything came from Dixon because he is regarded (for better or worse) as a genius. There is a chance other people could have been involved in ghosting for McTiernan and Rich. For example, we know from the Dixon Diaries that it is likely ‘Sammy’ Clyne ghosted for Rich on at least one occasion and this means that there might have been other ghost writers. However, this is not an issue we investigated. In this sense, the results reported in this study are a first step towards the use of computational linguistics to determine authorship of judgments in the High Court. Computational linguistics could be used to examine whether ghosting has been more widespread on the Court. Equally, computational linguistics has the potential to assist in areas other than research on the judiciary, including corporate bond contracts and securities fraud where authorship is often in dispute.⁹⁰

APPENDIX

Table A1: Part of Speech Tags from the Penn Tree Bank

#	Tag	Description	#	Tag	Description
1	CC	Coordinating conjunction	19	PRP\$	Possessive pronoun
2	CD	Cardinal number	20	RB	Adverb
3	DT	Determiner	21	RBR	Adverb, comparative
4	EX	Existential <i>there</i>	22	RBS	Adverb, superlative
5	FW	Foreign word	23	RP	Particle
6	IN	Preposition or subordinating conjunction	24	SYM	Symbol
7	JJ	Adjective	25	TO	<i>To</i>
8	JJR	Adjective, comparative	26	UH	Interjection
9	JJS	Adjective, superlative	27	VB	Verb, base form
10	LS	List item marker	28	VBD	Verb, past tense
11	MD	Modal	29	VBG	Verb, gerund or present participle

⁸⁹ See Erin M Harley, ‘Hindsight Bias in Legal Decision Making’ (2007) 25 *Social Cognition* 48. See also E Greene and L Ellis, ‘Decision Making in Criminal Justice’ in D Carson et al (eds), *Applying Psychology to Criminal Justice* (John Wiley & Sons, 2007).

⁹⁰ See Choi and Gulati, ‘Which Judges Write Their Opinions (and Should We Care)?’, above n 26, 1121–2.

12	NN	Noun, singular or mass	30	VCN	Verb, past participle
13	NNS	Noun, plural	31	VBP	Verb, non-3rd person singular present
14	NNP	Proper noun, singular	32	VBZ	Verb, 3rd person singular present
15	NNPS	Proper noun, plural	33	WDT	Wh-determiner
16	PDT	Predeterminer	34	WP	Wh-pronoun
17	POS	Possessive ending	35	WP\$	Possessive wh-pronoun
18	PRP	Personal pronoun	36	WEB	Wh-adverb