

A DATA DRIVEN APPROACH TO EVALUATING AND IMPROVING JUDICIAL DECISION-MAKING: STATISTICAL ANALYSIS OF THE JUDICIAL REVIEW OF REFUGEE CASES IN AUSTRALIA

DANIEL GHEZELBASH, * KEYVAN DOROSTKAR** AND SHANNON WALSH***

This article presents analysis of a database of over 6,700 applications for judicial review of refugee cases in the Federal Circuit and Family Court of Australia. The data reveals that the rate at which applications for judicial review are accepted by the Court varies widely based on the judge who hears the case and a number of other factors. While our findings are not necessarily a matter for concern, we argue that they do raise questions around the potential influence of cognitive and social biases in judicial decision-making, as well as in relation to the case management and resourcing of the Court. Drawing on recent research in the field of cognitive and behavioural sciences, we outline how statistics of the nature collected in our study could inform interventions and reforms aimed at addressing such biases and increase public confidence in the judicial system.

I INTRODUCTION

‘In an age of digital storage of information, and so the digital accessibility of information, there has arisen a greater demand for accountability of public institutions by reference to that information. This is neither to be feared or resented. It has to be recognised and taken into account.

* Associate Professor, Faculty of Law and Justice, University of New South Wales; Deputy Director, Kaldor Centre for International Refugee Law; Australian Research Council DECRA Fellow.

** MRES Candidate, Macquarie Law School.

*** LLB(Hons) Graduate, Macquarie Law School.

This research was partially funded by the Australian Government through the Australian Research Council (‘ARC’). Associate Professor Ghezelbash is the recipient of an ARC Discovery Early Career Award (project number DE220101189). Thanks are due to Matthew Groves, Gabrielle Appleby, Mary Crock and Saul Wodak for helpful comments, Harry Andresen for developing the technology and code used for our data collection, the team at the Kaldor Centre for International Refugee Law for assisting with the manual auditing of the data, and Jasmine Kassis for her research assistance. We also thank the anonymous reviewers for their deep engagement with our work and their many insightful comments and suggestions. All errors remain the authors’ own. The full dataset examined in this study along with similar data on decision-making at the Administrative Appeals Tribunal and Immigration Assessment Authority is available through the Kaldor Centre Data Lab <<https://www.kaldorcentre.unsw.edu.au/kaldor-centre-data-lab>>.

Accountability brings a need for being able to explain and justify how one is undertaking the task with which one is entrusted.'

– Chief Justice Allsop of the Federal Court of Australia¹

This study represents the first robust attempt at collecting and examining statistics on decision-making patterns of Australian judges in refugee cases. We draw on an original dataset covering over 6,700 applications for judicial review in the Federal Circuit and Family Court of Australia from 1 January 2013 to 11 March 2021.² The data was compiled using an automated computer code which was used to process each published decision and collect data on a wide range of variables in every case examined. We then ran statistical analysis examining the relationship between each factor and the success or failure of the judicial review application.

Reflecting the findings of similar studies in the United States ('US'), Canada and France,³ we found a large degree of variation in the success rates across individual judges. Our innovative use of computational methods to automatically code cases allowed us to examine a wide variety of other factors that may influence decision-making, including the respective caseloads of each judge, the time taken for judges to issue their judgment, the role of legal representation, the registry in which the application was filed, differences between the review of cases from the Administrative Appeals Tribunal ('AAT') and Immigration Assessment Authority ('IAA') and the gender of the judge.

While our findings are not necessarily a matter for concern, we argue that they do raise questions around the potential influence of cognitive and social biases in judicial decision-making, as well as in relation to the case management and resourcing of the Court. Our analysis in this regard is timely given the Australian Law Reform Commission's ('ALRC') recent report into judicial impartiality and the increased acceptance of the relevance of cognitive and behavioural science research into judicial decision-making.⁴ We outline how statistics of the nature collected in our study could inform interventions and reforms aimed at addressing cognitive and social biases, and to facilitate greater transparency that will increase public confidence in the judicial system. We also examine the question of whether

1 Chief Justice James Allsop, 'Courts as (Living) Institutions and Workplaces' (2019) 93(5) *Australian Law Journal* 375, 379.

2 Note that the Federal Circuit Court of Australia became known as the Federal Circuit and Family Court of Australia on 1 September 2021.

3 Jaya Ramji-Nogales, Andrew I Schoenholtz and Philip G Schrag, 'Refugee Roulette: Disparities in Asylum Adjudication' (2007) 60(2) *Stanford Law Review* 295; Jaya Ramji-Nogales, Andrew I Schoenholtz and Philip G Schrag, *Refugee Roulette: Disparities in Asylum Adjudication and Proposals for Reform* (New York University Press, 2009); Sean Rehaag, 'Troubling Patterns in Canadian Refugee Adjudication' (2007) 39(2) *Ottawa Law Review* 335; Sean Rehaag, 'Judicial Review of Refugee Determinations: The Luck of the Draw?' (2012) 38(1) *Queen's Law Journal* 1; Sean Rehaag, 'Judicial Review of Refugee Determinations (II): Revisiting Luck of the Draw' (2019) 45(1) *Queen's Law Journal* 1. See also data and analysis in respect to the French judges compiled by Michaël Benesty as part of the Supra Legum Project. The data is no longer publicly available following the French ban on statistical analysis of judicial decision-making discussed in n 12 and accompanying text below.

4 Australian Law Reform Commission, *Without Fear or Favour: Judicial Impartiality and the Law on Bias* (ALRC Report 138, December 2021) ('*Without Fear or Favour*').

statistical data could be used in the context of claims of apprehended bias in certain circumstances. In doing so, we seek to address and overcome the concerns and scepticism of the use of statistical data in this manner expressed in the recent jurisprudence of the Federal Court of Australia.⁵

Judicial decision-making is a complex undertaking involving many variables. At the outset, it is important to reiterate that we do not purport to draw any conclusive causal inferences about the role of the specific variables that we identify in influencing the outcome of a case. Our aim is much more modest. We identify correlations, provide context, and examine possible plausible explanations for the correlations. We then examine how the statistics may be used to inform interventions and reforms that may improve judicial decision-making and increase public confidence in the judicial system.

In a speech delivered in 2019, Chief Justice Allsop of the Federal Court of Australia provided some reflections on the appropriate uses of statistics and metrics of individual judicial behaviour.⁶ The Chief Justice cautioned against cherry-picking data points and expressed scepticism about metrics as a tool of measurement more broadly, warning that the ‘worth or accountability of courts as institutions is not metrically derivable or measurable’.⁷ His Honour noted, however, that such metrics could play an important role in *evaluating* the exercise of judicial power. While measurement is ‘an exercise in calculation’, evaluation involves the ‘drawing of value-laden conclusions from the balancing of considerations of the whole’.⁸ His Honour continued:

Many metrics or statistics for a judge, individually, or the court or part of the court, will be vital in raising questions (perhaps through comparative analysis about different periods of time or between different people, perhaps through an application of common sense) and in assessing what is happening. They may illuminate or suggest problems, personal or systemic; they may suggest solutions; they may help foster self-confidence, or help provoke a realisation of complacency; they may help the formulation of provisional hypotheses for investigation. But they will not give a measurement.⁹

In a similar vein, Opeskin and Appleby, writing in relation to the quantitative investigations of the Australian judiciary, caution that ‘while numbers can be revealing, they often give a partial account, and their value depends on sophisticated data collection and analytical methods accompanied by transparent explanation’.¹⁰

These critiques were front and centre in guiding our approach to data collection and analysis in this article. Our robust and comprehensive approach to data compilation and regression analysis overcomes concerns around cherry-picking and lack of analytical rigour. Moreover, we do not claim that the data we present here

5 *ALA15 v Minister for Immigration and Border Protection* [2016] FCAFC 30, [38]–[44] (Allsop CJ, Kenny and Griffiths JJ) (‘*ALA15*’); *CMU16 v Minister for Immigration and Border Protection* (2020) 277 FCR 201, 211 [36] (Jagot, Yates and Stewart JJ) (‘*CMU16*’).

6 Allsop (n 1).

7 Ibid 376.

8 Ibid.

9 Ibid 377.

10 Brian Opeskin and Gabrielle Appleby, ‘Responsible Jurimetrics: A Reply to Silbert’s Critique of the Victorian Court of Appeal’ (2020) 94(12) *Australian Law Journal* 923, 923.

is in anyway an objective measurement of judicial behaviour. Rather it is one piece of evidence that we hope can be used to contribute to a broader evaluation of how judicial review of refugee cases are dealt with in the Federal Circuit and Family Court of Australia. In the words of Chief Justice Allsop, '[i]t is the questions that are raised and the answers made and obtained from those questions that are the most valuable aspects of these statistics'.¹¹

Nonetheless, we acknowledge the very sensitive nature of this form of quantitative analysis of judicial behaviour and the risks of misuse or misinterpretation of the data. A similar project in France examining judicial decision-making in refugee cases caused significant controversy and contributed to the Government issuing a blanket ban on this form of research.¹² French judges raised concerns that the publication of statistics with individual names put pressure on them to move towards the average outcome and thus interfered with judicial independence.¹³ The French Government responded by introducing criminal sanctions stipulating that '[t]he identity data of magistrates and members of the judiciary cannot be reused with the purpose or effect of evaluating, analysing, comparing or predicting their actual or alleged professional practices'.¹⁴ Offending researchers may face up to five years in prison. While Australia does not have similar specific criminal provisions prohibiting statistical analysis of individual judicial decision-making, in some cases, the misuse of such data may amount to the common law offence of contempt by scandalising the court. This would occur where the research findings 'denigrates judges or the court so as to undermine public confidence in the administration of justice'.¹⁵

While acknowledging these potential risks, our intent is the exact opposite. We have a strong belief in the importance of the core judicial value of transparency in promoting public confidence in the courts.¹⁶ As Langford and Rask Madsen have

11 Allsop (n 1) 377.

12 See Michaël Benesty's *Supra Legum* project discussed at n 3 and accompanying text. For an example of similar work in other areas of French judicial decision-making which were taking place before the ban, see Christian Licoppe and Laurence Dumoulin, 'Judges, Algorithms and Jurisprudence: Initial Analyses of a Predictive Justice Experiment in France' (2019) 103 *Droit et Société* 535 <<https://doi.org/10.3917/drs1.103.0535>>.

13 Malcolm Langford and Mikael Rask Madsen, 'France Criminalises Research on Judges', *Verfassungsblog* (Blog Post, 22 June 2019) <<https://verfassungsblog.de/france-criminalises-research-on-judges/>>; Michaël Benesty, 'The Judge Statistical Data Ban: My Story', *Artificial Lawyer* (Blog Post, 7 June 2019) <<https://www.artificiallawyer.com/2019/06/07/the-judge-statistical-data-ban-my-story-michael-benesty/>>.

14 *Loi n° 2019-222 du 23 mars 2019* [Law No 2019-222 of 23 March 2019] (France) JO, 23 March 2019, art 33.

15 Judicial Commission of NSW, 'Criminal Trial Courts Bench Book Trial Procedure' (Bench Book, October 2002) [1-250], citing *R v Dunbabin; Ex parte Williams* (1935) 53 CLR 434; *Ex parte Attorney-General; Re Goodwin* (1969) 70 SR (NSW) 413; *Gallagher v Durack* (1983) 152 CLR 238. While the offence has been widely criticised and has largely fallen into disuse, there are recent examples of it being enforced: see, eg, *Ferguson v Dallow (No 5)* [2021] FCA 698.

16 Richard Devlin and Adam Dodek, 'Regulating Judges: Challenges, Controversies and Choices' in Richard Devlin and Adam Dodek (eds), *Regulating Judges: Beyond Independence and Accountability* (Edward Elgar, 2016) 1, 9; Monika Zalnieriute and Felicity Bell, 'Technology and the Judicial Role' in Gabrielle Appleby and Andrew Lynch (eds), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (Cambridge University Press, 2021) 116, 126 <<https://doi.org/10.1017/9781108859332.008>>.

argued, ‘using publicly available information to scrutinise the behaviour of the court system and its judges is in that view healthy for any democracy’.¹⁷ The courts are one of the institutional pillars of democracy and transparency around judicial decision-making enhances the authority and reputation of the courts.

The approach of the French Government in prohibiting this form of research is an outlier, and our research builds on and contributes to a burgeoning body of literature scrutinising statistics of decision-making in refugee cases around the world. As already mentioned, there is long history of this sort of research in the refugee space in the US and Canada.¹⁸ Rehaag, for example, publishes yearly statistics on all levels of Canada’s refugee determinations system.¹⁹ The Nordic Asylum Law and Data Lab is examining and comparing data on refugee status determination within and across Denmark, Sweden and Norway.²⁰ Similar research on disparities between the decision-making patterns of judges are also proliferating in other areas of law.²¹ We also seek to build on the small, but growing number of empirical and jurimetric studies of other areas of judicial decision-making in Australia,²² and address one of the key concerns raised in this literature with regards to the limited availability and publication of judicial data.²³

Our analysis proceeds in four Parts. In Part II, we briefly set out the process for assessing asylum claims in Australia, from the initial application, through to merits and judicial review. Part III sets out the methodology for collecting and analysing our data. Part IV provides an overview of our results. In Part V, we provide some analysis of the ramifications of the significant discrepancies between success rates before individual judges, and in particular the role that cognitive and social biases may be playing in decision-making, and how statistics may be used to counteract such biases. We identify how the statistics could be used as an intervention to mitigate implicit social and cognitive biases, as well as their potential utility in making out claims of apprehended bias against individual judges. We conclude in Part VI by highlighting the

17 Langford and Rask Madsen (n 13).

18 See above n 3 and accompanying text.

19 ‘Refugee Law Data’, *Refugee Law Lab* (Web Page) <<https://refugeelab.ca/projects/refugee-law-data/>>.

20 ‘Nordic Asylum Law & Data Lab’, *University of Copenhagen* (Web Page) <<https://asylumdata.ku.dk/>>. The project team includes scholars from University of Copenhagen (both Law and Computer Science), Uppsala University (Law and Medical Science) and Oslo University (Law). See also William Byrne et al, ‘Data Driven Futures of International Refugee Law’ *Journal of Refugee Studies* (forthcoming).

21 See, eg, Henrik Litleré Bentsen, ‘Court Leadership, Agenda Transformation, and Judicial Dissent: A European Case of a “Mysterious Demise of Consensual Norms”’ (2018) 6(1) *Journal of Law and Courts* 189 <<https://doi.org/10.1086/695555>>; Lucia Dalla Pellegrina, Nuno Garoupa and Fernando Gómez-Pomar, ‘Estimating Judicial Ideal Points in the Spanish Supreme Court: The Case of Administrative Review’ (2017) 52 *International Review of Law and Economics* 16 <<https://doi.org/10.1016/j.irl.2017.07.003>>; Erik Voeten, ‘The Impartiality of International Judges: Evidence from the European Court of Human Rights’ (2008) 102(4) *American Political Science Review* 417 <<https://doi.org/10.1017/S0003055408080398>>.

22 See, eg, Robin Creyke and John McMillan, ‘Executive Perceptions of Administrative Law: An Empirical Study’ (2002) 9(4) *Australian Journal of Administrative Law* 163; Robin Creyke and John McMillan, ‘Judicial Review Outcomes: An Empirical Study’ (2004) 11(2) *Australian Journal of Administrative Law* 82; Opeskin and Appleby (n 10); Gavin Silbert, ‘The First 24 Years of the Victorian Court of Appeal in Crime’ (2020) 94(6) *Australian Law Journal* 455.

23 Creyke and McMillan (n 22) 99–100; Opeskin and Appleby (n 10) 932.

important role that statistics of the nature collected in this article can play in promoting transparency and increasing public confidence in the judicial system.

II AUSTRALIA'S REFUGEE DETERMINATION SYSTEM AND THE ROLE OF THE FEDERAL CIRCUIT AND FAMILY COURT

Before examining the statistical data, it is important to provide some context about Australia's refugee status determination procedures and the role of the Federal Circuit and Family Court. Australia has separate procedures for processing asylum claims based on the applicant's mode of arrival. Those that arrive by plane go through the standard procedures, while those that arrive by boat without authorisation are subject to the so called 'fast track' procedures.²⁴ Broadly speaking, there are three steps that are common across both procedures. The application is initially assessed by the Department of Home Affairs by a delegate of the Minister. If the initial application is refused, merits review may be available at the Refugee and Migration Division of the AAT or the IAA. If merits review is unsuccessful, an applicant can seek judicial review of the tribunal's decision. The Federal Circuit and Family Court is generally the first forum for seeking such judicial review in refugee cases, with the potential to seek further review at the Federal Court of Australia, or the High Court of Australia. As a last resort, applicants can also apply directly to the Minister to intervene by exercising one of their public interest powers to grant the applicant a visa.²⁵

A Application to the Department

A number of important distinctions exist at the initial stage of applying for a protection visa for asylum seekers who arrived by boat ('fast track applicants'),²⁶ and those who arrived by plane. Fast track applicants are not entitled to apply for a protection visa unless the Minister exercises a personal and non-compellable discretion to allow this to occur (referred to as 'lifting the bar').²⁷ Secondly, fast track applicants are only eligible for Temporary Protection Visas lasting for three years or Safe Haven Enterprise Visas lasting for five years,²⁸ after which they must reapply and go through the refugee status determination process again. Successful

24 For a detailed comparison of the standard and fast track procedures, see Emily McDonald and Maria O'Sullivan, 'Protecting Vulnerable Refugees: Procedural Fairness in the Australian Fast Track Regime' (2018) 41(3) *University of New South Wales Law Journal* 1003 <<https://doi.org/10.53637/LQUA4141>>. For an assessment of the impact of the policy on the mental health of asylum seekers, see Mary Anne Kenny and Nicholas Proctor, 'The Fast Track Refugee Assessment Process and the Mental Health of Vulnerable Asylum Seekers' (2015) 23(1) *Psychiatry, Psychology and Law* 62 <<https://doi.org/10.1080/13218719.2015.1032951>>.

25 *Migration Act 1958* (Cth) ss 48B, 417, 195A ('*Migration Act*').

26 See *ibid* s 5 for a definition of fast track applicant.

27 *Ibid* s 46A.

28 *Migration Regulations 1998* (Cth) sch 1 reg 1403.

non-fast track applicants are generally granted permanent protection visas.²⁹ The process beyond these differences is largely the same. An application along with a statement of claims setting out the claims for protection is made to the Department of Home Affairs. Applicants are usually invited for an interview with a delegate of the Minister who will then decide whether to grant a protection visa.

B Merits Review

Where a decision-maker refuses to grant a protection visa, applicants can apply for the merits of their case to be reassessed. Merits review involves an independent reviewer standing in the shoes of the original decision-maker to make a decision afresh.³⁰ Non-fast track applicants have 28 days to seek merits review at the AAT.³¹ In conducting merits review, the AAT has the power to accept new information,³² and must invite an applicant to appear and give evidence.³³ The AAT also importantly has powers to substitute the decision of the department with its own (ie, issue a protection visa).³⁴

In contrast fast track applicants before the IAA do not have a right to an oral hearing and the majority of decisions are made ‘on the papers’.³⁵ The IAA is also not able to consider new information unless there are exceptional circumstances,³⁶ shifting the onus onto the asylum seeker applicant to provide all necessary information to the Department of Home Affairs at the initial stage.³⁷ Recent statistics show that the IAA has affirmed the Department of Home Affairs decision to refuse asylum claims in 91–4% of cases,³⁸ resulting in critics calling the process a ‘little more than a rubber stamp’ of the primary decision.³⁹ Lastly, the IAA does not have powers to replace the decision of the department and can only remit the case back with recommendations.⁴⁰ These limitations have led to numerous cases challenging whether the IAA is in fact conducting a full and independent merits

29 Ibid sch 1 reg 1401. Note, however, that asylum seekers arriving by air who are not immigration cleared are also excluded from applying for permanent protection visas: ibid sch 1 reg 1401(3)(d)(vi). For a critique of the costs and impact of temporary protection on refugees, see John van Kooy, ‘COVID-19 and Humanitarian Migrants on Temporary Visas: Assessing the Public Costs’ (Research Briefing Note No 2, Refugee Council of Australia, 29 July 2020).

30 *Minister for Immigration and Ethnic Affairs v Pochi* (1980) 4 ALD 139, 143 (Smithers J).

31 *Migration Act* (n 25) s 412.

32 Ibid s 424.

33 Ibid s 425.

34 Ibid s 415.

35 Ibid s 473DB.

36 Ibid s 473DD.

37 United Nations High Commissioner for Refugees, ‘Fact Sheet on the Protection of Australia’s So-Called “Legacy Caseload” Asylum-Seekers’ (Fact Sheet, 1 February 2018).

38 Asher Hirsch et al, Submission No 16 to Senate Legal and Constitutional Affairs References Committee, *The Performance and Integrity of Australia’s Administrative Review System* (March 2022).

39 Migration Law Program, ANU College of Law, Submission No 59 to Australian Law Reform Commission, *Inquiry into Traditional Rights and Freedoms: Encroachment by Commonwealth Laws* (December 2015).

40 *Migration Act* (n 25) s 473CC.

review.⁴¹ Ultimately, the High Court held that the IAA is engaged in a *de novo* consideration of the merits of the decision that has been referred to it, despite these limitations.⁴²

C Judicial Review

Where an applicant is unsuccessful at the AAT or IAA, they have the option to seek judicial review in relation to the findings of the tribunal.⁴³ Judicial review focuses on the narrow question of whether the decision-maker at the AAT or IAA made a serious legal error. In an attempt to completely oust judicial review in 2001, the executive introduced a privative clause in the *Migration Act*.⁴⁴ The High Court intervened, stating that any tribunal decision demonstrating jurisdictional error would fall outside the limitations of the privative clause and thus be subject to judicial review under section 75(v) of the *Constitution*.⁴⁵ This means that applicants seeking judicial review of refugee cases must demonstrate jurisdictional error.⁴⁶ This requires that the decision-maker has made ‘a decision outside the limits of the functions and powers conferred on him or her, or does something which he or she lacks power to do’.⁴⁷ Common grounds for judicial review in refugee cases include things such as denial of procedural fairness, misconstruction of a legal test, unreasonableness and illogicality, failure to deal with a claim or an integer of a claim or actual or apprehended bias.⁴⁸ The court does not concern itself with broader questions in relation to the findings of fact, or whether the tribunal reached the correct or preferable decision.

Applications for judicial review are generally heard by the Federal Circuit and Family Court in the first instance. This means the court is responsible for hearing the vast majority of refugee judicial review cases. While it is possible for

41 See *BMB16 v Minister for Immigration and Border Protection* (2017) 253 FCR 448, 458 (Besanko J); *Minister for Immigration and Border Protection v AMA16* (2017) 254 FCR 534, 555–6 (Griffiths J); *BVD17 v Minister for Immigration and Border Protection* (2019) 268 CLR 29.

42 *Plaintiff M174/2016 v Minister for Immigration and Border Protection* (2018) 264 CLR 217, 242 (Gageler, Keane and Nettle JJ).

43 For an overview of the largely failed attempts by the legislature to limit the scope of judicial review of migration cases, including in the context of fast track procedures, see Grant Hooper, ‘Three Decades of Tension: From the Codification of Migration Decision-Making to an Overarching Framework for Judicial Review’ (2020) 48(3) *Federal Law Review* 401 <<https://doi.org/10.1177/0067205X20927811>>. See also Mary Crock, ‘Judging Refugees: The Clash of Power and Institutions in the Development of Australian Refugee Law’ (2004) 26(1) *Sydney Law Review* 51.

44 *Migration Act* (n 25) s 474; *Migration Legislation Amendment (Judicial Review) Act 2001* (Cth).

45 *Plaintiff S157/2002 v Commonwealth of Australia* (2003) 211 CLR 476, 533–4 [161] (Callinan J). See also *Minister for Immigration and Multicultural Affairs v Yusuf* (2001) 206 CLR 323; *Abebe v Commonwealth* (1999) 197 CLR 510.

46 Note the *Administrative Decisions (Judicial Review) Act 1977* (Cth) cannot be relied upon in judicial review of refugee cases. See *Administrative Decisions (Judicial Review) Act 1997* (Cth) sch 1 s 3(da)–(db).

47 *Re Refugee Tribunal; Ex parte Aala* (2000) 204 CLR 82, 141 (Kirby J); *Hossain v Minister for Immigration and Border Protection* (2018) 264 CLR 123, 132 [23] (Kiefel CJ, Gageler and Keane JJ). See also *MZAPC v Minister for Immigration and Border Protection* (2021) 390 ALR 590 on the content and proof of establishing jurisdictional error.

48 Hooper (n 43) 413–17; Mary Crock and Laurie Berg, *Immigration, Refugees and Forced Migration: Law, Policy and Practice in Australia* (Federation Press, 2011) ch 19.

applicants to make further appeals in the Federal Court of Australia and the High Court of Australia, this only happens in a small percentage of cases.

III METHODOLOGY

For this study, we compiled an original database of judicial review decisions of refugee cases handed down by the Federal Circuit and Family Court of Australia from 1 January 2013 to 11 March 2021. The data was retrieved from the Australasian Legal Information Institute ('AustLII') database. The Court's annual report confirms that the AustLII database contains the complete record of the Court's published decisions.⁴⁹ It is important to note, however, that this does not represent all judicial review applications of refugee cases finalised by the Court. We do not capture cases that were resolved without a written judgment being issued. This includes instances where cases were discontinued, finalised by consent, or delivered *ex tempore* (orally) and no party to the case sought written orders. Information on these cases is not publicly available. We also could not capture data from cases which were not published or publicly available on AustLII at the time of obtaining the data. AustLII is a live platform that continually publishes cases on its website. Some cases within our date range may have been published on AustLII retrospectively after the date we obtained the data. To source refugee review decisions, we used the search term 'refugee' and 'protection visa' in the Federal Circuit and Family Court database from 1 January 2013 to 11 March 2021. In total, we identified 6,756 relevant judgments.

A computer program was written to convert these judgments into a plain text format. This process essentially ensured that the font and spacing of words in every judgment was the same. A judgment in the dataset was then parsed using Python language programming into the different syntactic components of the data points that we needed to extract. This involved a computer program analysing a string of symbols in natural language and separating the data according to the prescribed rules that had been coded. The prescribed rules required the computer program to first identify the required heading and then locate the corresponding text relevant to that heading and extract that information.

49 Federal Circuit Court of Australia, *Annual Report 2020–2021* (Report, 8 September 2021) 55 ('FCCA Annual Report').

One test case was broken down and analysed into the following components:

- Case name
- File Number
- Judge Name
- Hearing Date
- Location of the Hearing (eg, Sydney)
- Date Judgment Delivered
- First Respondent
- Second Respondent
- Applicant
- Catch Words
- Legislation
- Orders (reproduced directly from court documents)
- Representation

A majority of the judgments in the dataset followed the same structure and layout. Hence, a consistent pattern of where to find the required data could be established for all the cases. We then developed a Python program using Regular Expressions to analyse the cases. Regular Expressions identify and learn the allocated pattern and can locate the required headings and whether the corresponding information required was adjacent or below the headings. It was able to identify the patterns and extract the required information for each case. All this data was then exported into an excel spreadsheet.

As we were only interested in applications for judicial review of refugee determinations; cases which the Federal Circuit and Family Court categorised as involving matters other than refugee determinations were filtered out. We did this by manually reviewing the ‘catchwords’ and the ‘applicant’ sections of the data for each case. Section 91X(2) of the *Migration Act* dictates that the court must not publish the names of applicants for protection visas.⁵⁰ Hence, any data which contained the applicant’s identity was filtered out, as this clearly was not a refugee review application.⁵¹ Furthermore, the catchwords which provide a brief summary of the main issues of the case, were used to manually eliminate any further cases that were not in relation to protection visas.

A small number of cases categorised as refugee review decisions appeared to be matters relating to procedural and interlocutory issues including: applications for the extension of time, injunctions and situations where a party does not make an appearance for the hearing. These cases were retained because no means could be devised to exclude them consistently.

To determine whether an applicant had been successful or not in their judicial review; we created an excel formula which detected key words in the orders column of the data, revealing the outcome of the case. Where judges found in favour of the applicant, in a majority of cases, they issued a writ of certiorari

⁵⁰ *Migration Act* (n 25) s 91X(2).

⁵¹ Refugee applicants are referred to using alphanumeric codes in case names.

quashing the decision of the second respondent and usually a cost order against the first respondent. Where applicants for judicial review were unsuccessful, in a majority of cases, judges stated that the application was dismissed and cost orders were made against the applicant. We developed an Excel formula which identified the words ‘quash’ and ‘dismiss’ in the orders of each case. Where this language was not used in the orders, we developed another formula which identified who the cost orders were made against; we then read the complete orders of these cases manually to ensure accuracy.

The data went through two rounds of independent auditing. The auditing process involved manually checking every data point collected against the actual judgment. An initial independent audit of 5% of the cases found a 95% accuracy rate.⁵² Based on this feedback, some improvements were made to both the automatic and manual stages of the coding process and the dataset was updated. For example, the audit identified that the majority of judgments by Judge Emmett did not use the same structure and layout as the other judgments in our dataset. As our code relies on patterns and extracting the same data from the same location on the judgment, it was unable to accurately extract data from Judge Emmett’s decisions. Subsequently, we amended the code to better capture this variation and extracted that data again. A subsequent audit of 5% of the data revealed an accuracy rate of 99.91%.⁵³

The data was then analysed using Jamovi, a software built on the R statistical language. After the data was imported into Jamovi, a series of analyses were run to determine the relationship between different factors and the outcome of an applicant’s case. Different tests were used to examine the relationship between the decision made and other factors depending on whether the variables were categorical or continuous. Categorical variables are discreet and contain a finite number of categories. Examples include gender, decision outcome, and location. Continuous variables are numeric variables which can be counted or measured. Examples include the number of cases heard by a judge or the amount of time taken to make a decision. For relationships between two categorical variables, for example the decision made and the gender of the judge, the first stage of analysis involved using a chi-square test of independence to assess whether there was a significant association between each of the variables mentioned above and the outcome of the case. Once a statistically significant association was found, the strength of this association was determined by identifying the effect size using Cramér’s *V*. This is a number between 0 and 1 where a 0 indicates no association

52 The selection of the 5% of data that was audited was randomised through an excel formula and proportionate to the number of cases each judge had decided (ie, judges with a greater caseload were also equally proportionately represented in the number of cases that were audited in the 5% of cases selected). The audit was carried out externally by a team based at the Andrew and Renata Kaldor Centre for International Refugee Law, UNSW.

53 The audit followed the same process identified in the note above. The only errors identified in the audit were two cases where the status of legal representation was incorrectly noted. This was not a result of the automatic coding, but rather the manual conversion of the relevant extracted text into the standardised categories of represented, unrepresented or self-represented. Two of the 388 cases identified for the audit were no longer available on AustLII and could not be checked so were excluded for the purposes of these figures.

between 2 variables and 1 indicates that the variables are perfectly associated and are completely dependent on each other. A number greater than 0.6 is often interpreted as a 'strong association', between 0.3 and 0.6 as a 'moderate association' and between 0.0 and 0.3 as a 'small association'. However, it should be noted that the interpretation of effect sizes is context dependent.⁵⁴ Given the novelty of our research, there is no baseline or external criteria to interpret the importance of effect size in this context. Rather, we use the labels of association strength for the purpose of communicating and comparing results, both within this article, and across future research in this space.

For relationships between two continuous variables, for example the number of cases a judge heard and their average acceptance rate, Spearman's rank-order correlation coefficient, also called Spearman's rho, was used. This is a nonparametric test which assesses the correlation between two variables. Nonparametric tests are used where it is not assumed that data has come from prescribed models which are determined by parameters. As the data used did not fit the assumptions for parametric tests like Pearson's R, Spearman's coefficient was used to assess the correlation between all numeric variables. Spearman's rho can range from +1 to -1 when +1 indicates a perfect positive association, 0 indicated no correlation and -1 indicates a perfect negative correlation.

For relationships between one continuous and one categorical variable, where one variable is dependent on the other predictor variable, the Eta statistic was used. The Eta statistic describes the level of variance explained in the dependent variable by a predictor. The measure of association ranges from 0 to 1, with 0 indicating no association and 1 indicating a high degree of association. When the measure of association value is squared, it produces the coefficient of determination which indicates how much variance on the dependent variable can be accounted for by the predictor variable as a percentage.

For relationships between several variables, more than one of these tests was used. For example, a chi square test of association and Cramér's *V* were used to identify whether there was a correlation between legal representation and judges. This test revealed that legally represented applicants are not equally divided across judges. To assess whether this unequal allocation of represented applicants may be a factor in explaining variation in acceptance levels between judges, Spearman's rho was used to assess the correlation between each judge's average acceptance rate and their rate of represented applicants.

Two points of caution are worth noting before moving on to presenting the results. First, the nature of the data only allows for inductive statistical reasoning. Inductive reasoning begins with specific observations which are used to reach an overarching conclusion.⁵⁵ It cannot prove a causal link and can only predict the

54 James Rosenthal, 'Qualitative Descriptors of Strength of Association and Effect Size' (1996) 21(4) *Journal of Social Service Research* 37.

55 Zara O'Leary, *The Social Science Jargon Buster* (Sage Publications, 2011) 57.

most probable links based on the evidence at hand. This can be contrasted with a deductive statistical reasoning, which begins with a premise or idea, and then uses specific observations to *prove* that conclusions drawn from that premise are correct.⁵⁶ The data collected by this study does not capture the complex behavioural and organisational factors of judicial decision-making to support a deductive approach.

Second, consistency in judicial decision-making should not and cannot be used as a proxy for accuracy. Judicial review of refugee cases involves complex legal procedures and questions that do not lend themselves to uniquely black and white results. Hence in such a context it is a truism that ‘conscientious decision-makers, applying their minds to the same set of facts, may sometimes reasonably come to different conclusions’;⁵⁷ which can be both substantiated and justified. While the principle of consistency forms a key part of the concept of justice; the practical implications specifically in the refugee context mean that consistency should be ‘in the service of fair and just decision-making’.⁵⁸ A certain degree of variation is to be expected in a well-functioning judicial system. Even significant variations are not necessarily an indication of issues in the quality of decision-making. We, however, suggest that where statistically significant discrepancies do exist, they warrant further examination and analysis to identify potential explanations.

IV RESULTS

This section sets out the key findings from the dataset. Overall, applications for judicial review of refugee decisions were rarely successful. Out of 6,756 cases, only 519 succeeded in judicial review. That means that only 7.68% of cases were successful, and 91.77% resulted in the court upholding the decision of the tribunal to refuse the protection visa. Our algorithm was unable to ascertain the outcome of 0.55% of cases. The data also allows us to examine the relationship between various variables against success rate. These include the judge to whom a case is assigned, the size of the caseload decided by a judge, the time taken to issue a judgment, legal representation, the registry in which the application for review was lodged, whether the decision being reviewed was made by the IAA or AAT and the gender of the judge.

56 Ibid.

57 Hugo Storey, ‘Consistency in Refugee Decision-Making: A Judicial Perspective’ (2013) 32(4) *Refugee Survey Quarterly* 112, 114 <<https://doi.org/10.1093/rsq/hdt018>>.

58 Ibid.

Table 1: Judges Who Have Made 50+ Decisions

Judge Name	Count of orders	Success rate (%)	Success rate relative to median (%)
Judge Jones	65	23.08	185.19
Judge Riley	161	21.74	168.65
Judge Riethmuller	172	21.51	165.84
Judge Heffernan	79	17.72	119.00
Judge Kendall	100	16.00	97.73
Judge Barnes	222	13.51	67.00
Judge Lucev	170	12.94	59.93
Judge Wilson	104	12.50	54.48
Judge Young	85	11.76	45.39
Judge Smith	315	11.75	45.16
Judge Driver	792	11.36	40.43
Judge A Kelly	90	11.11	37.31
Judge Harland	54	11.11	37.31
Judge Humphreys	101	10.89	34.59
Judge Cameron	219	8.22	1.57
Judge McNab	113	7.96	-1.57
Judge Manousaridis	445	7.42	-8.36
Judge Mercuri	54	7.41	-8.46
Judge Burchardt	88	6.82	-15.74
Judge Jarrett	169	6.51	-19.56
Judge Baird	50	6.00	-25.85
Judge Egan	115	4.35	-46.27
Judge Lloyd-Jones	92	3.26	-59.70
Judge Raphael	65	3.08	-61.98
Judge Nicholls	406	2.71	-66.52
Judge Hartnett	186	2.15	-73.42
Judge Street	1144	2.01	-75.15
Judge Dowdy	183	1.64	-79.74
Judge Emmett	391	1.28	-84.20
Judge Vasta	165	0.61	-92.51

A Judges

The data reveals significant variation in the success rates across individual judges. In total, there were 52 judges in our dataset. We focus our analysis here, however, on the 30 judges who decided 50 or more refugee cases (see Table 1). The rationale is that the larger dataset reduced the likelihood of random variation. The overall success rate for judges in this group ranged from Judge Jones, at 23.08%, through to Judge Vasta, at 0.61%.

The data shows some significant statistical outliers of the success rates of individual judges. Judge Vasta has found in favour of the applicant once out of 165 refugee review cases he has presided over. Judge Emmett has decided in favour of the applicant in 5 cases out of 391 cases and Judge Street found in favour of the applicant in 23 cases out of 1,144 cases. In contrast, Judge Riley and Reithmuller have decided in favour of the applicant on 35 and 37 occasions out of 161 and 172 cases respectively.

A chi-square test of independence showed that there was a significant association between a decision-maker and the outcome of a case, $X^2(51, N = 6756) = 349, p < .001$. An analysis of effect size using Cramér's V , which determines how strongly variables are associated, reveals this association is of small strength (0.227). While we do note that every case is unique and should be assessed on its own merits, this high degree of statistically significant variability does raise questions which warrant further analysis. The docket system used to assign cases to judges should in theory provide a randomised sample of cases to each judge. Given that our sample only examines data from judges who have decided 50 or more cases, this further reduces the chance of random variation in terms of the relative merits of the cases heard by each judge. Potential factors contributing to this discrepancy are explored in Part V below.

B Distribution of Cases

The data demonstrates significant variation in the distribution of cases, with a handful of judges deciding the majority of judicial review of refugee cases. As already stated, 52 judges decided the 6,756 cases that constitute the full dataset. However, 57% of these cases have been decided by only 6 judges. Judges Street, Smith, Nicholls, Manousaridis, Emmett and Driver have collectively decided 3,493 cases out of the 6,756 case load. The spread of allocation of cases ranges from Judge Willis who heard only 1 case all the way to Judge Street who heard 1,144 cases. The average number of cases that a judge would hear across the dataset is 130 cases. There are a number of factors which explain this distribution. While the Court uses a docket case management process where matters are 'randomly allocated to a judge',⁵⁹ matters that require expertise in a specific jurisdiction are assigned to a judge who

is a member of the respective specialist panel.⁶⁰ Judicial review of refugee cases are assigned to the migration and administrative law panel. Hence the docket system randomly allocates the majority of the cases to judges within that panel. Secondly, as discussed further below, judges who finalise cases more quickly on average get assigned a higher number of cases under the docket system.

In determining whether the number of cases allocated to a judge impacts their acceptance rate, outliers need to be accounted for. For example, Judge Street, who has decided 1,144 cases, has an acceptance rate of only 2.0%. In order to account for this, the analysis in this Part considered the relationship between the number of cases a judge heard and their *average* acceptance rate. In other words, do we see a change in the average acceptance rates of individual judges as their caseload increases? This analysis shows that there is no significant relationship between the number of cases heard by a judge and their average acceptance rate. Given that the number of cases and acceptance rates are both numerical variables, their relationship can be measured using Spearman's rho. The *p*-value is over 0.054 which means the relationship between the variables is not significant and the Spearman's rho is irrelevant ($\rho(50) = 0.268$, $P = 0.054$). There is thus no significant relationship between the number of cases heard by a judge and their average acceptance rate.

C Time Taken to Issue a Judgment

There was significant variation across the dataset in terms of the time taken to deliver a judgment. The average amount of time between the hearing and when judgment was delivered was 64 days. The average for judges who made 50 or more decisions, however, increased to 77 days. Some notable outliers include: Judge Vasta who took 0.71 days on average to decide cases; Judge Street who took 1.34 days on average and Judge Lucev who took 276.37 days on average to deliver judgment after hearing. The very short turnaround by Judges Vasta and Street are presumably explained by their use of *ex tempore* judgments. These are judgments delivered orally immediately after the hearing of the matter. For example, of the 1,144 cases decided by Judge Street, 1,000 of those were delivered on the same day as the hearing, presumably *ex tempore*. As Chief Justice Allsop has noted, there are a variety of institutional factors and factors specific to each case which influence the length of time needed to reach a judgment.⁶¹

60 Ibid. '[T]he actual practice of allocating matters resembles much more closely a system of random allocation based on effective resource allocation principles': for further details on how the docket system operates at the Federal Circuit and Family Court, see *Without Fear or Favour* (n 4) 199–200.

61 Allsop (n 1) 376.

As such, variations in the time taken to reach decisions between judgments, even when extreme, may not necessarily be a cause for concern.⁶²

One observation which may raise questions worthy of further investigation is the correlation between the duration of time it takes to decide a case and the chances of success. Applicants are 10 times more likely to succeed in their judicial review application if judgment is delivered 2–3 months after hearing, compared to if judgment is delivered in under a month after hearing. The Eta statistic was used to determine whether there is an association between the outcome of a case and the length of time it takes a judge to decide a case. This test indicated that 19.6% of the variance in outcome can be accounted for by the length of time it takes to decide a case. Judges who have low success rates for applicants including Judges Vasta, Street and Emmett, take far less time to deliver judgments after hearing. These judges also deliver their judgments *ex tempore* more frequently than the rest of the judiciary. The risks associated with such heavy reliance on *ex tempore* judgments are explored in Part V.

D Legal Representation

Legal representation appears to significantly increase an applicant's chances of success. 40.5% of applicants had some form of legal representation (with the presence of a solicitor and/or barrister), 54.6% of applicants were self-represented, 4.3% made no appearance. The statistics show that applicants with legal representation are on average six times more likely to succeed than self-represented applicants. Self-represented applicants were successful in judicial review in just 89 cases out of 3,698 cases. In contrast, represented applicants were successful in 430 cases out of the 2,764 cases. Table 2 provides a breakdown of the number of represented and unrepresented applicants that each judge has presided over and the percentage of success of represented and unrepresented applicants before each judge. The mean of how many self-represented applicants is within each judge's caseload is 53.78% with a standard deviation of 9.07%. One hundred per cent of the data falls within 2 standard deviations from the mean, meaning that the distribution of represented and self-represented applicants is unlikely to explain the 280% variation in outcomes between individual judges set out in Table 1. Table 3 shows how much more likely an applicant is to succeed if they have legal representation before each judge. Represented applicants before Judge Driver are 47 times more likely to succeed than their unrepresented counterparts. This is a significant statistical outlier.

A chi-square test of interdependence confirms that there is a significant association between having legal representation and having a positive outcome, $\chi^2(1, N = 6080) = 375, p < 0.001$. The effect size was small (0.248).

62 On the issue of whether delays can constitute a jurisdictional error, see *WZASS v Minister for Immigration, Citizenship, Migrant Services and Multicultural Affairs* (2021) 282 FCR 516, 527 [52] (Katzmann, O'Bryan and Jackson JJ).

Table 2: Legal Representation

Judge	Total number of cases (represented)	Success rate of represented applicants (%)	Total number of cases (self-represented)	Success rate of self-represented applicants (%)
Judge Baird	19	10.53	28	3.57
Judge Barnes	63	33.33	145	6.21
Judge Burchardt	48	8.33	39	5.13
Judge Cameron	102	9.80	114	7.02
Judge Dowdy	48	4.17	125	0.80
Judge Driver	292	29.79	469	0.64
Judge Egan	61	8.20	54	0.00
Judge Emmett	93	3.23	247	0.81
Judge Harland	22	27.27	30	0.00
Judge Hartnett	79	5.06	106	0.00
Judge Heffernan	41	31.71	37	2.70
Judge Humphreys	60	16.67	38	2.63
Judge Jarrett	87	11.49	81	1.23
Judge Jones	31	35.48	33	12.12
Judge A Kelly	44	22.73	45	0.00
Judge Kendall	38	28.95	60	8.33
Judge Lloyd-Jones	28	10.71	64	0.00
Judge Lucev	44	34.09	109	6.42
Judge Manousaridis	146	15.75	293	3.41
Judge McNab	49	14.29	63	3.17
Judge Mercuri	34	11.76	20	0.00
Judge Nicholls	149	6.04	216	0.93
Judge Raphael	11	18.18	44	0.00
Judge Riethmuller	102	31.37	68	7.35
Judge Riley	84	34.52	76	7.89
Judge Smith	156	19.23	158	4.43
Judge Street	464	3.88	621	0.81
Judge Vasta	85	1.18	71	0.00
Judge Wilson	46	23.91	53	3.77
Judge Young	38	15.79	46	8.70

Table 3: Correlation between Representation and Increased Likelihood of Success before Each Judge

Judge	Likelihood of success if represented (%)
Judge Driver	46.58
Judge Harland	27.27
Judge A Kelly	22.73
Judge Raphael	18.18
Judge Mercuri	11.76
Judge Heffernan	11.73
Judge Lloyd-Jones	10.71
Judge Jarrett	9.31
Judge Egan	8.20
Judge Nicholls	6.52
Judge Wilson	6.34
Judge Humphreys	6.33
Judge Barnes	5.37
Judge Lucev	5.31
Judge Dowdy	5.21
Judge Hartnett	5.06
Judge Street	4.82
Judge Manousaridis	4.62
Judge McNab	4.50
Judge Riley	4.37
Judge Smith	4.34
Judge Riethmuller	4.27
Judge Emmett	3.98
Judge Kendall	3.47
Judge Baird	2.95
Judge Jones	2.93
Judge Young	1.82
Judge Burchardt	1.63
Judge Cameron	1.40
Judge Vasta	1.18

E Registry

The data shows a significant degree of variation in the average success rates across different Federal Circuit and Family Court registries (see Tables 4 and 5). Applicants are two times more likely to succeed in judicial review in Melbourne compared to Sydney. Three out of the six judges with the highest rates of success are located in Melbourne. Similarly, four out six judges with the lowest success rate for applicants are located in Sydney.

A chi-square test of independence showed that there was a significant association between registry location and case outcome, $X^2(15, N = 6756) = 84.4, p < 0.001$. An analysis of effect size using Cramér's V reveals this association is of small strength (0.112).

Table 4: Sydney Registry

Judge	Count of orders	Count of in favour of applicant	Percentage (%)
Judge Barnes	222	30	13.51
Judge Smith	301	36	11.96
Judge Driver	781	90	11.52
Judge Humphreys	97	11	11.34
Judge Cameron	214	18	8.41
Judge Manousaridis	445	33	7.42
Judge Baird	50	3	6.00
Judge Lloyd-Jones	92	3	3.26
Judge Raphael	62	2	3.23
Judge Nicholls	406	11	2.71
Judge Street	1083	22	2.03
Judge Dowdy	182	3	1.65
Judge Emmett	391	5	1.28
Total	4326	267	6.17

Table 5: Melbourne Registry

Judge	Count of orders	Count of in favour of applicant	Percentage (%)
Judge Jones	65	15	23.08
Judge Riethmuller	169	37	21.89
Judge Riley	161	35	21.74
Judge Wilson	91	12	13.19
Judge A Kelly	90	10	11.11
Judge Mcnab	113	9	7.96
Judge Burchardt	85	6	7.06
Judge Hartnett	182	4	2.2
Total	956	128	13.39

F AAT v IAA

The data also reveals a significant difference in the outcomes for judicial review between cases decided by the AAT and IAA. The dataset included a total of 4,726 cases which came from the AAT and the IAA.⁶³ Within that, 3,313 applicants had their merits review decided by the AAT (these are generally applicants who arrived by plane) of which 215 cases (6.5%) were successful in judicial review. There were 1,413 applicants from the IAA (dealing with boat arrivals subject to the fast track assessment process) of which 157 (11.1%) were successful. Tables 6 and 7 provide a breakdown of the total number of cases from the AAT and IAA which each judge has decided and how many were successful.

⁶³ The remaining cases relate to decisions made by the Refugee Review Tribunal, which was amalgamated into the AAT on 1 July 2015.

Table 6: Judicial Review of IAA Decisions

Judge	Total number of cases (IAA)	Success rate of applications from the IAA (%)
Judge A Kelly	15	13.33
Judge Baird	17	5.88
Judge Barnes	17	47.06
Judge Blake	21	4.76
Judge Brown	15	6.67
Judge Cameron	21	14.29
Judge Dowdy	15	6.67
Judge Driver	202	17.82
Judge Egan	72	1.39
Judge Emmett	50	4.00
Judge Hartnett	20	0.00
Judge Heffernan	23	30.43
Judge Howard	1	0.00
Judge Humphreys	64	15.63
Judge Jarrett	45	13.33
Judge Kemp	3	0.00
Judge Kendall	61	16.39
Judge Kirton	8	12.50
Judge Lucev	16	6.25
Judge Manousaridis	57	12.28
Judge McGuire	1	0.00
Judge McNab	14	28.57
Judge Mercuri	15	20.00
Judge Neville	3	0.00
Judge Nicholls	36	0.00
Judge Obradovic	9	33.33
Judge Riethmuller	31	25.81
Judge Riley	20	50.00
Judge Smith	81	17.28
Judge Street	331	2.72
Judge Vasta	73	0.00
Judge Wilson	22	13.64
Judge Young	34	14.71
Total	1,413	11.11

Table 7: Judicial Review of AAT Decisions

Judge	Total number of cases (AAT)	Success rate of applications from the AAT (%)
Judge A Kelly	69	10.14
Judge Baird	27	7.41
Judge Barnes	131	9.92
Judge Blake	19	10.53
Judge Brown	14	0.00
Judge Burchardt	20	0.00
Judge Cameron	83	6.02
Judge Coates	1	0.00
Judge Dowdy	145	1.38
Judge Driver	388	7.22
Judge Egan	42	9.52
Judge Emmett	130	1.54
Judge Harland	35	14.29
Judge Hartnett	114	2.63
Judge Heffernan	52	9.62
Judge Howard	14	7.14
Judge Humphreys	31	0.00
Judge Jarrett	60	1.67
Judge Jones	45	20.00
Judge Kendall	36	13.89
Judge Kirton	18	0.00
Judge Lucev	111	15.32
Judge Manousaridis	215	3.26
Judge McGuire	13	23.08
Judge McNab	94	5.32
Judge Mercuri	37	2.70
Judge Neville	15	20.00
Judge Nicholls	209	3.35
Judge Obradovic	6	16.67
Judge Riethmuller	99	20.20
Judge Riley	76	19.74
Judge Smith	147	12.24

Judge	Total number of cases (AAT)	Success rate of applications from the AAT (%)
Judge Street	615	1.79
Judge Tonkin	1	0.00
Judge Vasta	83	0.00
Judge Willis	1	0.00
Judge Wilson	78	10.26
Judge Wj Neville	1	0.00
Judge Young	38	13.16
Total	3,313	6.49

A chi-square test of independence confirmed that there is a significant association between whether the case came from the AAT or IAA and receiving a positive application outcome, $\chi^2(2, N = 6756) = 31.6, p < 0.001$. The effect size was small (0.0683).

Here it is important to note the discrepancy between our data and similar statistics published by the AAT. The AAT's annual reports include statistics on the number of the decisions which were subject to judicial review and the outcome of those cases. The statistics for the AAT do not distinguish between refugee cases and other migration matters so are not directly comparable. However, the data provided in relation to the IAA demonstrates a much higher success rate at judicial review for refugee applicants than what we found in our dataset. Since 2015, when the first IAA judicial review decision was handed down, the AAT Annual Reports show an overall success rate of 35.92% (see Table 8 below). This is significantly higher than the 11.1% set out in our data.

There are several explanations for this discrepancy. First, as already noted, the published decisions captured in our dataset do not represent the full set of judicial review decisions. According to the AAT annual reports, there had been a total of 2,670 IAA judicial reviews finalised, as compared to the 1,413 in our dataset. A small portion of the gap can be explained by the fact that the AAT's published figures cover the period through to 31 June 2021, while our data cuts off on 11 March 2021. The more significant reason is the fact that our dataset only covers *published* decisions. It does not include cases that were discontinued. It also does not include judicial review applications allowed by consent. It is this latter category which most likely has the biggest impact in terms of the discrepancy in success rates between the two datasets. While the breakdown between applications allowed by judgment and consent is not consistently reported in the annual reports, the 2017–18 report shows that 62 cases were finalised by consent and 26 by judgment. That means as much as 70% of successful judicial review cases are finalised by consent, and their absence from our dataset explains the discrepancy with the data published by the AAT. Two other contributors to the discrepancy are also worth noting. First, the data from the AAT annual reports looks beyond just the Federal Circuit and Family Court, reflecting the ultimate outcome when a case has been appealed to the Federal Court or Full Federal Court. Second, there are

around 240 cases that were set aside where the courts determined the IAA had no jurisdiction to conduct the review following the judgment in *DBB16 v Minister for Immigration and Border Protection*.⁶⁴ **These were likely finalised by consent and not captured in our dataset.**

Table 8: IAA Judicial Review Statistics Drawn from AAT Annual Reports

Year	Applications finalised	Allowed or set aside	Dismissed or discontinued	Success rate of applications (%)
2015–16	1	1	0	100.00
2016–17	53	19	34	36
2017–18	328	100	228	30.49
2018–19	925	449	476	48.50
2019–20	840	232	578	29.60
2020–21	523	158	365	30.20
Total	2,670	959	1681	35.92

Regardless of whether we use the AAT's published statistics or our data, the significant variation between the chances of success at judicial review between decisions of the IAA and AAT are cause for concern. This outcome is significant given that the IAA and fast track assessment process more broadly was specifically designed to limit the procedural and substantive rights of applicants and narrow the grounds available for judicial review. Moreover, it confirms concerns raised about the quality of decision-making at the IAA.⁶⁵ The foundational premise of the fast track system was to speed up asylum processing. The system is clearly failing in this regard, given the high number of cases being remitted back to the IAA for reconsideration. This aligns with the international research which emphasises that the best way to increase the speed and efficiency of processing is to invest in high quality robust refugee status determination procedures.⁶⁶

G Gender of the Judge

We also see a significant degree of variation between the average success rates before male and female judges. In our dataset, 5,447 decisions were made by male judges, with 363 decisions in favour of the applicant. This places the success rate before male judges at 6.7%. As for female judges, 1,309 decisions were made,

64 (2018) 260 FCR 447. The Full Federal Court found that certain people who entered Australia via the Territory of Ashmore and Cartier Islands between 23 January 2002 and 1 June 2013 should not have been processed through the fast track process, and such the IAA review of such decisions was invalid.

65 See, eg, McDonald and O'Sullivan (n 24).

66 Constantin Hruschka and Friedrich Ebert Stiftung, 'The Swiss Asylum Procedure: A Future Model for Europe?' (Q&A, January 2019); Dietrich Thränhardt and Bertelsmann Stiftung, *Speed and Quality: What Germany Can Learn from Switzerland's Asylum Procedure* (Report, 2016).

with 156 of those in favour of the asylum seeker applicant. That made the average success rate before female judges 11.9%. Cases decided by female judges were thus 1.78 times more likely to receive a favourable result.

The statistical significance of this relationship was confirmed by the chi-square test of independence, which showed a significant association between the gender of the judge and receiving a positive application outcome, $\chi^2(1, N = 6756) = 41.1, p < 0.001$. The effect size was small (0.0780).

V USING THE DATA

There are many implications and uses for the data set out above. In this Part we focus on just a few of the data points but intend to explore the further uses and ramifications in future research. Part V(A) focuses on understanding and addressing the significant discrepancy in success rates before individual judges and the potential role of cognitive and social bias in decision-making. In Part V(B) we turn our attention to some other questions which the data raises about the case management and resourcing of the Federal Circuit and Family Court.

A Reducing Bias in Judicial Decision-Making

We argue that some of the discrepancy in judicial review outcomes may potentially be explained by various forms of bias in judicial decision-making. Moreover, we outline some ideas in terms of how statistics of the nature collected in this study could be used to counteract such biases. Some of the variation in success rates in the judicial review of refugee cases can be explained by the respective merits of each case. While we acknowledge again that each case is unique and should be examined on its own merits, it is unlikely that the merits of each case alone can explain the statistically significant discrepancies identified in our data. The allocation of cases to judges through the docket system, should in theory provide a randomised sample of cases. The focus on judges who have heard more than 50 cases further reduces the likelihood of certain judges being allocated a significantly higher proportion of unmeritorious cases.

Moreover, as we discuss further below, there is increasing recognition of the potential influence of cognitive and social biases in judicial decision-making across the board, and significant interest in developing effective strategies and interventions in addressing such biases. These issues were recently examined in the ALRC's 'Review of Judicial Impartiality'.⁶⁷ The review examined whether existing laws governing bias are 'appropriate and sufficient to maintain public confidence in the administration of justice'.⁶⁸ However, the ALRC has noted that this cannot be assessed without reference to other systemic issues that 'impact decision-making and weaken public and litigant confidence' such as the potential

⁶⁷ *Without Fear or Favour* (n 4).

⁶⁸ *Ibid* 21.

for cognitive and social biases, including implicit bias to impact on the impartiality of decision-making.⁶⁹ In this section we examine the utility of our data in identifying and addressing two different types of bias. The first involves social and cognitive bias that influences all forms of human decision-making. The second is the much narrower subset of ‘legally-recognised bias’.⁷⁰

1 Cognitive and Social Biases

The ALRC’s Background Paper on Judicial Impartiality: Cognitive and Social Biases in Judicial Decision-Making (‘Background Paper J16’), sets out research from economists, legal academics, psychologists, and political scientists supporting the position that ‘judicial decision-making, like all human decision-making, is influenced by heuristics (or mental shortcuts), cognitive biases, and other forms of bias’.⁷¹ In this context, cognitive biases refer to ‘systemic tendencies in our thought process that can lead us to error’, and social biases refer to ‘automatically [formed] impressions of people ... based on the social group they are a member of’.⁷²

In some cases, bias may be explicit, in that a person holds certain ‘attitudes and stereotypes that are consciously accessible through introspection *and* endorsed as appropriate’.⁷³ Far more common, however, are implicit biases, which are ‘attitudes and stereotypes that are not consciously accessible through introspection’.⁷⁴ One of the key findings of the scientific research on bias is that biases exist even where decision-makers believe they are operating with impartiality and integrity.⁷⁵ The focus of this section is primarily on these forms of implicit biases.

Several decades of research in cognitive, personality and social psychology research have developed a distinction between two ways in which people think about information when making judgments.⁷⁶ Drawing on this research, Nobel Prize winning Professor Daniel Kahneman identifies two systems in the mind:

- System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control; and
- System 2 allocates attention to effortful mental activities that demand it, including complex computations. The operations of System 2 are

69 Australian Law Reform Commission, ‘Consultation Paper: Judicial Impartiality’ (2021) [76] (‘Judicial Impartiality Consultation Paper’).

70 This term is adapted from Gary Edmond and Kirsty A Martire, ‘Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedures and Decision-Making’ (2019) 82(4) *Modern Law Review* 633, 640 <<https://doi.org/10.1111/1468-2230.12424>>.

71 Australian Law Reform Commission, ‘Judicial Impartiality: Cognitive and Social Biases in Judicial Decision-Making’ (Background Paper J16, April 2021) [6] (‘Background Paper J16’). These issues were examined in further depth in the final report. See *Without Fear or Favour* (n 4) chs 4, 11.

72 Tom Stafford, ‘Biases in Decision Making’ [2017] (Winter) *Tribunals* 19, 19.

73 Jerry Kang et al, ‘Implicit Bias in the Courtroom’ (2012) 59(5) *University of California Los Angeles Law Review* 1124, 1132 (emphasis in original).

74 Ibid.

75 Edmond and Martire (n 70) 645–8.

76 For an overview, see Jonathan Evans and Keith Frankish, *In Two Minds: Dual Processes and Beyond* (Oxford University Press, 2009) <<https://doi.org/10.1093/acprof:oso/9780199230167.001.0001>>.

often associated with the subjective experience of agency, choice and concentration.⁷⁷

In the judicial decision-making context, Irwin and Real adapt this typology to distinguish between ‘blinking’ (heat-of-trial decisions) and ‘staring’ (carefully considered and weighed decisions).⁷⁸ Empirical research on judicial decision-making in the US context shows that judges, like all other humans, rely on a combination of these two systems of thinking in their decision-making.⁷⁹ However, as Judges Wistrich and Rachlinski note, System 1 is the main source of unwanted influences on judicial decision-making.⁸⁰ When engaging in the more intuitive System 1 thinking people rely on mental shortcuts like heuristics.⁸¹ These are ‘rules of thumb’ for solving problems and processing new information.⁸² These shortcuts are an essential part of human decision-making, but can also be influenced by implicit or unconscious biases.

It has been claimed that judges may be better equipped than most to overcome, or ‘compensate for’, the influence of unconscious bias due to the nature of their role and training.⁸³ This is based on a view that their legal training, experience and efforts somehow makes them able to resist the kinds of biases and predispositions that influence the decision-making of ordinary people.⁸⁴ A significant body of empirical research in the US refutes this view. Edmond and Martire draw on a wealth of scientific research to contend that ‘judges are likely to be vulnerable to many, and perhaps all, of the biases that influence ordinary human cognition’.⁸⁵ Similarly, Bradley’s study of American judges concluded that judges have the ‘same cognitive realities of human thought that sustain and plague all of us’, however they tend to appear less likely than others to be able to recognise and acknowledge this.⁸⁶ This has been referred to as the ‘*bias blind spot*’, which is in

77 Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011) 20–1. This work has been referred to by senior Australian judges. See, eg, Justice Stephen Gageler, ‘Why Write Judgments?’ (2014) 36(2) *Sydney Law Review* 189, 197.

78 John Irwin and Daniel Real, ‘Unconscious Influences on Judicial Decision-Making: The Illusion of Objectivity’ (2010) 42(1) *McGeorge Law Review* 1, 2.

79 Andrew Wistrich and Jeffrey Rachlinski, ‘Implicit Bias in Judicial Decision Making: How It Affects Judgment and What Judges Can Do about It’ in Sarah Redfield (ed), *Enhancing Justice: Reducing Bias* (American Bar Association, 2017) 87 <<https://doi.org/10.31228/osf.io/sz5ma>>.

80 Ibid 91.

81 ‘Background Paper J16’ (n 71) [6].

82 Edmond and Martire (n 70) 634 n 2.

83 Jeffrey J Rachlinski et al, ‘Does Unconscious Racial Bias Affect Trial Judges’ (2009) 84(3) *Notre Dame Law Review* 1195, 1195, 1197, 1210, 1221; Brian Barry, *How Judges Judge: Empirical Insights into Judicial Decision-Making* (Routledge, 2021) 174 <<https://doi.org/10.4324/9780429023422>>; ‘Background Paper J16’ (n 71) 12–13.

84 Edmond and Martire (n 70) 633.

85 Ibid 634, citing Redfield (n 79); Emma Cunliffe, ‘Judging Fast and Slow: Using Decision-Making Theory to Explore Judicial Fact Determination’ (2014) 18(2) *International Journal of Evidence and Proof* 139 <<https://doi.org/10.1350/ijep.2014.18.2.447>>; and see generally Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale University Press, 2008).

86 Anna Spain Bradley, ‘The Disruptive Neuroscience of Judicial Choice’ (2018) 9(1) *University of California Irvine Law Review* 1, 7, cited in Matthew Groves, ‘Bias by the Numbers’ (2020) 100 *Australian Institute of Administrative Law Forum* 60, 70.

effect, a bias about biases.⁸⁷ The bias blind spot leads judges to believe they are less susceptible to social and cognitive biases than other people.⁸⁸

As the ALRC has noted, '[recognition] that a judge is human does not mean that they cannot judge impartially. However, it may require additional personal and institutional strategies to remove and disrupt the influence of cognitive and social biases'.⁸⁹ Judicial acceptance of the possible influence of such biases is an important first step, making judges aware of and better able to counteract potential biases.⁹⁰ But this will likely not be enough. Implicit cognitive and social biases are notoriously difficult to counteract.⁹¹ Interventions which involve pre-informing people of the existence of an unconscious bias before asking them to complete a task have been shown to be ineffective,⁹² as have other interventions, including implicit bias training.⁹³ For example, a study of 829 companies over 31 years showed that bias training had no positive effects in the average workplace.⁹⁴

(a) Using Statistical Data as an Auditing Tool

One approach that does show promise in counteracting social and cognitive biases are interventions which encourage individuals to scrutinise their decision-making, thus exposing the more automatic System 1 thinking to the scrutiny of the more analytic and deliberative System 2 thinking. As discussed in the Background Paper J16, System 1 thinking is not adept for decisions requiring conscious deliberation.⁹⁵ If a person is distracted, rushed or tired, or if System 1 and System 2 thinking is in conflict, people tend to rely on System 1 thinking and invoke biases.⁹⁶ By encouraging judges to use System 2 thinking, decision-making is less likely to be affected by implicit biases. These interventions can be described as 'cognitive forcing strategies', which are mechanisms that interrupt reflexive cognitive responses.⁹⁷

87 Edmond and Martire (n 70) 649 (emphasis in original), citing Emily Pronin, Daniel Lin and Lee Ross, 'The Bias Blind Spot: Perceptions of Bias in Self Versus Others' (2002) 28(3) *Personality and Social Psychology Bulletin* 369 <<https://doi.org/10.1177/0146167202286008>>.

88 Edmond and Martire (n 70) 649, citing Richard West, Russel Meserve and Keith Stanovich, 'Cognitive Sophistication Does Not Attenuate the Bias Blind Spot' (2012) 103(3) *Journal of Personality and Social Psychology* 506 <<https://doi.org/10.1037/a0028857>>.

89 'Background Paper J16' (n 71) 4 [2]; Wistrich and Rachlinski (n 79) 104.

90 Justice Keith Mason, 'Unconscious Judicial Prejudice' (2001) 75(11) *Australian Law Journal* 676.

91 See generally Kahneman (n 77).

92 Carol T Kulik, Elissa L Perry and Anne C Bourhis, 'Ironie Evaluation Processes: Effects of Thought Suppression on Evaluations of Older Job Applicants' (2000) 21(6) *Journal of Organizational Behavior* 689 <[https://doi.org/10.1002/1099-1379\(200009\)21:6<689::AID-JOB52>3.0.CO;2-W](https://doi.org/10.1002/1099-1379(200009)21:6<689::AID-JOB52>3.0.CO;2-W)>.

93 Edouard Machery, 'Anomalies in Implicit Attitudes Research' (2021) *WIREs Cognitive Science* (advance); Elizabeth Paluck et al, 'Prejudice Reduction: Progress and Challenges' (2021) 72(1) *Annual Review of Psychology* 533 <<https://doi.org/10.1146/annurev-psych-071620-030619>>; Frank Dobbin, Alexandra Kalev and Erin Kelly, 'Diversity Management in Corporate America' (2007) 6(4) *Contexts* 21, 21–7 <<https://doi.org/10.1525/ctx.2007.6.4.21>>.

94 Dobbin, Kalev and Kelly (n 93) 21–7.

95 'Background Paper J16' (n 71) 6 [12].

96 Ibid 6 [13].

97 Edmond and Martire (n 70) 658.

One of the most effective interventions in encouraging this sort of self-reflection is the use of statistics of the type compiled in this study as a feedback tool for judges – a process known as post decision auditing.⁹⁸ It is very difficult to spot the influence of implicit cognitive and social biases on a single case. However, if similar decisions are logged across time and multiple decision-makers, that data can reveal patterns in decision-making.⁹⁹ The potential for such statistical data to be used to reduce cognitive and social biases in the judicial context has been recognised in multiple studies.¹⁰⁰

The effectiveness of such interventions is through generating ‘soft accountability’ pressures.¹⁰¹ If judges are accountable to explain and justify patterns in their decision-making, this will encourage them to make decisions ‘more carefully and accurately’.¹⁰² As Judges Wistrich and Rachinski note:

Unfortunately, judges operate in an institutional context that provides little prompt and useful feedback. Existing forms of accountability, such as appellate review ... primarily focus on a judge’s performance in a particular case, not on the systemic study of long-term patterns within a judge’s performance that may reveal implicit bias.¹⁰³

The interventions suggested below work by ‘promoting cognitively complex thinking and self-awareness’.¹⁰⁴ Again, Judges Wistrich and Rachniski argue:

Auditing can motivate judges to be more vigilant and thorough in deliberations, lessening their reliance on low-effort mental shortcuts that are often susceptible to unconscious biases. Auditing can also encourage judges to predict counter arguments while making decisions, thus helping them to identify flaws in their informational processing. Awareness of flaws can reduce overconfidence bias – a common tendency to overemphasise belief-affirming information – thus providing the added benefit of improving judges’ self-assessment abilities.¹⁰⁵

These forms of interventions are particularly effective in areas in which judges exercise substantial discretion. Edmond and Martire note, ‘[w]here decisions are open, where there is considerable scope for interpretation, where judges have discretion, this is where unconscious biases are most likely to unwittingly exert their potential discriminatory effects’.¹⁰⁶ This is precisely the type of decision-making which lies at the heart of the judicial review of refugee cases. Such cases turn on identifying whether a jurisdictional error has occurred.¹⁰⁷ While this is not a decision structured by discretion, jurisdictional error is a concept which is notoriously difficult to pin

98 Rachlinski et al (n 83) 1230; Irwin and Real (n 78) 9; Note this is also consistent with the key skills and qualities identified by the National Judicial College of Australia in ‘Attaining Judicial Excellence: A Guide for the NJCA’ (Guide, November 2019).

99 Jerry Kang, ‘What Judges Can Do about Implicit Bias’ (2021) 57(2) *Court Review* 78, 88.

100 Ibid; Kang et al (n 73) 1178; Wistrich and Rachlinski (n 79) 108–19; Chris Guthrie, Jeffrey Rachlinski and Andrew Wistrich, ‘Blinking on the Bench: How Judges Decide Cases’ (2007) 93(1) *Cornell Law Review* 1, 39.

101 Kang (n 99) 89.

102 Ibid.

103 Wistrich and Rachlinski (n 79) 108–9.

104 Ibid 109.

105 Ibid. See also Kang (n 99) 88.

106 Edmond and Martire (n 70) 663.

107 See cases cited at above n 47.

down.¹⁰⁸ As the High Court has noted, it is ‘neither necessary, nor possible, to attempt to mark the metes and bounds of jurisdictional error’.¹⁰⁹

Statistical data on individual decision-making patterns of Federal Circuit and Family Court judges can be used as a feedback tool in a number of different ways. The first approach would see this data used internally by the courts. This could involve simply providing judges with the statistical data and breakdown of past decisions to allow judges to assess trends and influences of cognitive and social biases.¹¹⁰

There are two potential limitations to this approach. Firstly, providing private feedback requires the individual to be self-motivated to address their own biases. Secondly, it opens the opportunity for individuals to engage in cognitive dissonance, a process by which individuals selectively interpret events to support their pre-existing beliefs or attitudes. Research in other areas of decision-making has found that providing feedback on behaviour can result in such self-deception.¹¹¹

The internal use of data by the courts would be more effective when combined with some form of peer review process where judges have an opportunity to account for the outcomes of their decision-making to other respected individuals within their profession.¹¹² There is empirical evidence from the behavioural sciences, that providing feedback on the consequences of behaviour and asking individuals to account for their behaviour to others is effective in countering bias against minority or disadvantaged groups.¹¹³ In the Federal Circuit and Family Court, this could take the form of periodic peer review and mentoring by the Chief Justice of the Court, or a panel of senior judges. This can be accompanied by professional development training targeted towards the identified needs of decision-makers.¹¹⁴

108 Mark Aronson, Matthew Groves and Greg Weeks, *Judicial Review of Administrative Action and Government Liability* (Thomson Reuters, 7th ed, 2021) 13–15 [1.110] (arguing that jurisdictional error has become a label rather than a coherent test). See also Mark Leeming, ‘The Riddle of Jurisdictional Error’ (2014) 38 *Australian Bar Review* 139; Mark Aronson, ‘Jurisdictional Error without the Tears’ in Matthew Groves and HP Lee (eds), *Australian Administrative Law: Fundamentals, Principles and Doctrines* (Cambridge University Press, 2007) 330.

109 *Kirk v Industrial Court (NSW)* (2010) 239 CLR 531, 573 [71] (French CJ, Gummow, Hayne, Crennan, Kiefel and Bell JJ).

110 Irwin and Real (n 78) 9.

111 Michael Auer and Mark Griffiths, ‘Cognitive Dissonance, Personalized Feedback, and Online Gambling Behavior: An Exploratory Study Using Objective Tracking Data and Subjective Self-Report’ (2018) 16 *International Journal of Mental Health and Addiction* 631 <<https://doi.org/10.1007/s11469-017-9808-1>>; Johnny Jermias, ‘Cognitive Dissonance and Resistance to Change: The Influence of Commitment Confirmation and Feedback on Judgment Usefulness of Accounting Systems’ (2001) 26(2) *Accounting, Organizations and Society* 141 <[https://doi.org/10.1016/S0361-3682\(00\)00008-8](https://doi.org/10.1016/S0361-3682(00)00008-8)>.

112 Irwin and Real (n 78) 9; Behavioural Insights Team and Macquarie University, Submission No 29 to the Australian Law Reform Commission Review, *Review into Judicial Impartiality* (30 June 2021).

113 Thomas E Ford et al, ‘The Role of Accountability in Suppressing Managers’ Preinterview Bias against African-American Sales Job Applicants’ (2004) 24(2) *Journal of Personal Selling and Sales Management* 113, 113–24.

114 Note that the National Standard for Professional Development for Australian Judicial Officers proscribes that ‘each judicial officer should be able to spend at least five days per calendar year participating in professional development activities’: National Judicial College of Australia (n 98). See also Australian Law Reform Commission, ‘Judicial Impartiality: Ethics, Professional Development, and Accountability’ (Background Paper J15, April 2021) 12. These interventions are also consistent with principles one and eight of the International Organisation for Judicial Training, ‘Declaration of Judicial Training Principles’

The second approach involves publishing the statistics publicly as we do in this article. A robust body of behavioural psychology research demonstrates that publicly publishing data is a more effective intervention than internal use of data as a feedback tool.¹¹⁵ Moreover, the transparency fostered by such an approach also has the potential to positively promote community trust in the judicial system. While our dataset could be used for this purpose, the influence of such data in countering implicit judicial bias would be much stronger if the data was compiled and used by the Federal Circuit and Family Court itself. The fact that data was scrutinised, collected and published by the Court (rather than academics) would be a more significant form of accountability. At the same time, its impact in increasing public confidence in the judicial system through transparency would be enhanced.

One of the main barriers to compiling such data has been the resources that would be required if this was done manually. However, as our research demonstrates, advances in computational methods, mean that this task can now be automated and undertaken with minimal investment and resources. Another benefit of the data being collected by the Court itself, is the fact that it could capture the full set of cases finalised, and not just those which are captured in published judgments (although depending on the nature of the records kept by the Court, this may require some manual coding).

Such an intervention has applications beyond the judicial review of refugee cases and can potentially be used in any area of judicial decision-making to counter social and cognitive biases generally, or with respect to any specific groups of participants in the law that may be at a higher risk of being impacted by such biases.¹¹⁶ The ALRC recognised the potential for statistics to be used in this way to address cognitive and social biases. One of the core recommendations in the final report of its review of judicial impartiality was that '[t]he Commonwealth courts (individually or jointly) should develop a policy on the creation, development, and use of statistical analysis of judicial decision-making.'¹¹⁷

2 *Legally-Recognised Biases*

There are also important safeguards in the law aimed at ensuring judicial impartiality. These include rules around circumstances in which a judge can be disqualified from hearing a case on the basis of bias. This Part examines the potential role of statistics of the form collected in this study as evidence of, or evidence in support of, such findings of bias. It is important to note that not all forms of cognitive and social biases outlined above will be sufficient to make out a claim of legally-recognised bias. Australia, like many other common law countries,

(Declaration, 2017). The principles recognise the importance of training to judicial independence and call for multidisciplinary training methods within and outside of the law incorporating skills, social context, values and ethics.

115 For a summary of this research, see Behavioural Insights Team and Macquarie University (n 112).

116 See, eg, the proposals to use this form of statistical data to address the influence of social and cognitive biases against First Nations people in judicial decision-making: Behavioural Insights Team and Macquarie University (n 112); National Justice Project, Submission No 44 to the Australian Law Reform Commission, *Review of Judicial Impartiality* (July 2021).

117 *Without Fear or Favour* (n 4) 488–96 (recommendation 13).

recognises two types of bias that may be relied on to disqualify a judge. Actual bias requires proof of the state of mind of the decision-maker that demonstrates they ‘approached the issues with a closed mind or had prejudged them ... [and] could not be swayed by the evidence in the case at hand’.¹¹⁸ In contrast, apprehended bias focuses on the appearances and impressions of how the matter is perceived from the outside. Two steps are required to make out a claim of apprehended bias:

First, it requires the identification of what it is said might lead a judge (or juror) to decide a case other than on its legal and factual merits. The second step ... [requires] an articulation of the logical connection between the matter and the feared deviation from the course of deciding the case on its merits.¹¹⁹

The test is to be assessed from the objective perspective of an imagined ‘fair-minded lay observer’, familiar with key elements of the case. This test is applied by the trial judge themselves when a claim of apprehended bias is raised, or by an appellate court on review.

The objective test for apprehended bias is a lower threshold to meet than for actual bias which requires evidence of the subjective views of the judge in question.¹²⁰ Rather than producing evidence on the state of mind of the relevant decision-maker, apprehended bias focuses on whether there is a risk the public may think they might be biased.¹²¹ Given this lower evidentiary burden, claims for apprehended bias are raised much more often than actual bias.¹²²

Attempts to date to rely on statistical analysis of a judge’s rulings to make out a claim of apprehended bias against judges have been unsuccessful.¹²³ However, given the rise of computational methods to collect and collate such statistics, future challenges relying on statistics are very likely. As Groves observes ‘[claims] based on statistical evidence are inevitable in the information age’.¹²⁴ We argue that two factors may contribute to the success of claims of apprehended bias based on statistics in future cases. The first is the fact that the data collected through computational methods, such as that presented in this article, is far more robust than what has been relied on in earlier unsuccessful cases. The second is the possible future change in approach when applying the apprehended bias test, and in particular the characteristics and knowledge attributed to the ‘lay minded fair observer’.

118 Aronson, Groves and Weeks (n 108) 617 [10.30], citing *Re Medicaments and Related Classes of Goods [No 2]* [2001] 1 WLR 700, 711 [37]–[39] (Lord Phillips MR). Note that these biases may not necessarily be explicit, and unconscious biases may be sufficient: *Sun v Minister for Immigration and Ethnic Affairs* (1997) 81 FCR 71, 135 (North J).

119 *Ebner v Official Trustee* (2000) 205 CLR 337, 345 [8] (Gleeson CJ, McHugh, Gummow and Hayne JJ).

120 But see Michael Kirby, ‘Grounds for Judicial Recusal Differentiating Judicial Impartiality and Judicial Independence’ (2015) 40 *Australian Bar Review* 195 <<https://doi.org/10.4324/9781315849034-17>> where judicial impartiality and independence are distinguished. Former Justice Kirby argues that both impartiality and independence are essential characteristics of a fair trial and must not be subsumed into the one concept.

121 Australian Law Reform Commission, ‘Judicial Impartiality: The Fair-Minded Observer and Its Critics’ (Background Paper J17, April 2021) 4 (‘Background Paper J17’).

122 ‘The Judge, the Public, and the Test for Apprehended Bias’, *Australian Law Reform Commission* (Web Page, 2 June 2021) <<https://www.alrc.gov.au/inquiry/review-of-judicial-impartiality/spotlight-on/judge-public-and-the-test/>>.

123 *ALA15* (n 5); *CMU16* (n 5). See also *Without Fear or Favour* (n 4) 384–6.

124 Groves, ‘Bias by the Numbers’ (n 86) 60.

The Full Federal Court dismissed the attempt to use statistical data to make out a claim of apprehended bias in *ALAI5 v Minister for Immigration and Border Protection* ('*ALAI5*').¹²⁵ The claim dealt with Judge Street, one of the judges in our dataset, focusing on his decision-making in migration cases. The statistical material provided to the Court covered a 6 month period, during which Judge Street had decided 254 migration claims.¹²⁶ The data demonstrated that Judge Street had only decided 0.79% of cases in favour of the applicant, with that figure dropping to 0% for contested cases.¹²⁷ Moreover, all 254 decisions had been delivered *ex tempore*. Figures drawn from the annual reports of the migration tribunals revealed that over a comparable period of time, the average success rates for judicial review cases in the Federal Circuit and Family Court were 10.8% for Migration Review Tribunal decisions and 12.2% for Refugee Review Tribunal decisions.¹²⁸

The reasons provided by Allsop CJ, Kenny and Griffiths JJ in dismissing the claim of apprehended bias fell into two broad categories. The first dealt with the quality and nature of the statistics that were relied on by the applicant. This included concerns that the statistics on the average success rates covered a different period than the statistics provided to the Court in relation to Judge Street's decision-making.¹²⁹ The data drawn from the annual reports covered the years immediately prior to Judge Street's appointment. This was because the updated statistics for the relevant period were not yet made available. Moreover, the Court was concerned that the data from the annual reports was not confined to the outcomes of Federal Circuit and Family Court decisions, but also included appeals to the Full Federal Court and the High Court.¹³⁰ A related concern was that the statistics did not account for the potential impact that of an earlier decision of the Full Federal Court which had criticised and overturned Judge Street's other migration decisions.¹³¹ As Groves notes, '[the] twin assumptions in this finding ... were that the judge's conduct before he was overturned in rather frank terms could not be compared to his conduct afterwards and also that the informed observer would accept such a distinction'.¹³²

The second set of arguments relied on to dismiss the application questioned whether any form of statistical data alone could make out a claim of apprehended bias. While the Court set out a number of discrete arguments against the relevance of statistics,¹³³ they all revolved around the central premise that 'raw statistics are generally likely to be irrelevant to the knowledge and information which is imputed to the hypothetical observer'.¹³⁴ The Court also raised doubts about the

125 *ALAI5* (n 5). For a detailed analysis and critique of the Court's reasoning in *ALAI5*, see Groves, 'Bias by the Numbers' (n 86).

126 *ALAI5* (n 5) [11] (Allsop CJ, Kenny and Griffiths JJ).

127 *Ibid*.

128 *Ibid*. Note that the Migration Review Tribunal and Refugee Tribunal were merged with the Administrative Appeals Tribunal on 1 July 2015.

129 *ALAI5* (n 5) [40] (Allsop CJ, Kenny and Griffiths JJ).

130 *Ibid*. See Part IV(F) above.

131 *Ibid* [44]–[45].

132 Groves, 'Bias by the Numbers' (n 86) 15.

133 These arguments are all examined and critiqued in detail: *ibid*.

134 *ALAI5* (n 5) [43] (Allsop CJ, Kenny and Griffiths JJ).

utility of data comparing one judge with others in the same court in making out a claim of bias.¹³⁵ Moreover, the Court dismissed the utility of statistics alone, without appropriate contextualisation and explanatory analysis,¹³⁶ reasoning that statistics would ‘normally ... need to be accompanied by a relevant analysis’ of the individuals decisions, so that the ‘statistics were placed in a proper context’.¹³⁷ The Full Court stated that this analysis may conclude that many of the decisions were rightly decided. However, it went on to reason that ‘even if some or all of the judgments were wrongly decided’, even that may not be sufficient, as it may be the result of ‘human frailty on the part of the judge’.¹³⁸ This it was argued, would be a ‘consideration which a fair-minded lay observer would take into account’.¹³⁹

In relation to the arguments about the quality of the data, the statistics presented in this article are far more robust and overcome the concerns raised by the Full Federal Court in *ALAI5*. The data provides the possibility of comparing an individual judge’s decision-making with other judges on the Court over any specific time period covered by our dataset. Our data also covers a much broader time period, capturing more than eight years of judgments delivered by the Court (rather than the six months of data relied on in *ALAI5*). Our data is also more focused in terms of the specific nature of the cases it includes. While the statistics relied on in *ALAI5* dealt with all migration decisions, our data focuses specifically on refugee cases. Our data also only captures the outcome of decisions at the Federal Circuit and Family Court, overcoming concerns in relation to the fact that the statistics drawn from the annual reports also included cases overturned upon review at the Full Federal Court and High Court. Our data also provides an opportunity to test the court’s assumption as to whether decisions of appellate courts in overturning and criticising decisions of a primary judge in any way impact their future decision-making. While we have not examined the data in relation to this point in this article, we intend to do so in future publications.

Perhaps most significantly, we also provide regression analysis demonstrating the statistical significance of the data provided. While the potential value of such regression analysis was not directly dealt with in *ALAI5*, it may in part address the Court’s critique in relation to the absence of contextualisation and explanatory analysis.¹⁴⁰ Courts in other common law countries have recognised the potential relevance and importance of regression analysis when relying on statistics in making out claims of apprehended bias. A similar claim of apprehended bias was examined by the Federal Court of Canada in *Turoczi v Canada (Minister of Citizenship and*

135 Ibid [38]. This aspect of the decision was reinforced by a differently constituted Full Federal Court in *CMU16* (n 5). That case dealt with a separate claim of apprehended bias against Judge Street which sought to rely on extracts of his decisions and Full Federal Court rulings which were critical of the judge’s decision-making in migration cases. The Court dismissed this evidence on the basis that earlier rulings of a judge were not admissible in later cases seeking to make out a claim of apprehended bias.

136 *ALAI5* (n 5) [44] (Allsop CJ, Kenny and Griffiths JJ).

137 Ibid [38].

138 Ibid.

139 Ibid.

140 Ibid [43].

Immigration).¹⁴¹ There, the applicant had attempted to rely on statistics in relation to the relative success rates of Immigration Refugee Board members in refugee cases.¹⁴² Justice Zinn found that the statistical data was not enough to make out a claim of apprehended bias. However, his Honour left the door open for the possible success of future cases where the statistics were combined with regression analysis which demonstrated the statistical significance of the member's rejection rate.¹⁴³

Overcoming the Full Federal Court's broader scepticism of statistics alone ever being sufficient to make out apprehended bias claims is more challenging. It would require a shift in the Court's approach to interpreting and implementing the common law test for apprehended bias. The test and its application have been the focus of significant criticism in recent years.¹⁴⁴ At a general level, the jurisprudence has been criticised as setting an inconsistent bar for making out claims of apprehended bias. As Groves notes, on the one hand, the test in *Ebner v Official Trustee* makes it clear that the apprehension of bias should be one of possibility rather than probability.¹⁴⁵ Yet, the courts have often restated the serious nature of a finding of apprehended bias, and the fact that such a finding should not be upheld lightly.¹⁴⁶ More specifically, concerns have been raised in relation to the knowledge and information judges impute to the fair-minded lay observer.¹⁴⁷ Numerous critics have noted the tendency to 'overload' the observer with specialised knowledge about the law and legal traditions,¹⁴⁸ stretching them 'virtually to a snapping point'.¹⁴⁹ As Groves notes, 'when the observer accepts institutional legal practices, as well as the apparently singular

141 [2012] FC 1423 ('*Turoczi*').

142 Sean Rehaag, '2011 Refugee Claim Data and IRB Member Recognition Rates', *Canadian Council for Refugees* (Web Page, 12 March 2012) <<https://ccrweb.ca/en/2011-refugee-claim-data>>. This was a precursor to the more recent studies by this author referred to at above n 3 and accompanying text. Note that Immigration Refugee Board is a merits review body which operates in a similar manner to the AAT. The case can thus be contrasted with *ALA15* (n 5) which dealt with the review of the decision of a Federal Circuit Court judge carrying out judicial review.

143 *Turoczi* (n 141) [15]. Another concern raised by Zinn J related to the allocation of cases at the IRB, and the fact that on the evidence it was unclear whether cases were randomly allocated to judges. This concerns does not apply with respect to our analysis of the Australian Federal Circuit Court, which uses the docket system to randomly allocate cases.

144 See, eg, 'Background Paper J17' (n 121); Matthew Groves, 'Clarity and Complexity in the Bias Rule' (2021) 44(2) *Melbourne University Law Review* 565; Andrew Higgins and Inbar Levy, 'Judicial Policy, Public Perception, and the Science of Decision Making: A New Framework for the Law of Apprehended Bias' (2019) 38(3) *Civil Justice Quarterly* 376.

145 *Ebner v Official Trustee* (2000) 205 CLR 337, 345 [7] (Gleeson CJ, McHugh, Gummow and Hayne JJ), cited in Groves, 'Bias by the Numbers' (n 86) 64 n 32.

146 See, eg, *R v Lusink; Ex parte Shaw* (1980) 32 ALR 47, 50 (Gibbs ACJ), cited in Groves, 'Bias by the Numbers' (n 86) 64 n 31.

147 See *Smits v Roach* (2006) 227 CLR 423, 456–7 [95] (Kirby J). See also *Johnson v Johnson* (2000) 201 CLR 488, 507–9 [52]–[54] (Kirby J) for the different ways in which the fair-minded lay observer has been conceptualised.

148 Groves, 'Bias by the Numbers' (n 86) 65; 'Background Paper J17' (n 121) 9–10; Andrew Higgins and Inbar Levy, 'What the Fair-Minded Observer Really Thinks about Judicial Impartiality' (2021) 84(4) *Modern Law Review* 881 <<https://doi.org/10.1111/1468-2230.12631>>.

149 *Smits v Roach* (2006) 277 CLR 423, 456–7 [95] (Kirby J).

ability of judges to remain impartial, that person is affirming legal traditions, judicial habits and the judges own perceptions of their abilities'.¹⁵⁰

At the same time, the decision in *ALAI5* demonstrates the judicial tendency to dismiss non-legal knowledge and norms as being capable of informing the observer. As already discussed, there is a rich body of literature from the cognitive sciences demonstrating the utility and accuracy of statistical data in demonstrating (and counteracting) cognitive and social biases in human decision-making.¹⁵¹ In light of these developments, it is surprising that the Court was so quick to dismiss the possibility that the informed observer could rely on such statistics when drawing inferences about the impartiality of judges and other officials. In the words of Groves:

The assertion of the Full Federal Court in *ALAI5*, that statistics do not speak for themselves and instead require detailed explanation, is entirely plausible. But sometimes there may be another reason statistics do not speak. It is because they shout – loudly enough for it to be possible that an informed observer *might* consider that those figures necessarily say something. Sometimes statistics are so extreme, so one-sided, that their sheer weight alone *might* say something even in the absence of a detailed analysis of the cases that comprise the statistical set. The difficult question that follows is whether courts can even conceive of that possibility, let alone hear it.¹⁵²

The ALRC's review into judicial impartiality found that the substantive law on actual and apprehended bias did not require amendment. It did, however, flag the potential use of statistics to ground reasonable apprehension of bias as one of the areas 'where further development or clarification through case law would be desirable'.¹⁵³

B Informing Other Reforms

The data set out in our study raises other questions that warrant further investigation and potential reforms in relation to the case management practices and resourcing of the Federal Circuit and Family Court. One potential area of concern is the uneven way in which the judicial review of refugee cases is distributed between judges. As already noted, 6 judges were responsible for 57% of all judgments examined in our dataset, with Judge Street alone handing down 16.5% of all decisions. This figure goes up to 20% when we focus on the period since Judge Street was appointed to the bench (1 January 2015). Judicial specialisation in specific types of cases is common in many courts. It has obvious benefits, including efficiency gains that flow from judges being intimately familiar with specific areas of law. However, such concentration also comes with risks. As has already been discussed at length, there is strong evidence that cognitive and social biases and other preferences, and life experiences of judges can impact judicial

150 Groves, 'Bias by the Numbers' (n 86) 69. Similarly, the Law Council of Australia has stated that 'disregarding a statistical analysis of a judge's decisions for the purposes of assessing actual or apprehended bias may sit uncomfortable with community expectations.'; Law Council of Australia, Submission No 37 to the Australian Law Reform Commission, *Review of Judicial Impartiality* (July 2021).

151 See Part V(A)(1).

152 Groves, 'Bias by the Numbers' (n 86) 78 (emphasis in original).

153 *Without Fear or Favour* (n 4) 32.

decision-making.¹⁵⁴ The larger the pool of judges (and the more diversity there is on the bench),¹⁵⁵ the more likelihood that these various preferences and biases balance each other at a systemic level. However, when decision-making is concentrated in the hands of a very small pool of judges, the biases and predispositions of a handful of judges can significantly tip the scale, raising concerns about the fairness of the system as a whole.

The heavy reliance on *ex tempore* judgments by certain judges exacerbates this issue. The docket system is designed to evenly distribute cases across the bench. However, when a judge decides a large proportion of their cases *ex tempore*, they hear more cases, significantly increasing the proportion of cases they decide as compared to judges who take longer to reach a decision. As discussed, the judges who rely heavily on *ex tempore* decisions, have some of the highest caseloads, and lowest success rates.¹⁵⁶ Research also shows that there is more risk of cognitive and social biases impacting decision-making when judges rely on *ex tempore* oral judgments, rather than written judgments.¹⁵⁷

Statistical data can also be used to inform discussions around the resourcing of the Federal Circuit and Family Court. While our data only represents refugee cases, further research can quantify the total number of cases heard by each judge across all subject matters. The ALRC Background Paper J16 notes that the very high caseload borne by judges in the Federal Circuit and Family Court may impact judges' ability to act impartially and to manage perceptions of impartiality.¹⁵⁸ Numerous other studies demonstrate that time pressures and stress in the context of decision-making leads people to consider alternatives less systematically and completely,¹⁵⁹ and correlate with less accurate decisions.¹⁶⁰

VI CONCLUSION: BOOSTING PUBLIC CONFIDENCE THROUGH TRANSPARENCY

The statistical data presented in this article provides new insights into the way Federal Circuit and Family Court judges decide refugee cases and some variables which may impact their decision-making. As we identify at the outset, the purpose was not to provide an objective measurement of decision-making. Rather, our aim was to identify potential questions and concerns that the data may reveal, as well as to explore the potential of statistical data to improve judicial decision-making and

154 See Part V(A)(1).

155 See below n 160 and accompanying text.

156 See Part IV(A) and (C).

157 'Background Paper J16' (n 71) 15.

158 Ibid 26.

159 Kang (n 99) 84 n 44, citing Giora Keinan, 'Decision-Making under Stress: Scanning of Alternatives under Controllable and Uncontrollable Threats' (1987) 52(3) *Journal of Personality and Social Psychology* 639 <<https://doi.org/10.1037/0022-3514.52.3.639>>.

160 Kang (n 99) 84 n 47, citing Robert Braun, 'The Effective Time Pressure on Auditor Attention to Qualitative Aspects of Misstatements Indicative of Potential Fraudulent Financial Reporting' (2000) 25 *Accounting, Organizations and Society* 243, 255 <[https://doi.org/10.1016/S0361-3682\(99\)00044-6](https://doi.org/10.1016/S0361-3682(99)00044-6)>.

inform potential reforms. While we acknowledge concerns and risks that statistical data may be misinterpreted to undermine confidence in the courts, our intention is the exact opposite. The courts are one of the institutional pillars of our democracy, and we strongly believe that the transparency provided through statistical analysis can enhance public confidence in the judicial system. While our data is a step in that direction, having the courts themselves, collect, use, and publish this data would have an even greater impact in fostering public confidence. We also note however, that using and publishing statistics is not some sort of panacea that will single-handedly transform public perceptions of the courts. This task will also require progressing other reforms recommended by the ALRC's report into judicial impartiality aimed at countering the opacity of the judicial system. This includes actioning long standing proposals calling for more transparency and independence in relation to how judges are selected and appointed,¹⁶¹ as well proposals to create a more transparent and rigorous mechanism for handling complaints in relation to judicial misconduct.¹⁶²

While our focus has been on the judicial review of refugee cases in the Federal Circuit and Family Court of Australia, it is our hope that our study can lay the groundwork for further research across all areas of judicial decision-making. Advances in computational methods for collecting such data is rapidly improving, meaning that we are likely to see much more research of this nature in the future. However, as we have stressed a number of times, the collection and use of this statistical data would have the most utility if done by the courts themselves. Not only would this allow them to control the narrative and provide the required context, but it would also enhance the effectiveness of such data in improving decision-making and public confidence in the judicial system.

161 *Without Fear or Favour* (n 4) 434 (recommendation 7); Australian Law Reform Commission, *Equality Before the Law: Women's Equality* (Report No 69, 21 December 1994) recommended the adoption of a more transparent process for appointing judges to the Federal Judiciary and the promotion of greater diversity in appointment; Mason (n 90).

162 *Without Fear or Favour* (n 4) 310 (recommendation 5); Gabrielle Appleby and Suzanne Le Mire, 'Judicial Conduct: Crafting a System That Enhances Institutional Integrity' (2014) 38(1) *Melbourne University Law Review* 1.