2017

'I Simply do not Believe...': A Case Study of Credibility Determinations in Canadian Refugee Adjudication

Sean Rehaag
Osagoode Hall Law School of York University, srehaag@osgoode.yorku.ca

Follow this and additional works at: http://digitalcommons.osgoode.yorku.ca/scholarly_works

Part of the Immigration Law Commons

Repository Citation
http://digitalcommons.osgoode.yorku.ca/scholarly_works/2625

This Article is brought to you for free and open access by the Faculty Scholarship at Osgoode Digital Commons. It has been accepted for inclusion in Articles & Book Chapters by an authorized administrator of Osgoode Digital Commons.
“I SIMPLY DO NOT BELIEVE...”:
A CASE STUDY OF CREDIBILITY DETERMINATIONS IN CANADIAN REFUGEE ADJUDICATION

*Sean Rehaag

I. INTRODUCTION

Refugee determinations often turn on a single question: Is the refugee claimant telling the truth? While there are other factors that refugee adjudicators must consider, determining whether the claimant’s story is credible remains central to virtually all refugee hearings. In light of the key role credibility assessments play in refugee determinations, scholars are paying increasingly more attention to how refugee adjudicators assess credibility.

This article contributes to the growing body of research on this subject by examining the full caseload of one refugee adjudicator at Canada’s Immigration and Refugee Board (IRB) over a three-year period. That adjudicator, David McBean, denied all the applications for refugee protection he heard during the first three years of his tenure on the IRB. This article examines McBean’s decision-making during this period as a case study to shed light on credibility assessments in Canada’s refugee determination process. The case study draws on quantitative and qualitative data obtained from the IRB using access to information requests. While McBean is, as the study demonstrates, an outlier in terms of the frequency with which he denies refugee applications, his reasoning with regard to credibility is nonetheless instructive.

The article begins by describing the role of credibility in Canada’s refugee determination process, including a discussion of existing scholarship in the area. Next, the article presents the case study, offering a quantitative and qualitative examination of McBean’s refugee determinations from 2008 to 2010, with a focus on how he approaches credibility assessments. The article then considers three sets of implications of the case study and ends with a brief conclusion.

II. CONTEXT: CREDIBILITY DETERMINATIONS

a. Credibility in Canada’s Refugee Determination Process

Canada’s refugee determination system was revised in 2012.1 Because this article involves a case study of refugee determinations made from 2008 to 2010, this section will describe the process that was in place at the time, noting where there are differences under the revised system.

Under both the prior and revised refugee determination systems, the process begins when a person makes a refugee claim to an immigration officer, either at a port of entry or at an

*Sean Rehaag is an Associate Professor at Osgoode Hall Law School. He holds an SJD from the University of Toronto, a LLB & BCL from McGill and a BA(Hon) from the University of British Columbia.

immigration office. The officer interviews the claimant to determine whether the claim is eligible to be referred to the IRB’s Refugee Protection Division (RPD).

Where claims are referred to the RPD, the claimant must submit documentation, including a Personal Information Form (PIF) under the prior system, or a Basis of Claim form under the new system. These forms include biographical information and the claimant’s account of the persecution feared in their home country (often called a “narrative”). In addition, claimants must submit supporting documents. These typically include information related to the claimant (e.g. identity documents, medical reports, police reports, etc.) and information related to conditions in the claimant’s country of origin (e.g. reports by media and human rights organizations). Other documents, such as IRB country condition national documentation packages and notes taken by immigration officers at the eligibility stage are usually also added to the record by the RPD.

Since 1985, refugee claimants in Canada have, as a matter of constitutional law, been entitled to an oral hearing, where they can present their case and respond to any case against them. At these hearings, the claimant’s testimony and the documents on the record are the key evidence considered. RPD hearings are conducted by specialized quasi-judicial administrative decision-makers, known as RPD Members, who are tasked with determining whether claimants meet the refugee definition, as set out in Canada’s immigration legislation. Hearings are, in most circumstances, meant to be non-adversarial.

One of the main purposes of the oral hearing is to facilitate credibility assessments, that is to say, to determine whether the claimant’s account of feared persecution is genuine. Although it is possible to obtain refugee protection notwithstanding negative credibility findings, this is rare. In most cases, if the RPD disbelieves the claimant’s story, the claim fails. However, it is not uncommon for the RPD to believe a claimant’s story and yet still refuse refugee protection.
on the basis that the claim does not meet an aspect of the refugee definition.\textsuperscript{16} While credibility is not, therefore, necessarily determinative, it is nonetheless often a central issue. In fact, according to case law, credibility is always at issue in refugee hearings.\textsuperscript{17}

\textbf{b. How Credibility Decisions are made at the RPD

Credibility determinations are factual findings.\textsuperscript{18} As with other types of factual findings in immigration proceedings, RPD Members have a great deal of flexibility in how they make these findings.\textsuperscript{19} For example, there are no formal rules of evidence,\textsuperscript{20} and courts are generally deferential towards RPD assessments of evidence on judicial review.\textsuperscript{21} Nevertheless, several principles governing credibility assessments can be found in the case law.

According to case law, claimants bear the burden of proof with regard to factual findings,\textsuperscript{22} and the civil standard (i.e. balance of probabilities) is the applicable standard of proof.\textsuperscript{23} At the same time, however, claimants benefit from a presumption of truthfulness, which means that the claimant’s sworn evidence is assumed to be true unless there is a good reason to think otherwise.\textsuperscript{24} Thus, if RPD members disbelieve claimants they must offer clear explanations as to why.\textsuperscript{25} In evaluating these explanations, courts have held that a lack of corroborating evidence is not, on its own, sufficient to justify negative credibility inferences.\textsuperscript{26} Courts have also indicated that sworn evidence cannot be disbelieved merely on the basis that it is self-serving.\textsuperscript{27}

Credibility findings are often based on a variety of factors. For example, the RPD frequently draws negative credibility inferences from unexplained inconsistencies or omissions in the claimant’s evidence, such as internal discrepancies between the claimant’s testimony, the claimant’s narrative, and information collected from the claimant by immigration officers.\textsuperscript{28} The RPD also often draws negative credibility inferences where the claimant’s evidence is

\begin{footnotesize}
\begin{itemize}
\item See e.g. \textit{Canadian Doctors for Refugee Care v Canada (AG),} 2014 FC 651 at paras 842-848, 28 Imm LR (4th) 1.
\item See e.g. \textit{Aymad-Antwi v Canada (Minister of Citizenship and Immigration),} [1995] FCJ No 1116 at para 7, 57 ACWS (3d) 332.
\item \textit{Lin v Canada (Minister of Citizenship and Immigration),} [2008] FC 1052 at para 13, 170 ACWS (3d) 399 [\textit{Lin}]; \textit{Camara v Canada (Minister of Citizenship and Immigration)} 2008 FC 362 at para 12, 167 ACWS (3d) 158 [\textit{Camara}].
\item Jones \& Baglay, \textit{supra} note 7 at 240.
\item IRPA, supra note 2, s 170.
\item Lin, \textit{supra} note 18 at para 13; \textit{Ortiz Garzon v Canada (MCI),} 2011 FC 299 at para 24, 199 ACWS (3d) 563; \textit{Camara, supra} note 18 at para 12; see also \textit{Federal Courts Act, RSC 1985, c F-7, s 18.1(4)(d).}
\item Alam, \textit{supra} note 22, at para 8 (note that the standard of proof for factual findings is not the same thing as the level of risk that must be shown under the refugee definition) (i.e. serious possibility or reasonable grounds); see also IRB, “Evidence”, \textit{supra} note 13 at 3-8.
\item IRB, “Credibility”, \textit{supra} note 24 at 20-22.
\item Sadeghi-Pari v Canada (Minister of Citizenship \& Immigration), 2004 FC 282 at paras 21, 37-38, Imm LR (3d) 150; Ortega Ayala v Canada (Minister of Citizenship \& Immigration), 2011 FC 611 at paras 20-21, 202 ACWS (3d) 561; see also, IRB, “Credibility”, \textit{supra} note 24 at 54-55.
\item IRB, “Credibility”, \textit{supra} note 24 at 67.
\item IRB, “Credibility”, \textit{supra} note 24 at 23-32.
\end{itemize}
\end{footnotesize}
inconsistent with country conditions information or other documents.\textsuperscript{29} Whether the identified inconsistencies are intrinsic or extrinsic, the RPD has been cautioned against engaging in over-vigilant or microscopic analyses of trivial or minor inconsistencies.\textsuperscript{30} The RPD also often makes negative credibility findings based on plausibility. That having been said, courts require that RPD Members offer more than just their subjective impressions of plausibility by requiring them to point to reliable and verifiable evidence against which the claimant’s evidence can be tested.\textsuperscript{31} Another ground the RPD frequently relies upon when making credibility determinations is the demeanour or manner in which the claimant testifies, such as where the claimant’s testimony is vague or hesitant. Generally speaking, though, courts prefer that credibility determinations not rely exclusively on claimant demeanour.\textsuperscript{32}

In addition to finding that specific factual assertions are not credible, the RPD can make general negative credibility findings. This is an overall finding that the claimant’s evidence is not credible. The RPD can make these findings without holding that every single aspect of the claimant’s evidence is untrue as long as the RPD reasonably disbelieves the totality of the evidence. The consequence of general negative credibility findings is that the claimant’s testimony is rejected, though the RPD must still consider any objective evidence that is untainted by the credibility finding.\textsuperscript{33}

Also, where the RPD is of the view that “there was no credible or trustworthy evidence on which it could have made a favourable decision,” the RPD must make a no credible basis (NCB) declaration.\textsuperscript{34} Such declarations limit the procedural rights available to claimants after the negative refugee determination.\textsuperscript{35} It is important to emphasize that NCB declarations and general negative credibility determinations are distinct. Courts have held that the threshold for NCB declarations is high and that the RPD should not routinely make such declarations merely because it concludes that claimants are not credible.\textsuperscript{36} Instead, NCB declarations are restricted to scenarios where there is not any trustworthy or credible evidence—that could possibly support a positive decision.\textsuperscript{37}

\textsuperscript{29} Ibid at 79-81.
\textsuperscript{30} Ibid at 85-86.
\textsuperscript{31} Monteiro v Canada (MCI), 2002 FCT 1258 (CanLII) at paras 14-15, [2002] FCJ No 1720 (QL); Yu v Canada (Citizenship and Immigration), 2015 FC 167 at para 10, [2015] FCJ No 138 (QL); see also, IRB, “Credibility”, supra note 24 at 33-37.
\textsuperscript{32} Rahal v Canada (Minister of Citizenship & Immigration), 2012 FC 319 (CanLII) at para 44, [2012] FCJ No 369 (QL); Gjergo v Canada (Minister of Citizenship & Immigration), 2004 FC 303 at para 22, 131 ACWS (3d) 508; Aguilar Zacarias v Canada (Minister of Citizenship & Immigration), 2012 FC 1155 at para 24, [2012] FCJ No 1252 (QL); see also, IRB, “Credibility”, supra note 24 at 38-40.
\textsuperscript{33} Sheikh v Canada (MEI) [1990] 3 FCR 238 at paras 7-8, 71 DLR (4th) 604; see also, IRB, “Credibility”, supra note 24 at 16.
\textsuperscript{34} IRPA, supra note 2, s 107(2) (note that under Canada’s new refugee determination system there is an additional procedure for declaring claims to be “manifestly unfounded” where claims are “clearly fraudulent”); see IRPA, supra note 2, s 107.1 (the test for the latter, and the distinction between NCB claims and manifestly unfounded claims has not yet been established in the case law).
\textsuperscript{35} See Ibid, s 110(2)(c) (For example, in the new system, NCB declarations prevent claimants from appealing to the IRB’s Refugee Appeal Division); Immigration and Refugee Protection Regulations, SOR/2002-227, s 231 (Under both the new system and prior systems, claimants whose cases are declared to have NCB are not entitled to automatic stays pending judicial review of their RPD decision).
\textsuperscript{36} Rahaman v Canada (Minister of Citizenship and Immigration), 2002 FCA 89 (CanLII) at para 51, [2002] 3 FCR 577 [Rahaman].
\textsuperscript{37} Levario v Canada (Citizenship and Immigration), 2012 FC 314 at para 19, 214 ACWS (3d) 562 [Levario].
c. Academic Commentary on Credibility Determinations

Significant scholarly attention has been paid to how refugee adjudicators make credibility determinations in Canada and elsewhere. One common theme in this literature is that credibility decision-making is poorly understood, including by decision-makers themselves. For example, scholars have pointed out that decision-makers seldom have the ability to accurately detect false testimony empirically verified, which is worrisome considering that most people dramatically overestimate their abilities in this regard.38

Another common theme is that assumptions underlying many credibility assessments may be contradicted by social-scientific evidence. For instance, scholars have cautioned against assumed connections between consistency in telling one’s story and the truthfulness of that account. Such assumptions about consistency are problematized by research on the complexity of memory and autobiographical accounts—complexity that is amplified where the person offering the account suffers from post-traumatic stress disorder or other mental health challenges, as is common in the refugee law context.39 Similarly, scholars have highlighted the dangers posed by refugee adjudicators relying on common sense or intuition to judge whether an account is plausible or credible, especially where the lived-experiences of decision-makers and claimants diverge dramatically.40 Scholars have also critiqued credibility assessments based on claimant demeanour because such assessments have long proven unreliable, particularly in cross-cultural contexts and in scenarios involving trauma.41

An additional theme in the literature on credibility relates to general views of refugees as liars and fraudsters. Some research focuses on this issue at a macro level, exploring how administrative, political and cultural factors have converged to construct a “culture of disbelief” towards refugees in recent years.42 Other research focuses on individual refugee-adjudicators, showing that while many adjudicators strive to approach each case with an open mind, others

40 See e.g., James P Eyster, “Searching for the Key in the Wrong Place: Why ‘Common Sense’ Credibility Rules Consistently Harm Refugees” (2012) 30:1 BU ILJ 1; Macklin, “Truth” supra note 38 at 138-139.
believe that most refugee claimants are liars, and some appear to pride themselves on their perceived ability to impugn the credibility of claimants appearing before them.\footnote{\textsuperscript{43} See e.g. François Crépeau & Dephine Nakache, “Critical spaces in the Canadian Refugee Determination System: 1989-2002” (2008) 20 Int’l J Refugees L 50 at 101-103;} Because of these sorts of concerns, scholars have called for further empirical research on how refugee adjudicators make credibility decisions and for evidence-based recommendations for improving decision-making in this area.\footnote{\textsuperscript{44} See e.g. Robert Thomas, “Assessing the Credibility of Asylum Claims: EU and UK Approaches Examined” (2006) 8 Eur J Migr & L 79 at 96; Herlihy et al, “Assumptions” supra note 39 at 365.} The remainder of this article represents one attempt to respond to this call.

\section*{III. \textbf{CASE STUDY: MCBEAN’S REFUGEE DETERMINATIONS (2008-2010)}}

David McBean is a refugee adjudicator at the IRB who was the subject of media reports when yearly refugee-claim statistics revealed that he denied all the refugee claims he heard during his first three years at the IRB from 2008 to 2010.\footnote{\textsuperscript{45} Nicholas Keung, “Getting asylum the luck of the draw?”, \textit{Toronto Star} (4 March 2011), online: <www.thestar.com>; Nicholas Keung & Jennifer Yang, “Refugee rejections justified, says board”, \textit{Toronto Star} (5 March 2011), online: <www.thestar.com>; Nicholas Keung, “Widowed, wounded, no refuge: Federal judge blasts IRB member with zero acceptance rate”, \textit{Toronto Star} (9 March 2011), online: <www.thestar.com>; “Treat refugees fairly”, Editorial \textit{Toronto Star} (10 March 2011), online: <www.thestar.com>.} This article uses McBean’s decision-making during this period as a case study of credibility assessments in Canada’s refugee determination process.

The case study begins with a quantitative overview of outcomes in McBean’s refugee decisions, compared to outcomes in refugee claims decided by other adjudicators during the same period. The quantitative overview is based on electronic data obtained from the IRB’s internal database through an access to information request.\footnote{\textsuperscript{46} Letter from Eric Villamaire, Director, Access to Information and Privacy to Professor Sean Rehaag (4 April 2014).} This data covers all principal applicant\footnote{\textsuperscript{47} \textit{Refugee Decisions by IRB member David McBean (2008-2010)}, online: <www.yorkspace.library.yorku.ca/xmlui/handle/10315/13675> (thus, the data treats RPD decisions for a family with one principal applicant and one or more associated applicants as a single decision, rather than as multiple decisions). refugee claims decided on the merits\footnote{\textsuperscript{48} \textit{Ibid} (thus, the data cover only RPD decisions resulting in a grant or a refusal of refugee protection, excluding withdrawn, abandoned and otherwise resolved claims).} from 2008 to 2010. There are 39,905 cases in the dataset, and the relevant datapoints provided for each case include: file number, date of the decision, name of the decision-maker,\footnote{\textsuperscript{49} \textit{Ibid} (during the period of the study, most cases were decided by a single adjudicator. However, the IRB occasionally assigned multiple adjudicators for training purposes. In the dataset, only the identity of the first listed adjudicator is considered).} country of origin,\footnote{\textsuperscript{50} \textit{Ibid} (some claimants may have multiple countries of origin. Only the first country of origin listed in the IRB’s database is included in the data).} and outcome.\footnote{\textsuperscript{51} For a general discussion of this methodology and its limitations see Sean Rehaag, “Troubling Patterns in Canadian Refugee Adjudication” (2008) 39:2 Ottawa Law Review 335 [Rehaag, “Troubling”].} Next, the case study undertakes a qualitative assessment of all the written reasons in McBean’s principal applicant refugee determinations decided on the merits from 2008 to 2010. Redacted copies of these
written reasons were obtained from the IRB through an access to information request. This produced a dataset of 174 written reasons, which were digitized and reviewed.

Three points should be made before considering the quantitative and qualitative results of the case study. First, the study’s methodology was selected because most RPD decisions are unpublished and decisions that are selected for publication are not likely to be representative of the RPD’s full caseload. Researchers studying decision-making in this area must go beyond standard legal research methodologies—beyond examining published decisions in legal databases. Obtaining unpublished data directly from the IRB through access to information requests is one way to overcome this methodological challenge.

Second, 2008 to 2010 was selected for the case study because McBean’s zero percent recognition rate during this period attracted media attention. It should be noted, however, that according to more recent IRB data—after the media stories about his recognition rates were published—McBean granted refugee status in nine out of 185 principal applicant claims decided on the merits from 2011 to 2015: one in 2011, one in 2012, three in 2013, one in 2014, and three in 2015.

Third, it may be helpful to consider McBean’s history with the IRB. In 2007, he was appointed to his first three-year term as an RPD Member, effective 15 October 2007. In announcing his appointment, Citizenship and Immigration Canada described his experience as follows: “Prior to his appointment, Mr. McBean was an Operations Service Manager as well as a Refugee Protection Officer with the IRB. He received his Bachelor of Arts from Brandon University and a Bachelor of Laws from Osgoode Hall Law School. He was called to the Ontario Bar in 1991.”

52 Supra note 47 (an access to information request sought copies of all written reasons for decisions issued by McBean in 2008, 2009 and 2010. In response to this request, the IRB provided 1303 records (i.e. pages) containing 174 redacted decisions; ATIP A-2010-00168 (14 June 2011) (one of the decisions (TA9-03220) was included twice. The duplicate was removed from the dataset. At least one of McBean’s decisions appears to have been left out of the IRB’s response to the access to information request. A redacted copy of that decision was obtained from counsel for the claimant in the case and added to the dataset; Richard Addinall, Personal correspondence by email (28 December 2011), on file with the author).


55 Supra note 45 (all sources).

56 Sean Rehaag, “UPDATED 2011 Refugee Claim Data and IRB Member Recognition Rates” (3 August 2012), online: <www.ccrweb.ca/en/2011-updated-refugee-claim-data> (note that data for this year included principal applicants and dependents. Thus, one needs to filter out non-principle applicants from the raw data file available on this website).


Bar in 1995." McBean was later reappointed for a five-year term, effective 15 October 2010. As of the time of writing, that term has expired and he is no longer serving as an RPD Member.

**a. Quantitative Findings**

**i. Recognition rates and NCB Rates**

Table 1 summarizes the outcomes in McBean’s decisions during the period of the study, compared to outcomes in decisions made by all RPD members during the same period.

<table>
<thead>
<tr>
<th>Year</th>
<th>McBean, David</th>
<th>All RPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>2008</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>2009</td>
<td>0</td>
<td>71</td>
</tr>
<tr>
<td>2010</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>2008-2010</td>
<td>0</td>
<td>172</td>
</tr>
</tbody>
</table>

Source: IRB ATIP A-2013-10523 [4 April 2013]

As is evident in the table, McBean refused all of the 172 refugee claims he heard from 2008 to 2010. Refusing all claims heard over a three-year period is extremely unusual. The RPD average recognition rate during this period (51.9 percent) was much higher than McBean’s zero percent recognition rate. In fact, McBean is the only RPD Member deciding ten or more cases during the period of the study who refused all the claims heard.

As is also evident in Table 1, McBean made NCB declarations far more frequently than is typical, doing so in 40.7 percent the cases he decided, whereas the equivalent rate among all RPD Members was only two percent.

**ii. Country of Origin**

In considering McBean’s exceptionally low refugee claim recognition rate and his unusually high NCB rate, one factor that should be taken into consideration is country of origin. During the period of the study, RPD Members specialized in cases from particular countries. Because claimants from different countries may be more or less likely to have well-founded claims, country specialization may understandably affect recognition rates.

Table 2 breaks down the outcomes in McBean’s decisions by country of origin. The table includes only countries for which McBean decided at least five claims. Table 2 also notes average recognition rates and NCB rates at the RPD for the relevant countries during the same period.

---

64 See also Appendix for a table covering all of McBean’s decisions.
As can be seen in Table 2, McBean mostly heard cases from countries with below average recognition rates. One can also see this phenomenon at work using weighted country of origin average recognition rates, through which it is possible to calculate expected recognition rates for individual RPD Members in light of the mix of countries for the cases they decided. McBean’s expected recognition rate from 2008 to 2010 was 17.2 percent. This partly explains why his recognition rate (zero percent) is so much lower than the overall RPD average recognition rate for the same period (51.9 percent). However, this also demonstrates that, although country of origin is a relevant factor, it does not fully account for McBean’s unusually low recognition rates (i.e. it does not explain why McBean’s recognition rate was zero percent when the expected recognition rate was 17.2 percent). The same pattern is evident in Table 2. With the exception of Portugal, for every country of origin for which McBean decided at least five cases, his recognition rate was below the RPD average for that country of origin.

A similar point can be made with respect to the frequency with which McBean makes NCB declarations. For each country of origin from which he heard at least five cases, he was much more likely to make NCB declarations than were his colleagues deciding cases from similar countries. McBean’s NCB rate (40.7 percent) also dramatically exceeds his expected NCB rate (10.1 percent) based on weighted average NCB rates at the RPD for the set of countries in the cases he decided.

---

Weighted country of origin averages are calculated as follows:

\[ n = \text{Number of countries from which McBean decided cases from 2008 to 2010} \]
\[ N_i = \text{Number of cases decided by McBean from country i from 2008 to 2010} \]
\[ R_i = \text{Average RPD recognition rate for country i from 2008 to 2010} \]

\[
\frac{\sum_{i=1}^{n} N_i \times R_i}{\sum_{i=1}^{n} N_i} = \frac{(82 \times 0.127) + (11 \times 0.00) + (8 \times 0.235) + ...}{82 + 11 + 8 + ...} = 0.1718
\]
iii. Explanations for Unexpectedly Low Recognition Rates

In my view, there are three possible explanations for McBean’s unexpectedly low recognition rates and his exceptionally high NCB rates. The first possible explanation is that while McBean decides cases in more or less the same way as other RPD Members, he was randomly assigned an unusually high number of unfounded cases from the particular countries of origin in which he specializes.

This first explanation, while not impossible, is highly improbable. To see why this is the case, we can calculate the probability of McBean making 172 negative decisions, given the mix of countries of origin in the cases he was assigned, and assuming that all cases from any particular country are independent, that RPD members are interchangeable, and that cases from any particular country are randomly assigned to adjudicators specializing in claims from that country. One way to calculate this probability would be to assume that the likelihood of success in any given claim from a particular country is constant at the RPD average recognition rate for that country over the period of the study, irrespective of who serves as the decision-maker. In such circumstances, the probability of a decision-maker deciding eighty-two negative cases and zero positive cases from Mexico from 2008-2010 is approximately one in 70,000.\textsuperscript{66} If one does

\textsuperscript{66} The probability is calculated as follows:

\[
\begin{align*}
\text{n} & = \text{Number of Mexican cases decided by McBean from 2008 to 2010} \\
\text{RPDnum} & = \text{Number of Mexican cases decided by all decision-makers at the RPD from 2008 to 2010} \\
\text{RPDneg} & = \text{Number of Mexican cases refused by all decision-makers at the RPD from 2008 to 2010} \\
\left(\frac{\text{RPDneg}}{\text{RPDnum}}\right)^n & = \left(\frac{5279}{6050}\right)^{82} \\
& = 0.00001397 \\
& = \frac{1}{71,566}
\end{align*}
\]

The above probability calculations were made on a with replacement basis because we assumed that the likelihood of any given case from a particular county succeeding was constant at the RPD average for that country. An alternative approach would be to calculate the probability on a without replacement basis. For such calculations one would make the following assumptions: outcomes in each claim would be exactly the same regardless of the decision-maker (in other words, each claim is properly decided), and within particular countries cases are assigned randomly to decision-makers. On these assumptions, we know exactly how many claims are positive and negative for any given country. We then ask what the probability of a decision-maker being randomly assigned a certain number of negative cases and no positive cases from a pool with a specific number of positive and negative cases from a given country. In McBean’s case, for Mexico we would calculate the probability of being randomly assigned 82 Mexican negative cases and 0 Mexican positive cases from a sample of 6,050 Mexican cases of which 5,279 are negative as follows:

\[
\begin{align*}
& = \frac{5,279}{6,050} \times \frac{5,278}{6,049} \times \frac{5,277}{6,048} \ldots \times \frac{5,198}{5,969} \\
& = 0.00001289
\end{align*}
\]
the same calculations for all the countries from which McBean decided at least five cases, the probability shrinks down to approximately one in thirteen billion.\textsuperscript{67} Moreover, if one includes all 172 cases McBean decided, the probability is extremely small: approximately one in one quadrillion.\textsuperscript{68} The first explanation can, therefore, safely be rejected.

A second explanation would be that McBean decided cases in more or less the same way as other RPD Members, but that the cases he was assigned from particular countries were not randomly selected, and were instead selected on the basis that they were unlikely to be well-founded—or, alternatively, that they were selected on other grounds that are related to their being unlikely to be well-founded. It should be emphasized that this would require selection based on factors beyond country of origin, which, as we have seen, cannot account for McBean’s low recognition rates. If, for example, cases from Mexico were screened on their merits at an initial stage in the refugee determination process, and only claims that appeared prima facie unfounded were assigned to McBean, this might account for his zero percent recognition rate in the eighty-two Mexican cases he heard from 2008 to 2010. This might also account for his above average NCB rate in Mexican cases.

While this second explanation is, in principle, plausible, the IRB has, to my knowledge, never suggested that it uses such a screening process to assign cases (beyond noting that members specialize in particular countries).\textsuperscript{69} Moreover, according to IRB responses to access to information requests, there are no paper or electronic records at the IRB suggesting such a screening process is used either for McBean or more generally.\textsuperscript{70} It would be surprising if the IRB systematically screened cases, and did so in a way that resulted in McBean being assigned only prima facie unfounded cases, but that is not reflected in any public or internal IRB records.\textsuperscript{71} As such, the second explanation should also be rejected.

\[
= \frac{1}{77,600}
\]

In my view, the without replacement assumptions are simpler to understand and easier to defend. However, because they produce lower probabilities, I have decided to err on the side of caution and report the with replacement probabilities.

\textsuperscript{67} That is to say, given the assumptions set out above, the probability (on a with replacement basis) of McBean making all negative decisions in cases from Mexico (5,279/6,050)\textsuperscript{82}, Portugal (33/33)\textsuperscript{11}, Trinidad \& Tobago (101/132)\textsuperscript{8}, Poland (56/86)\textsuperscript{8}, the Philippines (176/207)\textsuperscript{6}, the Czech Republic (111/171)\textsuperscript{6} and Saint Vincent (536/877)\textsuperscript{6}, works out to 1/12,793,871,024. Or, alternatively, on a without replacement basis, this works out to 1/19,282,206,094.

\textsuperscript{68} Or, alternatively, on a without replacement basis, one in 1.7 quadrillion.

\textsuperscript{69} For a discussion of how patterns in case assignment may affect recognition rates, including explanations offered by the IRB, see Rehaag, “Troubling”, supra note 51 at 348-362.

\textsuperscript{70} According to the IRB, there are no records relating to “How cases are assigned to David McBean”; see ATIP A-2012-00176 (22 January 2013). The IRB did identify 125 pages of records in relating to “How cases [are] assigned to RPD members.” ATIP A-2012-00177 (16 April 2013) (none of these records indicates that cases are assigned after any kind of merit screening. The records do include occasional references to particular cases being assigned only to “experienced” RPD members); see e.g., \textit{ibid} at 361 (There is also a recurring phrase “All members may be assigned to all countries, all counsel unless they have specific parameters indicated.” In some places there appear to be parameters listed (typically for the Western region, whereas McBean decided cases from Central region) but the specific information has been redacted); see e.g, \textit{ibid} at 348.

\textsuperscript{71} See Rehaag, “Troubling”, supra note 51 at 348-349 (In other words, the only factor related to patterns in case assignment identified by the RPD Member to explain variations in recognition rates was regional specialization. A similar point can be made about explanatory notes provided by the IRB when it releases data about outcomes in the refugee determination process under access to information requests).
Given the implausibility of the first explanation and the lack of evidence for the second explanation, the most likely explanation is the third and final one. This explanation is that McBean does not decide cases in the same way as other RPD Members. That is to say, he does not apply the same tests and the same standards as his colleagues at the RPD when he decides refugee claims. If this explanation is right, then, from 2008 to 2010, McBean denied at least some claims that would have been granted if other RPD Members heard them. This explanation would account not only for McBean’s lower than expected recognition rates and higher than expected NCB rates, but also for the fact that he, alone among RPD Members deciding ten or more cases from 2008 to 2010, refused refugee status in every single case that he heard.

b. Qualitative Findings

To get a better understanding of why, from 2008 to 2010, McBean’s recognition rates were so much lower than would be expected based on country of origin averages and why his NCB rates were also unexpectedly high, written reasons for all of his decisions during this period were obtained from the IRB through an access to information request. According to my review of the 174 written reasons, McBean denied the applications for refugee protection in every case. He also made NCB declarations in seventy-eight cases (44.8 percent).\(^\text{72}\)

There are several patterns in the reasons McBean deployed denying claims, some of which are especially evident if one examines the frequency of particular credibility determinations types, and if one considers the number of times specific phrases and terms appear in the 174 cases reviewed.\(^\text{73}\)

i. Overview of Credibility Determinations

As can be seen in Table 3, credibility is a key factor in most of McBean’s refugee determinations. He explicitly found claimants to be generally lacking in credibility in 116 cases (66.7%). In thirty-seven cases (21.3%) he denied the claim based on grounds other than credibility without making explicit credibility findings, and yet nonetheless indicated that this should not be taken to mean that he believed the claimants to be credible. To express this holding, he used various versions of the phrase, “Even if I were to accept the claimant’s evidence as true, which I do not necessarily do, the claim fails.” In a further twenty cases (11.5%), McBean did not say anything about credibility and denied the case on other grounds. In only one case (0.6%), McBean explicitly

<table>
<thead>
<tr>
<th>Credibility Determination</th>
<th>Number of Cases</th>
<th>Proportion of Cases (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generally lacking in credibility</td>
<td>116</td>
<td>66.7</td>
</tr>
<tr>
<td>No determination, with caveat</td>
<td>37</td>
<td>21.3</td>
</tr>
<tr>
<td>No determination</td>
<td>20</td>
<td>11.5</td>
</tr>
<tr>
<td>Credible</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>174</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

Source: ATIP A-2010-00168 (14 June 2011)

\(^{72}\) These figures differ slightly from similar statistics calculated based on the quantitative data provided by the IRB. This may reflect different ways of including or excluding cases from datasets. For example, one can calculate decision-dates from the date when a decision was signed by the adjudicator or from the date when the signed decision was processed by the IRB or mailed to the claimant. This may also reflect a small number of data entry errors in the datasets, either at the IRB or in my review of the cases. For example, in the IRB’s database, one case was not recorded as including a NCB declaration, yet such a finding was clearly made in the written reasons; See RPD File NO. TA8-19574 (private proceeding) at para 13, online: <www.yorkspace.library.yorku.ca/xmlui/handle/10315/13675>.

\(^{73}\) The number of times particular terms or phrases appear were calculated based on searches of the digitized reasons for decisions. Digitization was not manually verified, thus it is possible that the terms or phrases in fact appear more often but were not recognized as such due to uncorrected digitization errors.
indicated that he believed the claimant’s account, but nonetheless denied the claim on other
grounds. Interestingly, he was not the sole decision maker in this last case, because a panel of
three RPD members decided it. It is, therefore, apparent that McBean seldom believes the stories
recounted by refugee claimants.

**ii. Grounds for Credibility Determinations**

In the cases where McBean made negative general credibility findings, he often seemed
preoccupied by differences between various versions of the claimant’s evidence. This can be
seen in the phrase that he typically uses to begin his credibility analysis: “It was apparent
throughout the hearing that there were a number of serious discrepancies in the claimant’s
evidence when the oral testimony was compared to the Personal Information Form (PIF) and the
other documents available.” The phrase “It was apparent throughout the hearing that there were a
number of” appears ninety-four times in the 174 cases reviewed. Regardless of whether he used
this particular phrase, in all 116 cases where he found the claimant to be generally lacking in
credibility, he explicitly noted differences between various versions of the claimant’s account.
Often, McBean characterized the differences in the evidence as contradictions or discrepancies.
The term “contradiction” appears 129 times and the terms “discrepancy” and “discrepancies”
appear 228 and 287 times respectively. At other times the differences are characterized as
omissions. The term “omission” appears 244 times. Whether characterized as contradictions,
discrepancies, or omissions, McBean usually identified the differences in claimant evidence by
comparing the claimant’s testimony with documents submitted in support of the claimant’s
application (including the PIF), or with documents prepared by the government, such as notes
taken by immigration officers at the time when claimants first apply for refugee protection. He
placed particularly heavy reliance on the latter; there are 296 references to immigration officer
notes.

Another pattern in McBean’s credibility determinations is that the discrepancies he
identified often relate to specific quantitative details in the claimant’s evidence, including dates,
times and number of assailants. The following passages are typical in this regard:

[The claimant] was asked how many men there were in the car and he said three
came out. When asked if there were more in the car, he said there were a total of
three. When asked why his PIF would indicate that there were four men in the car,
he said that he had not included the driver. I do not find this explanation for this
contradiction to be satisfactory.

In both the PIF and in oral testimony, [XXX] was menaced by a trio of [XXX].
However, as noted at the hearing, in the notes of the Immigration Officer made at
the time that [XXX] made his claim, it is stated that he was menaced by two
people. [XXX] stated that he meant that one person spoke and the other two were
violent and that he thought that he had said that there were three [XXX]. I do not
find [XXX]’s explanations satisfactory.

---

74 RPD File NO TA7-15564 (private proceeding), online: <www.yorkspace.library.yorku.ca/xmlui/handle/10315/13675>.
75 Supra note 47 (see decisions: TA7-10912/12205/12253/122540).
76 Supra note 47 (see decisions: TA7-09899/TA7-09982/TA7-09983/TA8-05827/TA8-05828/TA8-05829).
The claimant stated that his problems started with the family of criminals in [XXX] 2004. However, in the interview with the [Immigration Officer], the claimant stated that his problems started in [XXX] 2003. When confronted with the discrepancy, the claimant stated that the [XXX] 2003 date was incorrect. The claimant could not explain why the date was incorrect other than to say generally that he was nervous during the interview with the [Immigration Officer]. I do not find this explanation satisfactory.\textsuperscript{77}

In oral testimony, the claimant agreed with the statement that he had been kidnapped for 10 days, yet when pressed for the dates he stated from [XXX] through [XXX]. However, as noted at the hearing in the Immigration Officer Interview Notes, he stated that he had been kidnapped for 15 days. The claimant stated that all of the periods were approximate and that he did not want to correct the Refugee Protection Officer (RPO) on her mistaken calculation. I do not find the claimant's explanations satisfactory.\textsuperscript{78}

The claimant stated as per her PIF, that she was assaulted at 8:00 a.m. However, as noted at the hearing, the Public Ministry document stated the incident took place at 10:00 a.m.\textsuperscript{79}

McBean also frequently found that the failure to mention certain details when other less significant details were included in a particular version of a claimant’s story amounted to an omission that undermined the claimant’s credibility. Often he noted that these accounts are otherwise “quite detailed” (a phrase used 26 times in this way). This description is sometimes used even when the account in question is actually rather brief, including one case where the claimant’s narrative was only twelve lines long.\textsuperscript{80}

When claimants attempted to explain contradictions, discrepancies, or omissions by saying that they did not know what level of detail to provide in the documents they submitted, McBean discounted these arguments, often by saying something like: “The directions for filling out the PIF narrative [were] quite clear in that all significant incidents are to be included. I find that this omission further undermines the claimant's credibility.” The phrase “directions for filling out” is used in this way sixty-two times in the 174 cases reviewed. Similarly, when claimants attempted to explain contradictions, discrepancies, or omissions by saying that they deferred to counsel who prepared their supporting documents, McBean rejected these arguments on the grounds that the claimant agreed in writing and at the hearing that the PIF was complete and accurate. The phrase “complete and accurate” is used in this way twenty-two times.

Along similar lines, when claimants tried to explain contradictions, discrepancies, or omissions by pointing to mental health challenges that affect their ability to testify, McBean often dismissed these explanations. In some cases, McBean discounted these arguments because no expert psychological evidence was presented about impairing psychological conditions. For example, in one case he noted, “While counsel asked me to take into account the claimant's

\textsuperscript{77} Supra note 47 (see decisions: TA7-11451).
\textsuperscript{78} Supra note 47 (see decisions: TA7-11051/58/59/60/61).
\textsuperscript{79} Supra note 47 (see decisions: TA9-05027).
\textsuperscript{80} Supra note 47 (see decisions: TA7-11417/18/19).
emotional state of mind when considering the claimant's testimony, no psychological evidence was produced to indicate that the claimant would have any significant difficulty telling his story. The phrase “no psychological evidence was” is used in this way seven times. However, in other cases, he disregarded expert psychological reports because he believed that the claimant duped the medical professionals preparing the reports. For instance, in one case he writes, “I am mindful of the psychological report with respect to the female claimant on file. However, this report is based on a story that I simply do not believe so I give it little weight.”

McBean’s preoccupation with inconsistencies in claimant evidence is not only reflected in the cases reviewed, but he also acknowledged that this is his standard means of assessing credibility. As he puts it in a particularly telling passage from a decision where he did accept expert psychological evidence: “[Due to the claimant's mental disabilities], I totally agree with counsel for the claimant that this was not a typical hearing and that credibility could not simply be assessed by the usual 'Here you said A, there you said B, please explain why the two answers are different' types of questions.”

In addition to this “usual” way that McBean tested credibility, a few other patterns are evident in the cases reviewed. One is that he often found claimants’ testimony to be vague or evasive and drew negative credibility inferences on this basis. The term “vague” appears forty-two times and the term “evasive” appears thirty-four times. Another is that he frequently found that supporting documents provided by the claimant were not genuine. The terms “forged” and “forgery” / “forgeries” appear eighteen times and twenty-eight times respectively. He also often relied on delay in making a refugee claim to infer that the claimant’s fear of persecution is not genuine. The term “delay” appears fifty-eight times.

After noting some or all of these problems regarding the claimant’s credibility, McBean typically concluded his credibility analyses by using a variation of the following passage: “Given the serious discrepancies, omissions and other problems with respect to major issues, I find that the claimant was generally lacking in credibility. I simply do not believe that, on a balance of probabilities, that any of the significant events that the claimant alleged happened [...] actually happened and as such the claim fails.” The phrase “Given the serious” is used in this way eighty-nine times. Similarly, the phrases “I simply do not believe” and “do not believe” appear 140 times and 203 times, respectively.

While McBean frequently indicates that he does not believe claimants, it is also noteworthy that not once in the 174 reviewed decisions does he indicate that sworn testimony offered by claimants was presumed to be true.

81 Supra note 47 (see decisions: TA8-01701).
82 Supra note 47 (see decisions: TA8-09146/TA8-13838).
83 Interestingly, after this passage McBean goes on to rely on inconsistencies to deny the claim: “Small details and big details changed from telling to telling. Dates changed from one decade to another. This was entirely expected, given the psychological evidence on file [...] Even Dr. [XXX] notes that in three years of treatment the claimant has never given a fully consistent history of his past. While it is clear that something happened to the claimant in the past, given his current permanent mental condition, I doubt that we will ever truly know what that something was. I find on a balance of probabilities that all of these incidents mentioned by the claimant are merely manifestations of the claimant's psychiatric condition and that they never actually occurred.”); supra note 47 (see decision: TA6-08970).
84 Emphasis added.
iii. Copied Alternative Grounds for Denying Refugee Claims

In most of McBean’s decisions, credibility was not the sole reason for refusal, and, as noted earlier, credibility was sometimes not a reason for refusal at all. When refusing cases on alternative grounds, McBean frequently copied passages directly from his other decisions, without attribution or any indication that the passages were copied. Often these were lengthy passages about the availability of state protection against particular types of persecution or about the availability of a viable internal flight alternative for particular groups of claimants. At other times the copied passages were brief and were used to refute specific factual allegations or to characterize a specific type of harm as falling outside the scope of refugee protection.

c. Case Study Conclusions

According to the quantitative data obtained from the IRB, McBean denied refugee applications and made NCB declarations much more frequently than would be expected based on country of origin averages from 2008 to 2010. The most likely explanation for these unexpected rates is that he decides cases differently than his colleagues, and that some claimants, who had their claims denied by McBean, would have obtained refugee protection if their cases had been assigned to other adjudicators. This raises serious fairness concerns.

The qualitative data is also troubling. In my view, having reviewed over 1,300 pages of written reasons for decisions by McBean, there are reasons to be concerned about the tests and standards he uses. In particular, frequency with which he simply does not believe stories recounted by claimants is surprising. Similarly, it is worrisome that he only indicated that he generally believed the claimant in one of the 174 cases he decided (and in that case he was sitting on a panel of three RPD members). It is also troubling that he typically offered for holding that claimants generally lack credibility. In particular, he regularly drew negative credibility inferences from relatively minor inconsistencies, and he was very quick to discount explanations for these inconsistencies. He also relied heavily on notes taken by immigration officers at the time that refugee claims were first made—a practice that the Federal Court has warned can lead to unreasonable credibility determinations unless the RPD carefully considers the highly constrained circumstances in which such notes are prepared.

85 Supra note 47 (For example, for lengthy copied passages regarding the availability of state protection against anti-Roma persecution in the Czech Republic, see decisions: TA8-20470/TA8-20518/TA8-20519; TA7-15160/TA7-15188/TA7-15189; TA8-10888/TA8-10948/TA8-10949/TA8-10950; TA8-18021; TA8-19249/TA8-19301/TA8-21528/TA8-21529/TA8-21530).
86 Supra note 47 (For example, for lengthy copied passages regarding Mexico City being a viable internal flight alternatives for sexual minorities facing persecution elsewhere in Mexico, see: TA6-07832; TA6-08970; TA7-03282; TA7-08664; TA7-09246; TA7-11162; TA8-01468; TA8-06422.)
87 Supra note 47 (this passage appears 11 times: “it is not always easy to trace people in Mexico. There is no comprehensive personal database in Mexico, and access to the main existing databases requires a court order and/or written permission from the public prosecutor’s office.”).
88 Supra note 47 (this passage appears 15 times, with slight changes to reflect the number and gender of claimants: “The claimant alleged that he is a victim of crime based on a criminal vendetta. As such, his claim under section 96 of the IRPA fails for lack of nexus to any of the Convention grounds.”).
89 Kanapathipillai v Canada (MCI), [1998] FCJ No 1110 (QL) (FCTD) at para 8; Lubana v Canada (MCI), 2003 FCT 116 (CanLII) at para 13; Triona Aguirre v Canada (MCI), 2008 FC 571 (CanLII) at paras 27-31; Wu v Canada (MCI), 2010 FC 1102 (CanLII) at para 16; Hamdar v Canada (MCI), 2011 FC 382 (CanLII) at paras 44-49; Cao v Canada (MCI), 2012 FC 694 (CanLII) at paras 3-7; Cetinkaya v Canada (MCI), 2012 FC 8 (CanLII) at para 51.
After having read all of McBean’s decisions from 2008 to 2010, it is clear that he takes an adversarial approach to adjudicating refugee claims. His hearings, at least as he describes them in his written decisions, appear to amount to cross-examinations aiming to impugn the claimant’s credibility. In cases that survive this cross-examination, he denies the claims on other grounds. In such cases, it is concerning that he often copies passages from his other decisions without attribution—suggesting that he has standard templates for refusal, raising questions about whether he approaches each case with an open mind and properly turns his mind to the particularities of each case. This practice also raises questions about whether claimants are provided copies of all the materials that are being used against them, and are afforded the opportunity to explain how their cases may differ from the cases that McBean copies from without attribution.

In short, based on this quantitative and qualitative examination of McBean’s decisions from 2008 to 2010, there are serious reasons to be concerned about the frequency with which McBean denies refugee claims, and the reasoning he deploys to justify rejecting these claims.

IV. DISCUSSION

Several implications flow from two basic observations. The first relates to the broader refugee law context, and the second comes from the case study. First, with regard to the broader context, the stakes involved in refugee determinations are high because uncorrected false negatives by definition leave people at risk of deportation to face persecution, torture or death. This has led the Supreme Court of Canada to hold that the constitutional right to life, liberty and security of the person are engaged in the refugee determination process. As a result, the process must comply with principles of procedural justice.

Second, notwithstanding the constitutional requirement that the refugee determination process comply with principles of fundamental justice, there is evidence that outcomes in refugee claims come down to the luck of the draw and hinge in part on who is assigned to hear the claim. One example of such evidence from this case study showing at least some refugee adjudicators demonstrate much more proclivity for refusing claims than others. As a result, they make sustained efforts to find reasons to disbelieve claimants and frequently refuse refugee claims on the basis that claimants are not credible. There are at least three sets of implications that result from combining these observations.

a. Removing Outlier Decision-Makers: Reasonable Apprehension of Bias

One way to think about the luck of the draw in refugee adjudication is to view inconsistent adjudication as an exception from the norm. Such a perspective is reflected in a common reaction to empirical evidence of the kind offered in this case study. This reaction is concerned mostly with the rule of law, which includes the principles that like cases should be decided alike and that outcomes should be based on evidence and law rather than on extraneous

---

90 See supra note 62 (it is worth recalling that as a lawyer and a former Refugee Protection Officer, McBean has training and experience in cross-examinations).
91 Singh, supra note 9.
extra-legal factors.\textsuperscript{93} On this view, if refugee claim outcomes hinge, at least some of the time, on who happens to decide the case, this amounts to a breakdown in the rule of law.

Reasonable apprehension of bias is one legal doctrine that might be deployed to correct such breakdowns in the rule of law. According to this doctrine, a decision can be overturned if a reasonable and properly informed person, having thought through the matter realistically, would conclude that the decision-maker, “whether consciously or unconsciously, would not decide fairly.”\textsuperscript{94} Notice that the focus of this doctrine is not on whether the decision-maker is, in fact, biased. Rather the focus is on whether an informed person might reasonably believe that the decision-maker is biased.

The case law on whether (and if so when) one can infer a reasonable apprehension of bias from evidence that a decision-maker is predisposed towards particular outcomes is decidedly mixed. This is not surprising, as these issues raise difficult questions that go to the heart of adjudication, including: what does it means to be impartial when decision-makers, like all human beings, have views and perspectives that no doubt impact their decision-making—sometimes consciously and sometimes not?\textsuperscript{95}

The Supreme Court of Canada grappled with this question in \textit{R v S (RD)}.\textsuperscript{96} The case involved a racialized young man who was accused of assaulting a white peace officer. The accounts of the incident by the accused and by the peace officer differed substantially. In delivering her reasons for acquitting the accused, the judge in the case drew on her knowledge of her local community and noted that police officers had been known to overreact to non-white groups and to mislead the court. This led the crown to allege a reasonable apprehension of bias on the part of the judge.\textsuperscript{97}

On appeal, the Supreme Court divided sharply—in a variety of different ways—on the question of the test for reasonable apprehension of bias, and on the application the test to the facts of the case. Despite these differences, the justices largely agreed on the principle that, while judges are required to be impartial, they cannot be expected to be neutral. For example, L’Heureux-Dubé and McLachlin JJ noted that a reasonable person expects that “triers of fact will be properly influenced in their deliberations by their individual perspectives on the world in which the events in dispute in the courtroom took place.”\textsuperscript{98} What matters is not whether judges have perspectives and views (they inevitably will), but whether they are able to keep an open mind. In other words, to use Jennifer Nedelsky’s terminology, whether they are able to maintain “the capacity for ‘enlargement of mind’ that makes autonomous, impartial judgment possible.”\textsuperscript{99} Along similar lines, Cory J. noted that “not every favourable or unfavourable disposition attracts the label of bias or prejudice.”\textsuperscript{100} He went on to say, “it is not enough to show that a particular ... [decision-maker] has certain beliefs, opinions or even biases. It must be demonstrated that those

\textsuperscript{93} This principle is regarded as central to the rule of law from a variety of theoretical perspectives; see e.g. Lon Fuller, \textit{The Morality of Law}, rev ed (New Haven: Yale University Press, 1969) at 81-91 (see especially Fuller’s 8th principle of legality); HLA Hart, \textit{The Concept of Law}, 3rd ed (Oxford: Clarendon Press, 1961) at ch 8; John Rawls, \textit{A Theory of Justice} (Cambridge, Mass: Belknap Press, 1971) at 164-165.


\textsuperscript{96} [1997] 3 SCR 484, 151 DLR (4th) 193 \textit{[R v S]}.

\textsuperscript{97} \textit{Ibid}.

\textsuperscript{98} \textit{Ibid} at para 39.

\textsuperscript{99} Nedelsky, \textit{supra note} 96; \textit{R v S}, \textit{supra note} 97 at para 42.

\textsuperscript{100} \textit{R v S}, \textit{supra note} 97 at para 105.
beliefs, opinions or biases prevent the ... [decision-maker] from setting aside any preconceptions and coming to a decision on the basis of the evidence." Or, as the Supreme Court paraphrased in a later case, the "essence of impartiality lies in the requirement of the judge to approach the case to be adjudicated with an open mind."

In this context, the question to be asked is whether the sort of evidence set out in the present study could be the basis of an argument for reasonable apprehension of bias—whether arguments about refugee adjudicators failing to approach cases with an open mind might be used to address challenges that outlier refugee adjudicators may pose to the rule of law.

This argument was partly addressed in Turoczi v Canada (Minister of Citizenship and Immigration), where a refugee claimant asked the Federal Court to overturn a McBean decision due to reasonable apprehension of bias. The applicant introduced evidence about McBean’s zero percent recognition rates and his expected recognition rates in light of weighted country of origin averages. Also introduced into evidence was a newspaper article quoting a former Chair of the IRB who said that a zero percent recognition rate “certainly hints at bias, that this member has an attitude about either particular claimants from particular countries or claimants in general.” Based, on this evidence, the applicant asserted that “in the eyes of the reasonable person ... the number of acceptances by the Board Member is astronomically different from his colleagues [and that] one must be wilfully blind to ... not conclude that there is a reasonable apprehension of bias.

The Federal Court, however, disagreed, finding that the applicant failed to establish a reasonable apprehension of bias because the statistics cited were not “sufficiently informative.” In the court’s view, the statistics merely established different acceptance and rejection rates across decision-makers, without offering sufficient analysis of the potential causes of these differences. As the court put it:

Although the statistical data presented by the applicants may raise an eyebrow for some, the informed reasonable person, thinking the matter through, would demand to know much more, including:

- Were all of the figures, including, importantly, the weighted country origin averages, properly compiled?
- Did the RPD randomly assign cases within each country of origin? If not, how did the RPD assign cases?
- Can factors affecting the randomness of case assignment be reliably adjusted for statistically?
- If so, what are the adjusted statistics, and what is their significance?
- If the RPD did randomly assign cases, what is the statistical significance of the Member’s rejection rate?

101 Ibid at para 107.
103 2012 FC 1423, 224 ACWS (3d) 429 [Turoczi].
104 Ibid at paras 2-4.
105 Ibid at para 16.
106 Ibid at para 5.
108 Ibid at paras 12-19.
Beyond the Member’s relative performance within the RPD, is there anything objective impugning the Member’s decisions (i.e. that suggests they are wrongly decided)?

Accounting for appropriate factors (if that is possible), are the Member’s decisions more frequently quashed on judicial review than would be expected?

Has the Member made recurring errors of a certain type, e.g. on credibility, state protection, etc., that bear a semblance to the impugned decision?

In short, the informed reasonable person, thinking the matter through, would demand a statistical analysis of this data by an expert based upon and having taken into consideration all of the various factors and circumstances that are unique to and impact on determinations of refugee claims before he or she would think it more likely than not that the decision-maker would not render a fair decision.109

In coming to the conclusion that there was no reasonable apprehension of bias, the court also noted that the applicant did not raise any substantive concerns with the decision under review.110 In addition, the court indicated that the initial refugee claim was so weak that McBean was “practically obliged ... to decide as he did ... [A]n informed person, viewing the matter realistically and practically—and having thought the matter through—would conclude that there was very little likelihood that any member would have decided the claim differently.”111

The court was correct in finding that the evidence presented in this case did not meet the threshold for establishing a reasonable apprehension of bias. Allegations of reasonable apprehension of bias based on empirical evidence of inconsistent decision-making should be approached with the same caution and rigour that counsel contemplating constitutional test case litigation would employ. Without a solid evidentiary foundation, including expert analysis, such cases will, and should, fail.

However, two aspects of the court’s reasoning are problematic.112 First, the underlying merits of the refugee claim under review were not relevant to the question of whether there was a reasonable apprehension of bias and should not have been considered by the court. To see why this is the case, consider the following hypothetical. Imagine that a judge hearing a case has a financial interest in the outcome of the case (suppose, for example, the judge is a major shareholder in a corporation being sued). Next, imagine that the outcome would likely have been the same regardless of who serves as the judge (the plaintiff’s legal case against the corporation is weak). In these circumstances, is there a reasonable apprehension of bias? The answer, in my

---

109 Ibid at para 15.
110 Ibid at para 18.
111 Ibid.
112 It is worth noting that when dealing with large datasets, statistical significance may not tell us very much about whether observed differences are meaningful, thus other measures may be preferable; see e.g., Lawrence Kreiger & Kennon Sheldon, “What Makes Lawyers Happy?: A Data-Driven Prescription to Redefine Professional Success” (2015) 83:2 Geo Wash L Rev 554, n 82 (there are also tricky questions to be asked about how to interpret statistical significance when dealing with full data on a population rather than with data for samples of a population).
Despite the weakness of the plaintiff’s case, the plaintiff is nonetheless entitled to have the case heard by a judge who does not have a financial interest in the litigation. Along similar lines, irrespective of the merits of refugee claims, claimants are entitled to have their claims heard by decision-makers who approach cases with open minds.

Second, the Court’s questions about recurring errors in McBean’s decisions and the frequency of his decisions being overturned on judicial review confuse two different matters. These questions relate to the lawfulness of decisions (i.e. whether a decision-maker commits reviewable errors). Reasonable apprehension of bias, however, is not about reviewable errors. Indeed, if a party has to demonstrate reviewable errors by a decision-maker in order to succeed with a reasonable apprehension of bias argument, there would be no point to alleging reasonable apprehension of bias in the first place. Instead, the party could simply ask the Court to overturn the decision based on reviewable errors. Reasonable apprehension of bias is meant to address more than reviewable errors. A decision’s lawfulness does not tell us much about the quality of the decision—just as the fact that an architect’s designs comply with the building code does not tell much about the quality of the architect’s work. In both cases, lawfulness (i.e. no reviewable errors, no violations of the building code) does, of course, matter. But there is much more to quality decision-making (and to good architecture) than lawfulness.

Even if we set aside these conceptual problems, it is not at all clear, on a practical level, how one should interpret the relative frequency of a decision-maker being overturned on judicial review due to reviewable errors. In the refugee law context, decision-makers can easily create barriers to judicial oversight. For example, when decision-makers issue NCB declarations, claimants do not benefit from automatic stays on removal pending judicial review. This means that it is less likely that claimants will apply for or follow through with judicial review. Along similar lines, courts are deferential towards RPD credibility determinations on judicial review, meaning that decision-makers denying claims due to credibility may be overturned less frequently than those who deny claims on other grounds. In fact, given the case with which refugee adjudicators can insulate their decisions from judicial review, one could plausibly assert that a low rate of decisions being overturned on judicial review is as indicative of a problem as a high rate. Such an assertion could be based on the argument that decision-makers in high-stakes refugee adjudication should prepare reasons with the aim of facilitating review—by for example, making their uncertainties explicit, clearly and persuasively articulating the strongest counter-arguments, and attempting to offer a fulsome account of their reasoning. Where, instead,

-----

113 My view accords with the position of the Canadian Judicial Council, which advises that, due to reasonable perceptions of bias,

114 Immigration and Refugee Protection Regulations, SOR/2002-227 (prior to SOR/2012-272 amendments), s 231(2) (prior to SOR/2012-272 amendments) (under the revised system, claims that are declared to have no credible basis or to be manifestly unfounded cannot be appealed to the Refugee Appeal Division and claimants who cannot access RAD do not benefit from automatic stays of removal pending judicial review); see IRPA, supra note 2, s 110(2)(c); IRPR, supra note 35, s 231 (as amended by SOR/2012-272).

115 Supra note 21.
decision-makers aim to craft review-proof decisions, they misunderstand their role in the refugee determination process, adopting the role of an advocate for particular outcomes, rather than that of an administrative adjudicator explaining their decisions.

Another practical difficulty relates to a lack of transparency in Federal Court decision-making that makes it challenging to supplement quantitative data about outcome rates on judicial review with qualitative data. Applications for judicial review of refugee determinations are subject to a leave requirement,\textsuperscript{116} and the Federal Court denies leave in the vast majority of applications—almost always without providing reasons.\textsuperscript{117} As a result, there are relatively few cases through which one could examine written reasons to see how the Federal Court treats particular aspects of an RPD Member’s body of decision-making. Thus, if one wants to examine an RPD Member’s record on judicial review one must rely largely on the kind of bare statistics that the court cautions against in \textit{Turoczi}.

Therefore, the Supreme Court was incorrect in suggesting that evidence of reviewable errors and outcome rates on judicial review are helpful for assessing reasonable apprehension of bias. That being said, because the Court indicated that this information is relevant, data about judicial review of McBean’s RPD decisions was obtained from the IRB and from Federal Court dockets.\textsuperscript{118} Out of 166 McBean decisions from 2008 to 2010 for which data is available, there were eighty-two applications for judicial review. That is to say, 49.4% of McBean’s cases were subject to applications for judicial review. The frequency of applications for judicial review was much lower in McBean’s decisions involving NCB declarations (twenty out of seventy cases, 28.6%) than in his decisions that were denied without such a declaration (sixty-two of ninety-six cases, 64.6%). Of the eighty-two applications for judicial review, thirty-six (43.9%) were not perfected. Of the forty-six applications for judicial review that were perfected, the Federal Court granted leave in eight (17.4% of perfected applications, 9.8% of all applications for judicial review). Finally, of the eight applications where leave was granted, judicial review was granted in three (37.5% of applications where leave was granted, 6.5% of perfected applications for judicial review, and 3.7% of all applications for judicial review). As a point of comparison, from 2005 to 2010 there were 22,914 applications for judicial review brought by unsuccessful refugee claimants. Of these, 4,666 (20.4%) were not perfected. Of the 18,248 that were perfected, 3,250 (17.8% of perfected applications, 14.2% of all applications) were granted leave, and, of those, judicial review was granted in 1,415 cases (43.5% of applications where leave was granted, 7.8% of perfected applications for judicial review, and 6.2% of all applications for judicial review).\textsuperscript{119} In other words, in cases that were subject to applications for judicial review, McBean’s rates of

\textsuperscript{116} IRPA, \textit{supra} note 2, s 72.


\textsuperscript{118} Letter from R Gould to D Younes (18 June 2012); the data obtained from the IRB included 181 RPD decisions where McBean was involved and listed the outcomes in these decisions and, where these decisions were subject to applications for judicial review, the Federal Court docket number. The Federal Court dockets contain information about whether applications were perfected, and about outcomes in those applications. In some cases, documents from the court record were obtained to confirm details from the docket. Of the 181 cases listed in the IRB’s data, 14 were eliminated from the dataset. One claim (TA6-12048) was listed three times so the two extras were eliminated, one claim was linked to an incorrect Federal Court docket number (TA7-10912) so that claim was eliminated, two claims were linked to a single Federal Court docket number (TA7-06791 & TA6-16901) so these were combined, and 10 claims (TA6-02529, TA6-07941, TA6-14902, TA7-02899, TA7-07808, TA7-09551, TA7-12527, TA7-12734, TA7-13170 & TA6-07231) did not appear to be cases where McBean made the final determination so these were excluded. This left a dataset of 166 claims.

\textsuperscript{119} Rehaag, “Luck”, \textit{supra} note 118 at 51.
reversal in perfected applications were broadly similar to the rates for other RPD Members, though claimants were much less likely to perfect applications for judicial review in cases involving McBean than in cases involving other decision-makers.  

Substantively, credibility determinations were raised as an issue in all of the applications for judicial review involving McBean’s decisions where leave was granted. In Arqueta v Canada (Minister of Citizenship and Immigration), the Court found that McBean unreasonably drew negative credibility inferences from a lack of corroborating media accounts, unreasonably discounted medical evidence, unreasonably relied on alleged discrepancies between notes taken by an immigration officer and the PIF / testimony at the hearing, unreasonably found that supporting documents were forgeries, and unreasonably found that there was no corroborating evidence. In Xu v Canada (Minister of Citizenship and Immigration), the Court held that McBean unreasonably found the claimant generally lacking credibility, unreasonably characterized certain statements as contradictory when in fact there was no contradiction, unreasonably rejected plausible explanations for a lack of supporting documents, including media reports, unreasonably relied on differences between notes taken by an immigration officer and the PIF to draw negative credibility inferences, and unreasonably rejected explanations offered by the claimant regarding alleged inconsistencies. In Fernandez Ortega v Canada (Minister of Citizenship and Immigration), the Court found that McBean unreasonably determined a claimant lacked credibility based on minor and inconsequential details of the claimant’s testimony, and unreasonably relied on credibility problems identified with regard to one claimant to draw negative credibility inferences about another claimant. In addition, in one case that was upheld on other grounds, the Court found

---

120 One possible explanation for the lower rates of perfection could relate to legal aid practices regarding when legal aid will be provided to refugee claimants seeking to judicially review negative RPD decisions. If judicial reviews of cases with negative credibility findings are unlikely to succeed before the Federal Court, they are also unlikely to be funded by legal aid programs with merit-screening processes. As such, one might argue that, to the extent that cases with particular types of findings (e.g. negative credibility determinations, no credible basis findings) are less likely to succeed on judicial review, they are also less likely to attract public funding for counsel for judicial review, and, thus, applications for judicial review in these cases would also be less likely to be perfected; see Sean Rehaag, “The Role of Counsel in Canada’s Refugee Determination System: An Empirical Assessment” (2011) 49 Osgoode Hall Law Journal 71 [Rehaag, “Counsel”].

121 2011 FC 1146, 4 Imm LR (4th) 333.
122 Ibid at paras 28-30.
123 Ibid at paras 31-32.
124 Ibid at paras 33-36.
125 Ibid at para 36.
126 Ibid at para 37.
127 Bingrou Xu v Canada (Minister of Citizenship) (4 March 2011), Toronto IMM-4394-10 (FC).
128 Ibid at para 5.
129 Ibid at para 6.
130 Ibid at para 7-8.
131 Ibid at para 9.
132 Ibid at para 10.
133 2009 FC 754, 182 ACWS (3d) 988 (this decision involved two separate claims that were heard together. The court overturned the decision for one claim but upheld the other claim) [Fernandez].
134 Ibid at para 5.
135 Ibid at para 6.
136 Guzman Hernandez v Canada (Minister of Citizenship & Immigration), 2010 FC 953, 197 ACWS (3d) 201.
McBean’s evidentiary determinations to be flawed.\textsuperscript{137} At the same time, however, in the four remaining cases, McBean’s credibility and evidentiary findings were explicitly upheld.\textsuperscript{138}

What, then, should one make of McBean’s record on judicial review? Unfortunately, it would seem that we do not know very much about the Federal Court’s views of McBean’s decision-making from 2008 to 2010. Only a handful of his decisions during this period were fully considered by the Federal Court. Most were not subject to applications for judicial review, and many of those that were subject to applications for judicial review were not perfected. In the cases that were perfected, most were denied leave without any explanation from the court. In this context, it is difficult to determine whether McBean’s average rates of reversal in perfected applications and the above average rates of non-perfection are due to his decisions being generally free from reviewable errors or due to the use of strategies—such as making NCB declarations or general no credibility findings—to craft review-proof decisions and to otherwise reduce the likelihood of successful applications for judicial review.

And what, if anything, should we infer from the small number of applications for judicial review that were fully considered by the court? On the one hand, all of the successful applications for judicial review involved credibility determinations that were rejected by the court in a context where courts are reluctant to interfere with credibility assessments. On the other hand, the court explicitly upheld McBean’s credibility determinations in other cases—presumably rejecting arguments impugning many more of his credibility assessments in cases where leave was denied. In my view, it is difficult to draw meaningful conclusions from all of this, which is one reason why I disagree with the court that this sort of information is important or helpful.

At the end of the day, while I have concerns about some of the further information that the court in \textit{Turoczi} suggests a reasonable person would demand about McBean’s recognition rates, the most interesting feature of the court’s reasoning is not the specific information that the court says a reasonable person would demand. Rather, the most interesting aspect of the court’s decision is that the decision leaves open the possibility of making a reasonable apprehension of bias argument based on expert analysis of patterns in an administrative decision-maker’s cases.

In leaving open the possibility of reasonable apprehension of bias arguments on these grounds, the court in \textit{Turoczi} invites increased engagement with empirical legal scholarship of the kind set out in this article. What makes this especially interesting is that once such engagement begins, it is hard to know where it will end. If \textit{Turoczi} is right, and reasonable apprehension of bias can, in principle, be substantiated using expert evidence that outcomes result from the identity of the adjudicator rather than the merits of the case, then courts can expect litigants to increasingly seek out empirical evidence of inconsistent decision-making. Moreover, this growth in demand for empirical legal research will occur during a period when information about decision-making—and sophisticated new tools to assess such information—is increasingly available in an era of big data.\textsuperscript{139}

\textsuperscript{137} \textit{Ibid} at paras 4-5 (the court held that McBean was not in a position to determine what had in fact happened to the claimant, and that as such it was an “overstatement” for McBean to hold that the incidents recounted by the claimant “never actually occurred.”).


\textsuperscript{139} Consider, for example a recent study of over 150,000 US asylum determinations, which found that an application’s likelihood of success changed depending on whether prior cases decided by the adjudicator were granted or refused – most likely because adjudicators underestimate the probability of streaks, leading them to adjust
My sense is that courts would generally prefer avoiding increased engagement with empirical legal scholarship about patterns of decision-making in the context of reasonable apprehension of bias. Such engagement would force courts to grapple with difficult theoretical questions about the nature of administrative decision-making and adjudication, including the thorny relation between subjectivity, neutrality, and impartiality discussed above in the context of \( R \times S \) (RD). Simply put, the current case law on reasonable apprehension of bias is both unsettled and unsettling; adding empirical legal scholarship to the mix is likely to make this area of law even more difficult. Moreover, such engagement would inevitably spill over from the refugee law and administrative decision-making contexts into other types of decision-making, including adjudication in courts. It would also raise practical challenges, such as where to draw the line in terms of an acceptable degree of subjectivity. Is there, for example, a reasonable apprehension of bias when, due to the identity of the adjudicator, a refugee claimant’s chance of success decreases by one hundred percent, by fifty percent, by twenty percent, by five percent, by one percent, by 0.1 percent?

However, courts cannot simply ignore persuasive empirical evidence where the outcomes in high-stakes refugee adjudication hinge at least partly on the identity of the adjudicator rather than the merits of the claims. This creates an opening, both for courts and for other institutions, to consider mechanisms that could respond to inconsistent decision-making without necessarily relying on reasonable apprehension of bias. These mechanisms will be examined next.

\textit{b. Mitigating the Impact of Inconsistent Decision-Making: Asymmetric Refugee Procedures}

An alternative way to think about the luck of the draw in refugee law decision-making is to view this phenomenon as the norm in adjudication rather than as the exception. Some degree of subjectivity is an unavoidable feature of adjudication. It is not surprising that outcomes in adjudication turn in part on the identity of the adjudicator. Moreover, this subjectivity cannot be mitigated through the doctrine of reasonable apprehension of bias. If adjudication inevitably involves subjectivity, replacing one subjective adjudicator with another subjective adjudicator will not reduce subjectivity in adjudication.

If one begins with the assumption that some degree of subjectivity in adjudication is inevitable, then the key question to ask about high stakes adjudication is not how to eliminate adjudicative subjectivity. Instead, it is necessary to consider how its pernicious effects can be minimized. Asymmetric procedural rules are an especially promising way to do this. Such asymmetries involve making policy choices about what type of outcomes are most important to avoid, and then setting up rules that reduce the likelihood of those outcomes. Thus, for example, in Canadian criminal law, there is a strong policy preference for minimizing false positives (i.e. wrongful convictions) even at the expense of increasing false negatives (i.e. wrongful convictions) to avoid streaks. Daniel Chen, Tobias Moskovitz and Kelly Shue, “Decision-making Under the Gambler’s Fallacy: Evidence from Asylum Judges, Loan Officers and Baseball Umpires” (2016) Fama-Miller Working Paper, available online: <http://ssrn.com/abstract=2538147>. Or consider a similar study of 25,000 US asylum determinations made the day after an NFL team played a game in the adjudicator’s home town, which found that adjudicators were more likely to grant asylum when the local team had won than when the team had lost; see Daniel Chen, Tobias J Moskovitz & Kelly Shue ““Decision-making Under the Gambler’s Fallacy: Evidence from Asylum Judges, Loan Officers and Baseball Umpires” (2014) National Bureau of Economic Research Working Paper No 22026, online: <www.nber.org/papers/w22026>.

\textit{Supra} notes 97-102 (and accompanying text).
Several asymmetric criminal law procedural rules give effect to this policy preference. For instance, the Crown bears the burden of proof, the standard of proof for conviction (beyond a reasonable doubt) is extremely demanding, and convictions require unanimity among jurors. Similarly, the rules on when the Crown can appeal acquittals are more restrictive than the rules regarding individuals seeking to appeal convictions.

There are good reasons for a similar type of policy preference in the refugee law context whereby false negatives (which can be understood from this perspective as refusing refugee protection when most adjudicators would grant protection) should be minimized even if this increases the likelihood of false positives (which can be understood as granting refugee protection when most adjudicators would refuse protection). False-negative refugee determinations may result in refugees being deported to face persecution, torture, or death. These involve high stakes, comparable to the consequences of false convictions in the most serious criminal proceedings. Moreover, false-negative refugee determinations not only have devastating consequences for individual refugees, but they may also leave Canada in breach of international refugee law.

To be sure, false positives may also have important consequences. For example, frequent false positives may undermine the integrity of the Canadian immigration system and may decrease public support for refugee programs. However, these consequences are generally less pressing than the comparable consequences of false acquittals in serious criminal proceedings. Thus, the argument for asymmetric procedures is, if anything, more compelling in the refugee law context than it would be in the criminal law context. Moreover, such a policy preference would be in keeping with the basic objectives of Canadian refugee law, which, according to Canadian immigration legislation, “is in the first instance about saving lives and offering protection to the displaced and persecuted.”

There are many procedures that could minimize false-negative refugee determinations in light of empirical evidence about the luck of the draw in refugee adjudication. However, Canada’s refugee determination system does not currently operate in a manner that minimizes such false negatives. McBean’s decisions from 2008 to 2010 provide a good example of this problem.

Consider, for example, that claimants whose cases were decided by McBean from 2008 to 2010 did not have access to an administrative appeal on the merits to the IRB’s Refugee Appeal Division (RAD). Since 2002, Canada’s immigration legislation has included such an appeal, at least on paper. However, the RAD was not actually implemented until Canada’s refugee determination system was revised in 2012. Additionally, when the appeal was finally implemented, several groups of claimants were barred from accessing the RAD. These bars included claimants from designated countries of origin (DCO bar), claims resulting in NCB declarations (NCB bar), claims declared to be manifestly unfounded (MUC bar), claims by

---

142 Don Stuart, Canadian Criminal Law, 7th ed (Toronto: Thomson Reuters Canada Limited, 2014) at 42ff.
143 Ibid.
146 Of course frequent false negatives could also undermine the public’s confidence in the process.
147 IRPA, supra note 2, s 5(2)(a) (emphasis added).
foreign nationals who transited to Canada via an exception in the Canada-US Safe Third Country Agreement (Safe 3rd bar), and claims by designated foreign nationals (DFN bar).\textsuperscript{148} Because of these limits, most claimants in McBean’s decisions from 2008 to 2010 would not have benefited from the appeal set out in Canada’s revised refugee legislation. In fact, of the 172 McBean cases examined in the quantitative section of this paper, 149 (85.5\%) would have been denied access to the RAD due to the DCO and NCB bars alone.\textsuperscript{149} This reveals a serious problem with the revised refugee legislation: one of the most significant outliers on the negative side in refugee adjudication from 2008 to 2010 would be largely free from oversight in the internal appeal at the RAD if McBean made exactly the same decisions under the revised system. While this problem has been mitigated to some degree by a recent Federal Court case finding the DCO bar unconstitutional,\textsuperscript{150} the problem has not entirely been resolved. That is because even without the DCO bar, seventy (40.7\%) of McBean’s 172 decisions during this period would be covered by the NCB bar.\textsuperscript{151}

Moreover, even if claimants whose cases were denied by McBean from 2008 to 2010 had access to the RAD in its current form, it is not clear how this would have provided much assistance. Due to the level of deference that may be shown by the RAD to first instance RPD credibility determinations, as well as the deference paid to the type of decisions that McBean made, it is unclear what help, if any, the RAD would have provided.

There is a growing body of jurisprudence on how the RAD should approach appeals of RPD decisions. Initially, the RAD applied a highly deferential approach based on the “reasonableness” standard of review as articulated in jurisprudence relating to judicial review of administrative decision-making.\textsuperscript{152} According to this standard of review, if a decision falls within a range of acceptable outcomes and is justified in a transparent and intelligible manner, a reviewing court will uphold the decision, even if the court would have come to a different conclusion had it approached the case on a de novo basis.\textsuperscript{153} More recently, the Federal Court of Appeal in Canada v Huruglica found that this approach is not applicable when the RAD hears appeals from the RPD because the deferential reasonableness standard aims to govern the relationship between the courts and administrative tribunals, whereas the RAD is an appellate body within an administrative tribunal.\textsuperscript{154} Instead, the Federal Court of Appeal found that the RAD should assess RPD decisions on a standard of correctness, where there is no issue of credibility of oral evidence.\textsuperscript{155} The Federal Court of Appeal explicitly left it to the RAD to develop its own jurisprudence on the degree of deference that it should afford the RPD in respect of its credibility assessments, noting that this jurisprudence will need to develop on a case-by-

\textsuperscript{148} Ibid, s 110.
\textsuperscript{149} Of the 172 claims McBean decided, 24 would have been covered by the NCB bar, 79 would have been covered by the DCO bar, and 46 would have been covered by both. The DCO figures are calculated based on countries designated as of 15 October 2014. No data was available on whether any of the 172 claims would have been covered by the safe third bar. However, it is likely that many would be, considering that 23.1\% of claims referred to the IRB in 2013 were covered by this bar; see Grant & Rehaag, supra note 93 at 46-48 (none of the 172 claims would have been subject to the DFN or MUC bars as neither procedure existed at time the claims were decided).
\textsuperscript{151} For further discussion, see Grant & Rehaag, supra note 93 at 261-264.
\textsuperscript{152} See e.g. X (Re), 2013 CanLII 76473 (CA IRB) at paras 10-29.
\textsuperscript{153} Dunsmuir v New Brunswick, 2008 SCC 9 (CanLII) at paras 43-64, [2008] 1 SCR 190.
\textsuperscript{154} Canada (Citizenship and Immigration) v Huruglica, 2016 FCA 93 (CanLII) at para 47, [2016] FCJ No 313.
\textsuperscript{155} Ibid at para 103.
However the Federal Court of Appeal also noted that the level of deference owed in particular cases should recognize “there may be cases where the RPD enjoys a meaningful advantage over the RAD in making findings of fact or mixed fact and law, because they require an assessment of the credibility or weight to be given to the oral evidence it hears.”

While the general findings in Huruglica regarding the correctness standard on matters of law and mixed fact and law other than credibility are a step in the right direction, this last passage is problematic in light of the case study of McBean’s decision-making from 2008 to 2010. Recall that in 66.7% of these cases, McBean found claimants to be generally lacking in credibility. What this suggests is that if the RAD defers to negative RPD credibility determinations, then the RAD cannot serve to minimize false-negative refugee determinations caused by outlier decision-makers who are prone to disbelieving claimants. If RAD is to serve as a meaningful form of oversight to correct false-negative refugee decisions caused by such outlier decision-makers, it is essential that the RAD avoid overly deferential standards of review for negative RPD credibility determinations.

A similar point can be made about judicial review. If claimants either cannot access the RAD or if the RAD does not meaningfully reconsider negative credibility determinations, then the only recourse available to correct false-negative refugee determinations resulting from this sort of subjective first instance decision-making is judicial review in Federal Court. McBean’s decision-making from 2008 to 2010, however, reveals problems with the judicial review process. Out of 166 McBean decisions during this period for which data is available, only eight were fully reviewed by the Federal Court (4.8%). Moreover, when these decisions were considered, the Federal Court adopted a deferential standard of review towards credibility findings (reasonableness); as was previously noted, 66.7% of McBean’s decisions involved negative credibility determinations. Taken together, this means that Federal Court judicial review did not, in the vast majority of cases, meaningfully reduce the risks of false-negative refugee determinations caused by McBean’s outlier decision-making from 2008 to 2010.

The current refugee determination system does not do a good job of limiting the most pernicious aspects of subjectivity in refugee adjudication (i.e. false-negative determinations). However, it would be a simple matter to bring in revisions that would dramatically enhance that system.

One option would be to return to the pre-2002 practice whereby refugee claims were heard by panels of two adjudicators, with differences in refugee claim outcomes between the adjudicators being resolved in favor of the claimant. Such a practice would mean that claims could not be denied merely because of a single outlier decision-maker (though they could be granted by a single outlier decision-maker). Moreover, this practice could actively mitigate the harm caused by outlier decision-makers by matching them with more typical decision-makers.

---

156 Ibid at para 74.
157 Ibid at para 70.
158 There is cause for concern in this regard in a very recent post-Huruglica Federal Court decision, which upheld RAD’s deference to RPD credibility determinations; see Gebremichael v Canada (Citizenship and Immigration), 2016 FC 646 at para 14, 267 ACWS (3d) 670.
159 See Fernandez, supra note 134 at para 3.
161 François Crépeau & Delphine Nakache, “Critical Spaces in the Canadian Refugee Determination System: 1989–2002” (2008) 20:1 Int J Refugee Law 50 at 85 (there is some evidence that this was precisely how the two-panel-
Another option would be to rework the new appeal at the RAD to achieve similar effects. For example, all claimants refused by the RPD could be provided with access to the RAD to ensure that there is at least some substantive oversight of negative first instance RPD decisions. Additionally, there could be restrictions preventing the government from appealing positive RPD decisions. If the government cannot appeal positive decisions, then the RAD could review RPD decisions on an entirely de novo basis that would show no deference to factual and credibility findings made by the RPD. This would effectively replicate the pre-2002 practice whereby two decision-makers would need to agree to deny a claim, whereas a single decision-maker could grant a claim. Alternatively, if both claimants and the government could appeal RPD decisions, then a more complex standard of intervention would need to be applied. That is because if both positive and negative RPD decisions could be appealed and if the RAD heard appeals on an entirely de novo basis then the system would have simply replicated the subjective adjudication/false-negative problem at a higher level. Instead of refugee claims hinging on the luck of the draw at the RPD, they would hinge on the luck of the draw at the RAD.

If the RAD were accessible both to claimants and to the government then an asymmetric standard of intervention is needed. One possible asymmetric standard would be for the RAD to adopt a deferential approach to factual and credibility findings that assist the claimant, while adopting a non-deferential approach to factual and credibility findings that disadvantage claimants. Again, such a policy would essentially replicate the pre-2002 practice of having two decision-makers hear claims, with differences between the decision-makers being resolved in favour of the claimant.

These are only a few examples of possible policies that could mitigate the harms of subjectivity in refugee adjudication. The point here is not that any particular policy should be adopted. Rather, it is to argue that, if one begins with the assumption that inconsistent refugee law decision-making is inevitable in light of unavoidable subjectivity in adjudication (or if one is not prepared to remove outlier decision-makers through the doctrine of reasonable apprehension of bias), then asymmetric procedures should be used to minimize the most serious impact of the luck of the draw in high stakes refugee adjudication.\footnote{See also Hilary Evans Cameron, \textit{When in Doubt: The Law of Fact-Finding in Canadian Refugee Status} (SJD Thesis, University of Toronto, 2016) [unpublished] (for a more extended argument about the need for asymmetric procedures – that is for preferring mechanisms that limit false negatives even at the expense of increasing false positives).}

\textit{c. Taking Responsibility for Inconsistent Decision-Making: Good-Faith Decision-Makers}

A third way of thinking about the luck of the draw in refugee adjudication is to ask what all this looks like from the perspective of decision-makers, and specifically, from the perspective of decision-makers who are trying their best to make fair and consistent refugee decisions.

Decision-makers seeking guidance in this area can look to a number of sources. These include Federal Court case law, formal instruments used by the IRB to enhance consistency, and informal mechanisms such as IRB training.\footnote{For a discussion of means that the IRB has explored to enhance consistency in refugee adjudication, see Rehaag, “Troubling”, \textit{supra} note 51 at 340.} All these mechanisms can be helpful in enhancing consistency by encouraging decision-makers to follow similar processes and to develop and member panels were used by the IRB and it is also worth recalling that the only case from 2008 to 2010 in which McBean explicitly found a claimant to be credible involved a panel of three members).
apply shared understandings of the refugee definition. In some cases, these mechanisms can also usefully highlight problematic practices in credibility assessments. For example, documents prepared by the IRB Legal Service do a good job summarizing the case law on credibility determinations and offering helpful reminders about the many ways that credibility determinations can go wrong.\textsuperscript{164} Along similar lines, training from experts on sexual minority refugee determinations has helped reduce (though not eliminate) some of the more egregious reliance on LGBTQ stereotypes when the IRB assesses the veracity of a minority refugee claimant’s asserted sexual orientation.\textsuperscript{165}

What none of these mechanisms really do, however, is provide clear guidance about how decision-makers can reliably determine when refugee claimants are telling the truth. The reason for this is straightforward: There is simply no way to do this. That is, there are no procedures, no instructions, and no formulas that would allow decision-makers—certainly those confronted with the highly constrained refugee determination context—to reliably detect truth and falsehood.\textsuperscript{166}

It would be understandable if, from the perspective of decision-makers who are called upon to make credibility determinations, such an observation seems unsatisfying. However, this observation has an important implication that, if taken seriously, can improve refugee decision-making. Indeed, this observation can help identify not only the main problem evident in the case study of McBean’s decision-making from 2008 to 2010, but also the most promising solution.

In finding this solution, a proper starting point can be found in the work of Audrey Macklin.\textsuperscript{167} Reflecting on her experience as an IRB Member, Macklin examines how refugee decision-makers who want to make good quality decisions should approach credibility determinations.\textsuperscript{168} She runs through the various means typically used by IRB Members to assess credibility—including consistency, plausibility and demeanour—and argues that none of these allow adjudicators to effectively distinguish truth from falsehood.\textsuperscript{169} This leads her to suggest that, “[m]y tenure as a decision maker has taught me that if there is an objective reality out there, I’m unlikely to discern it except incidentally.” In other words, there is, according to Macklin, no mechanism that refugee adjudicators could use to reliably and consistently assess credibility.

However, Macklin does not stop there. Instead, she suggests that adjudicators have something to gain once they realize “that it is a mistake to overestimate our own capacity to distinguish truth from falsehood, fact from fiction.”\textsuperscript{170} This realization leads to the insight that “credibility determination is not about ‘discovering’ truth. It is, rather, about making choices—what to accept, what to reject, how much to believe, where to draw the line—in the face of

\textsuperscript{164} See also Immigration and Refugee Board, \textit{Assessment of Credibility in Claims for Refugee Protection}, 31 January 2004 (Ottawa: Refugee Protection Division, 2004), online: \textless http://www.irb-cisr.gc.ca/Eng/BoaCom/references/LegJur/Documents/Credibe.pdf\textgreater .


\textsuperscript{166} Hilary Evans Cameron, “Substantial deference and tribunal expertise post-Dunsmuir: A new approach to reasonableness review” (2014) 27:1 Can J Admin L & Practice 1 at 17-18 (for a similar argument, including references to social scientific evidence to bolster this argument).

\textsuperscript{167} Macklin, “Truth” supra note 38 at 134.

\textsuperscript{168} Ibid.

\textsuperscript{169} Ibid at 137-139.

\textsuperscript{170} Ibid at 139.

\textsuperscript{171} Ibid at 140.
empirical uncertainty." If credibility determinations are about choices, then refugee adjudicators who want to make good quality decisions need to be thoughtful about these choices and to examine why they make particular choices. As Macklin puts it, making good quality credibility determinations "requires [of decision-makers] that we do not confine ourselves to interrogating the claimant, but that we also interrogate ourselves."

If one examines McBean's decision-making from 2008 to 2010 through this lens, then the biggest problem is not just that McBean is much more likely than his colleagues to deny refugee claims. Nor is it that he so frequently disbelieves claimants. Rather, it is that the reasons he offered for disbelieving claimants do not seem to have much to do with the actual reasons why he found claimants not to be credible. If one wanted to know why a claimant who appears before McBean is found not to be credible one has to go beyond the reasons offered. That is because, as the case study has shown, it is likely that, on the exact same facts, at least some of the claimants who were disbelieved by McBean would have been believed by at least some other refugee adjudicators. In other words, the claimant's credibility is not something that can be attributed merely to the claimant; adjudicators also play a key role. If adjudicators play a key role, but they write reasons that do not say anything about that role, then the reasons provided do not reflect the actual reasons for the decision. In Macklin's terms, the reasons do not interrogate or explain the adjudicator's choices. Rather, the reasons obscure those choices.

This also points to the most promising solution to the problems, which is evident in McBean's decisions from 2008 to 2010: Decision-makers should not merely ask themselves whether they have been persuaded by the claimant's evidence. They should also ask themselves whether any of their colleagues would reasonably have been persuaded by the claimant's evidence, including colleagues who are much more likely to grant refugee protection. If a decision-maker personally disbelieves the claimant but is nonetheless of the view that at least one colleague would reasonably believe the claimant then, absent unusual circumstances, the decision-maker should take the claimant's evidence to be true. And if there are unusual circumstances, such that the decision-maker wants to find the claimant's evidence not to be credible despite acknowledging that another decision-maker would reasonably believe that evidence, then the decision-maker should explain why.

Such an approach would accomplish two things simultaneously. First, it would force decision-makers to articulate and interrogate their choices about "where to draw the line ... in the face of empirical uncertainty." If adjudicators have to explain why their judgment differs from that of their colleagues, it would ensure that, at a minimum, these differences are better understood—and in doing so, this would facilitate oversight either on appeal at the RAD or on judicial review in the Federal Court. Second, it would make it more difficult for decision-makers to deny claims based on credibility. Such denials would generally only occur in the clearest cases. This would reduce the luck of the draw and increase fairness, while limiting the likelihood of false-negative decisions, which, as we saw in the prior section, is a key way to mitigate the most pernicious effects of the luck of the draw in refugee adjudication.

172 Ibid.
173 See generally Michael Kagan, "Is Truth in the Eye of the Beholder? Objective Credibility Assessment in Refugee Status Determination" (2003) 17:3 Geo Immigr LJ 357-415 (this could be understood as an "objective" rather than "subjective" test for credibility. That is, the adjudicator is not asking whether she or he believes the claimant, but whether there are reasonable grounds to believe the claimant in a context where claimants are presumed to be telling the truth and in which evidentiary findings are incredibly fraught).
174 Macklin, "Truth" supra note 38 at 140.
This approach essentially requires humility and self-reflection among refugee adjudicators. It requires that adjudicators approach their task not with the aim of preparing written reasons that will withstand legal challenge (i.e. not to merely set out the legal authority for their decision) but rather to make good-faith efforts to explain their decisions, including the role that they as decision-makers play in their decisions, as well as to facilitate appeals or reviews of their decisions. This is admittedly quite demanding—particularly in a context where decision-makers are under enormous pressure to decide large numbers of cases quickly. This is what is required of good-faith decision-makers who seriously consider research related to the luck of the draw—including the findings of this case study—and the incredibly high stakes of the decisions they are being asked to make.

V. CONCLUSION

This article seeks to contribute to the growing body of scholarship about credibility determinations in refugee adjudication. It does so through the case study of a single outlier decision-maker at the Canadian Immigration and Refugee Board. The case study shows that, over a three-year period, this decision-maker was much more likely than his colleagues to deny refugee claims from similar countries. The study also demonstrates that this decision-maker frequently found that refugee claimants who appeared before him were liars and frauds. He simply did not believe them.

The article also sets out several implications of the study, including in the areas of reasonable apprehension of bias, institutional design, and approaches to credibility decision-making. But perhaps the most important implication of the study is this: More research is needed for how refugee adjudicators make credibility determinations, particularly research that engages from a variety of perspectives with social scientific methods, and research that offers practical suggestions using insights drawn from those methods. As was argued, credibility determinations are the key to many, if not most, refugee decisions. Those who conduct research on refugee law issues, as well as those who are involved in refugee decision-making as adjudicators, administrators or policy makers, need a much better understanding of how such determinations are, can, and should be made.

Finally, I would like to conclude on a personal note. I struggled with writing this article, which has been percolating for several years now. The challenge for me was not in conducting the research. Rather, the challenge was in thinking through what I wanted to learn from the research—why some possible lessons mattered more to me than others. I wondered, as I was writing, whether I simply believe refugee claimants, and, if so, whether that is as much of a problem as simply not believing them. I wondered whether the research, which uses the language and methods of empirical legal studies, might be advocacy for refugees in disguise (and if so, whether this is a problem). I wondered whether quantitative research of this kind oversimplifies complex social phenomena. I wondered whether latching onto inconsistencies in refugee determination statistics is like being preoccupied with inconsistencies between dates in a claimant’s testimony and whether such preoccupations may do more to obscure than to clarify. And I wondered whether there was something unfair about using public resources and a privileged position as a tenured law professor to publish a critique of a (now former) refugee adjudicator.

I continue to wonder about all of this. But, ultimately, I think the best answer to these sorts of questions is to set out the research as fairly as I can and then to hope that others take up
my call to conduct further research, including research that might expose some of the limits of this case study and of the implications that I have sought to draw from it.

VI. APPENDIX

<table>
<thead>
<tr>
<th>Country</th>
<th>McBean</th>
<th>All RPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Mexico</td>
<td>0</td>
<td>82</td>
</tr>
<tr>
<td>Portugal</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Poland</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Philippines</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Saint Vincent</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>El Salvador</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Germany</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Guyana</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Italy</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Grenada</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Guatemala</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Saint Lucia</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Hungary</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Peru</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Sweden</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Barbados</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Belize</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Bolivia</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Brazil</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dominica</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>France</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ireland</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Jamaica</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Norway</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Spain</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>USA</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Venezuela</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: IRB ATIP A-2013-10523 (4 April 2013)

* Only countries from which McBean decided at least one case