BERKELEY JOURNAL OF INTERNATIONAL LAW

VOLUME 30 2012 NUMBER 1

CONTENTS

Articles

THE REMOVABILITY OF NON-CITIZEN PARENTS AND THE BEST INTERESTS OF CITIZEN CHILDREN: HOW TO BALANCE COMPETING IMPERATIVES IN THE CONTEXT OF REMOVAL PROCEEDINGS
Patrick Glen..........................................................1

FICTITIOUS STATES, EFFECTIVE CONTROL, AND THE USE OF FORCE AGAINST NON-STATE ACTORS
Brian Finucane ................................................................................................................35

THE EVOLVING DEFINITION OF THE REFUGEE IN CONTEMPORARY INTERNATIONAL LAW
William Thomas Worster .........................................................................................94

THE “RIGHT TO BE FORGOTTEN”: RECONCILING EU AND US PERSPECTIVES
Steven C. Bennett .......................................................................................................161

VALUATION IN INVESTOR-STATE ARBITRATION: TOWARD A MORE EXACT SCIENCE
Joshua B. Simmons ..................................................................................................196

Reviews

REVIEW OF TWILIGHT OF IMPUNITY BY JUDITH ARMATTA
Maya Karwande..........................................................251
The Removability of Non-Citizen Parents and the Best Interests of Citizen Children: How to Balance Competing Imperatives in the Context of Removal Proceedings

Patrick Glen
The Removability of Non-Citizen Parents and the Best Interests of Citizen Children: How to Balance Competing Imperatives in the Context of Removal Proceedings

By
Patrick Glen*

I. INTRODUCTION

With comprehensive immigration reform effectively stalled in the United States (US) Congress and ripples of discontent still roiling domestic and international financial and economic markets, illegal immigrants again have become a convenient target for an emergent nativist sentiment in the United States and abroad. In Europe, this sentiment has taken on ugly overtones, complete with threatened and real violence against new arrivals, both legal and illegal.1 The rhetoric has not always been as heated in the United States, but the preceding three years have seen a sharp rise in state legislation both to curb illegal immigration and to make it easier to discern those who are present in the United States illegally. State laws from Arizona,2 Georgia,3 and Alabama4 have taken

---

* Adjunct Professor, Georgetown University Law Center; Attorney, Office of Immigration Litigation, United States Department of Justice. The views and opinions reflected in this article do not represent those of the federal government or Department of Justice.


2. See S.B. 1070, 49th Leg., 2d Reg. Sess. (Ariz. 2011). The Ninth Circuit has enjoined enforcement of large sections of the Arizona law. See United States v. Arizona, 641 F.3d 339, 344 (9th Cir. 2011). Arizona since has filed a petition for certiorari, which, as of the date of this writing, has not been acted on by the Supreme Court. Ariz. Asks High Court to Rule on SB 1070, ARIZ. REPUBLIC, Aug. 11, 2011, at A1.


center stage in this drama, but these states are by no means alone. One specific
effect of this general reaction against illegal immigrants is a renewed focus on
the conferral of US citizenship through the *jus soli* principle, that is, birth in this
country.  

On January 5, 2011, State Legislators for Legal Immigration, an organization
composed of state lawmakers from across the United States, unveiled a model
citizenship law that would define who is a citizen under both state law and the Federal Constitution, thereby taking aim at *jus soli* citizenship. The proposed model law represents an attempt to reverse a century-old constitutional interpretation of the citizenship clause of the Fourteenth Amendment, which provides that “All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside.” The Supreme Court last interpreted the scope of the citizenship clause in 1898, when it held, with very limited exceptions, that the Constitution establishes *jus soli* citizenship—the conferral of US citizenship on those born in this country.

The model law attempts to limit the prevailing interpretation of the citizenship clause by defining the phrase “subject to the jurisdiction thereof”:

[S]ubject to the jurisdiction of the United States has the meaning that it bears in Section 1 of the Fourteenth Amendment to the United States Constitution, namely that the person is a child of at least one parent who owes no allegiance to any foreign sovereignty, or a child without citizenship or nationality in any foreign country. For the purposes of this statute, a person who owes no allegiance to any foreign sovereignty is a United States citizen or national, or an immigrant accorded the privilege of residing permanently in the United States, or a person without citizenship or nationality in any foreign country.

To the extent that the aforementioned model citizenship law is contrary to the Supreme Court’s decision in *Wong Kim Ark*, and fails to offer any compelling justification for reversing that constitutional interpretation, its fate would seem clear.

---

7. U.S. CONST. amend. XIV, § 1.
10. See Glen, supra note 9, at 103-07 (discussing whether there are any grounds that would
This proposal, and others like it, however, has gained momentum by taking aim at so-called “anchor babies”—children born to undocumented immigrants in the United States who, by that birth, become US citizens and provide, by the logic of these laws’ proponents, an anchor by which undocumented immigrants may remain in the United States. Whatever view one takes on this specific issue, it is clear that there is a significant population of undocumented immigrants in the United States with at least one US citizen child. “Almost 9 million people live in families with at least one unauthorized immigrant. Included in the population of unauthorized immigrants are 3.8 million parents of US citizen children. Parents of US citizen children, therefore, make up 37% of the adult population of unauthorized immigrants.” As the issue currently stands, the main focus is on the removal of non-citizen parents who have citizen children and to what extent the interests of those children should affect the removability of the parents. Thus, the subject of the instant article is on how the interests of citizen children should weigh in the balance of determining whether a non-citizen parent or parents should be removed.

Rather than approach the issue solely from the perspective of US immigration law and policy, this article offers a comparative assessment of US domestic law and policy with that of the United Kingdom. This approach is well-founded in general considerations, including the fact that the US and UK have a shared common-law history, have risen to similar levels of economic development, are devoted to rule-of-law principles, and are both attractive destinations for immigrants seeking a better life. More specifically, the comparison is apt because the US and UK immigration systems share the same broad baseline considerations in determining the best interests of citizen children in adjudicating the immigration claims of their non-citizen parents. Nonetheless, the approach of each country in weighing those considerations is quite distinct, as the subsequent sections of this article will make clear. In considering the differences in approach between these two countries, one can glean both what works and areas of improvement in each system, thereby leading to a synthesis that includes the best of each approach.

Part I provides a brief sketch of UK immigration law and addresses the recent landmark decision of the UK Supreme Court in ZH v. Secretary of State for
the Home Department. That decision holds that the best interests of the citizen child are a primary consideration in determining whether the United Kingdom can remove a non-citizen consistent with its obligations under the European Convention on Human Rights.14 By prioritizing the interests of citizen children in the removal proceedings of their parents, ZH serves as the template for future cases raising such issues. Part II addresses US immigration law on the removability of non-citizens and the potential avenues for relief from removal to which they may turn. Contrary to the claims of most critics of the prevailing interpretation of the Fourteenth Amendment, US citizen children do not provide an effective “anchor” for non-citizen parents in the United States. This review will make clear that not only does a citizen child not provide an automatic trump to the removal of a non-citizen parent, but also the fact of having such a child rarely gives rise to actual relief from removal. Finally, Part III offers pros and cons regarding the approaches taken by the United Kingdom and the United States, and Part IV attempts to synthesize the best parts of each approach in proposing a new framework for weighing the interests of citizen children confronted with removal of a non-citizen parent.

The competing interests in this issue are weighty and the best framework is likely to leave all parties unhappy in some respects. Nonetheless, there are ways to balance these interests effectively and equitably while also ensuring strict enforcement and compliance with the US immigration system. Moreover, the US government seems ready to accept more balancing of interests in immigration decisions: recently, the Obama administration emphasized the discretionary nature of the Department of Homeland Security’s enforcement authority, which supports the feasibility of taking a more nuanced approach in circumstances where the removal of a non-citizen parent implicates his or her citizen children.15

II.
REMOVING NON-CITIZEN PARENTS OF CITIZEN CHILDREN:
THE UK APPROACH

The United Kingdom’s main domestic immigration statute is the 1971 Immigration Act, as amended.16 Section 3 of the Act contains two provisions relating to the removability of non-citizens from the United Kingdom. First, “[a] per-
Glen: The Removability of Non-Citizen Parents and the Best Interests of

son who is not a British citizen is liable to deportation from the United Kingdom if—(a) the Secretary of State deems his deportation to be conducive to the public good; or (b) another person to whose family he belongs is or has been ordered deported. 17 Second, “a person who is not a British citizen shall also be liable to deportation from the United Kingdom if, after he has attained the age of seventeen, he is convicted of an offence for which he is punishable with imprisonment and on his conviction is recommended for deportation by a court empowered by this Act to do so.” 18 If a non-citizen is liable to deportation, the Secretary of State for the Homeland may order him deported. 19 Nonetheless, “[a] deportation order against a person may at any time be revoked by a further order of the Secretary of State, and shall cease to have effect if he becomes a British citizen.” 20

ZH’s case came before the administrative agency and UK courts via an article 8 application to enforce the “[r]ight to respect for private and family life” under the European Convention on Human Rights. 21 Article 8 provides:

1. Everyone has the right to respect for his private and family life, his home and correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Article 8 is enforceable in the UK’s domestic administrative agencies and courts via the Human Rights Act of 1998. 22 For the UK Supreme Court, the issue in ZH had both general and specific dimensions. The “over-arching issue” was the “weight to be given to the best interests of children who are affected by the decision to deport one or both of their parents from this country.” 23 However, the Court saw within this issue a much more specific question: “[I]n what circumstances [was] it permissible to remove or deport a non-citizen parent where the effect will be that a child who is a citizen of the United Kingdom will also have to leave?” 24

17. Id. at § 3(5).
18. Id. at § 3(6).
19. Section 5(1) of the Immigration Act provides that, as to a person liable to deportation under sections 3(5) and 3(6) of the Act, “the Secretary of State may make a deportation order against him, that is to say an order requiring him to leave and prohibiting him from entering the United Kingdom; and a deportation order against a person shall invalidate any leave to enter or remain in the United Kingdom given him before the order is made or while it is in force.” Id. at § 5(1).
20. Id. at § 5(2).
24. Id.
ZH, a female Tanzanian citizen and national, arrived in the United Kingdom in December 1995, where she met a British citizen with whom she had two children—a daughter, born in 1998, and a son, born in 2001. Both children were British citizens because they were born in the United Kingdom and had a British citizen father. Between 1997 and 2000, the British Government denied ZH asylum three times, once in an application under her own name and twice in applications filed with false identities. In 2001, ZH filed an initial article 8 application with the administrative immigration tribunal, but the tribunal denied her application and dismissed her appeal. In 2004, ZH applied for leave to remain in the United Kingdom pursuant to the “one-off family concession,” but the immigration tribunal denied this application in 2006 due in large part to her prior fraudulent asylum claims. Before the disposition of that application, ZH also applied to remain in the United Kingdom pursuant to the “seven year child” concession, but the immigration tribunal denied her application, again based on her fraudulent asylum claims. In 2005, while ZH was pursuing these various applications, she and the children’s father separated, but he continued to see the children regularly, visiting each month for about a week. However, he was diagnosed with human immunodeficiency virus (“HIV”) in 2007. At the time of ZH’s proceedings, he was living with his parents and current wife on disability allowance and was “reported to drink a great deal.”

Following the father’s diagnosis, ZH again sought to forestall her removal via an article 8 application, but she was unsuccessful before the immigration judge. Subsequently, a senior immigration judge granted reconsideration of ZH’s claim and found that the initial immigration judge failed to consider adequately the children’s British citizenship or to take “into account the rights of the children and the effect of the mother’s removal upon them.”

Nonetheless, after considering the evidence, the senior immigration judge

25. Id. at [2].
26. Id. at [5].
27. Id. See generally UK BORDER AGENCY, APU NOTICE: ONE-OFF EXERCISE TO ALLOW FAMILIES WHO HAVE BEEN IN THE UK FOR THREE OR MORE YEARS TO STAY, 2004, 4/2003 (U.K.), http://www.ukba.homeoffice.gov.uk/sitecontent/documents/policyandlaw/asylumpolicyinstructions/apunotices/oneoffexercise.pdf?view=Binary (discussing the “one-off family concession,” which allowed families who had been in the UK for three or more years to stay when certain eligibility criteria were met).
28. ZH, [2011] UKSC 4 at [5]. The colloquially termed “seven year child concession” was a policy implemented in order to determine “whether enforcement action should proceed or be initiated against parents of a child who was born [in the UK] and has lived continuously to the age of seven or over, or where, having come to the UK at an early age, they have accumulated seven years or more continuous residence.” See 485 PARL. DEB., H.C. (6th series) (2008) 49 (U.K.), http://www.publications.parliament.uk/pa/cm200809/cm Hansrd/cm081209/wnstext/81209m0002.htm?o81209430008019. That policy was withdrawn as of December 9, 2008. 485 PARL. DEB., H.C. (6th series) (2008) 49 (U.K).
30. Id. at [6].
upheld ZH’s removal as proportionate. 31 The judge recognized that ZH’s removal would place a heavy burden on the children’s relationship with one of their parents: if the children accompanied ZH to Tanzania, it would “substantially interfere” with their relationship with their father; if the children remained with their father in the United Kingdom, it would “substantially interfere” with their relationship with their mother. 32 Nevertheless, he concluded that ZH’s removal was not a disproportionate response to her unlawful presence for three reasons. First, the judge emphasized that ZH was “seriously lacking in credibility.” 33 Second, he considered ZH’s pregnancies irresponsible given her precarious immigration status in the United Kingdom and viewed the current dilemma as a foreseeable result of the couple’s prior choices. 34 Third, he found that the children could live with either parent without practical difficulty—essentially ignoring the father’s HIV status and financial difficulties. 35 Given the above factors and the general imperative of immigration control, the judge found that an interference with the children’s family life did not outweigh the importance of removing ZH. 36

The Court of Appeal for England and Wales upheld this decision. Counsel for ZH argued that the children’s citizenship acted as a “trump card,” prohibiting their mother’s removal. 37 The court rejected this proposition as too absolutist and inconsistent with governing law. 38

Before the UK Supreme Court, counsel for ZH retreated from the absolutist position presented before the Court of Appeal and argued simply “that insufficient weight [was] given to the welfare of all children affected by decisions to remove their parents and in particular to the welfare of children who are British citizens.” 39 The Secretary of State decided that ZH’s removal would be disproportionate but questioned what the appropriate governing principles would be in cases where the removal of non-citizen parents had the potential to affect citizen children.

The opinion of the Court began by reciting the appropriate standards by which to adjudicate an article 8 application. In such an application the authorities are required to consider the rights of all potentially affected family members, not just the specific applicant herself. 40 In the words of Lady Hale:

31. Id. at [7]-[10].
32. Id. at [7].
33. Id. at [8].
34. Id. at [8].
35. Id. at [9]-[10].
36. Id. at [10].
37. Id. at [11].
38. Id.
39. Id. at [12].
40. Id. at [14] (“Together these members enjoy a single family life and whether or not the removal would interfere disproportionately with it has to be looked at by reference to the family unit
The decision-maker has to balance the reason for the expulsion against the impact upon other family members, including any alternative means of preserving family ties. The reason for deporting may be comparatively weak, while the impact on the rest of the family, either of being left behind or of being forced to leave their own country, may be severe. On the other hand, the reason for deporting may be very strong, or it may be entirely reasonable to expect the other family members to leave with the person deported.41

Thus, the determination of an article 8 application is intricately tied to the specific facts and circumstances of the applicant.42 Nonetheless, the Court noted that such applications generally arise in one of two circumstances: (1) where the non-citizen is a long-settled resident of a country and has committed a criminal offense that subjects him to removal or (2) “where a person is to be removed because he or she has no right to be or remain in the country.”43 ZH’s case fell under the second situation, as it was her illegal presence alone that subjected her to the possibility of removal.

According to the European Court of Human Rights (ECtHR), in an article 8 proceeding, a national court must decide whether the removal of a non-citizen will interfere with “the right to respect for family life.” If so, the removal must be “necessary in a democratic society and proportionate to the legitimate aim pursued.”44 In making this determination, the starting point for the state authorities and the reviewing courts “is the right of all states to control the entry and residence of aliens.” The state’s legitimate aim likely will be “the economic well-being of the country in controlling immigration, although the prevention of disorder and crime and the protection of the rights and freedoms of others may also be relevant.”45 Other pertinent factors include the length of the applicant’s stay in the country from which he or she is to be expelled; the concerned persons’ nationalities; the applicant’s family situation, including the length and closeness of any marriage; whether there are children from their marriage and their ages; whether family life was established knowing the precariousness of the immigration situation; and the difficulties attendant upon the applicant’s removal to and life in the country of nationality.46 Finally, and most pertinent to this inquiry, authorities must consider “the best interests and well-being of the children,” specifically the children’s potential difficulties in the applicant’s destination country and the “solidity of social, cultural and family ties” with both

42. EB v. Sec’y of State for the Home Dep’t, [2008] UKHL 41, [12] (“Thus the appellate immigration authority must make its own judgment and that judgment will be strongly influenced by the particular facts and circumstances of the particular case.”).
43. ZH, [2011] UKSC 4 at [18].
44. Id. at [17].
45. Id. at [18].
46. Id. at [17]-[18] (citing Uner v. The Netherlands, 45 Eur. Ct. H.R. 421 (2007)).
the host country and the destination country.\(^{47}\)

Despite the ostensibly strict standard of review of article 8 applications, the ECtHR has found violations of article 8 even in egregious cases of illegal presence with children involved. For instance, in *Rodrigues da Silva Hoogkamer v. The Netherlands*, the ECtHR found a Brazilian mother’s expulsion from the Netherlands disproportionate, despite the fact that she started her family in full knowledge of her illegal presence.\(^{48}\) In *ZH*, the UK Supreme Court explained the ECtHR’s article 8 decisions as a response to increasing international awareness of the importance of children’s rights in making legal determinations under both international and domestic law.\(^{49}\)

The United Kingdom is party to the United Nations Convention on the Rights of the Child, and that instrument requires that “[i]n all actions concerning children . . . the best interests of the child shall be a primary consideration.”\(^{50}\) The United Kingdom implemented this obligation into domestic law via the Borders, Citizenship and Immigration Act, which provides that all immigration determinations must “regard . . . the need to safeguard and promote the welfare of children who are in the United Kingdom.”\(^{51}\)

In addition to reviewing relevant ECtHR decisions and United Nations (“UN”) documents, the UK Supreme Court also looked at case law from other countries. The Court cited with approval the Federal Court of Australia’s reasoning in a similar case:

[The Tribunal] was required to identify what the best interests of Mr. Wan’s children required with respect to the exercise of its discretion and then to assess whether the strength of any other consideration, or the cumulative effect of other considerations, outweighed the consideration of the best interests of the children understood as a primary consideration.\(^{52}\)

In its reliance on this case, the UK Supreme Court put forth its belief that the best interests of the children must be considered first, but other considerations could outweigh it in determining the proportionality of removal.\(^{53}\) Having examined the ECtHR decisions, UN documents and foreign case law, the UK Supreme Court ultimately concluded that the best interests of the child are a primary consideration, but not the primary consideration, in a removability proceeding.\(^{54}\)

Having established that the best interests of the child will be a primary consideration in determining a non-citizen parent’s removal, the UK Supreme Court


\(^{49}\) *See ZH*, [2011] UKSC 4 at [21]-[28].


\(^{52}\) *Wan v. Minister for Immigration & Multicultural Affairs* [2001] FCA 568, ¶ 32 (Austl.).

\(^{53}\) *ZH*, [2011] UKSC 4 at [26].

\(^{54}\) *Id.* at [29]-[33].
then identified what is encompassed by the phrase “best interests of the child.” According to the Court, this standard required asking “whether it [was] reasonable to expect the child to live in another country.” 55 To answer this question, the Court listed several relevant factors: the level of the child’s integration in his present country and the length of absence from the other country; living and caretaking arrangements in the other country; and the strength of the child’s relationships with parents or other family members which would be severed if the child moved away.56 The Court noted that the nationality of the children will be an important factor in gauging the children’s best interests even if citizenship will not operate as a “trump card” against the parent’s removal.57

Applying these principles to the facts of the case, the Court noted that both of ZH’s children were British citizens by birth and descent and that “they ha[d] an unqualified right of abode” in the United Kingdom.58 They had lived in Britain, were receiving a British education, had social ties to Britain, and had a positive relationship with their father.59 As British citizens, the children had rights which “they w[ould] not be able to exercise if they move[d] to another country.”60 The Court then examined whether the children’s interests, while a prima consideration, could be “outweighed by the cumulative effect of other considerations.”61 The Court acknowledged numerous factors that weighed against ZH, namely “the need to maintain firm and fair immigration control . . . [ZH’s] appalling immigration history and the precariousness of her position when family life was created.”62 Nonetheless, the Court found that “there really [was] only room for one view,” as the children would have had to leave the United Kingdom with their mother based on a legal transgression for which they bore no responsibility.63 Accordingly, the Court allowed ZH’s appeal, and thus forestalled her deportation.64

Lord Hope’s and Lord Kerr’s concurring opinions further emphasized the child’s interests in UK deportation proceedings.65 In his opinion, Lord Hope

55. Id. at [29].
56. Id.
57. See id. at [30] (“Although nationality is not a ‘trump card’ it is of particular importance in assessing the best interests of any child.”).
58. Id. at [31].
59. Id.
60. Id. at [32].
61. Id. at [33].
62. Id.
63. Id. (quoting Simon Brown L.J. in Edore v. Sec’y of State for the Home Dep’t [2003] 1 WLR 2979, [26]).
64. Id. at [38]. The Court also established the possibility that the government should directly consult the children in certain circumstances regarding their interests in specific cases. Id. at [34]-[37].
65. Id. at [39] (opinion of Lord Hope); Id. at [45] (opinion of Lord Kerr).
took aim at two errors he perceived in the decision of the Court of Appeals.\footnote{Id. at \[40\].}
First, although he agreed that the Court properly rejected the assertion that a child’s citizenship will trump the state’s interest in the removal of a non-citizen parent, he argued that it nonetheless failed to correctly weigh the fact of citizenship. According to Lord Hope, citizenship may not be a trump, but:

\[\text{[I]}\text{t will hardly ever be less than a very significant and weighty factor against moving children who have that status to another country with a parent who has no right to remain here, especially if the effect of doing this is that they will inevitably lose those benefits and advantages for the rest of their childhood.}\footnote{Id. at \[40\]-\[41\] (opinion of Lord Hope).} \]

Second, both the administrative tribunal and the Court of Appeals examined the interference with the family unit in light of the mother’s own immigration transgressions, while “[t]he best interests of the children melted away into the background.”\footnote{Id. at \[42\].} Lord Hope explained that neither consideration would be dispositive: weighing both factors together would dictate the outcome of the removal proceeding:

There is an obvious tension between the need to maintain a proper and efficient system of immigration control and the principle that, where children are involved, the best interests of the children must be a primary consideration . . . The fact that the mother’s immigration status was precarious when they were conceived may lead to a suspicion that the parents saw this as a way of strengthening her case for being allowed to remain here. But considerations of that kind cannot be held against the children in this assessment. It would be wrong in principle to devalue what was in their best interests by something for which they could in no way be held responsible.\footnote{Id. at \[44\].}

Lord Kerr’s opinion went further in its deference to the child’s interests. He believed that the best interests of the child should dictate the ultimate disposition of a given case unless there were substantial countervailing considerations—in essence, the best interests would create a rebuttable presumption against the removal of a non-citizen parent.\footnote{Id. at \[46\] (opinion of Lord Kerr).}

Succinctly distilled, the UK Supreme Court concluded that the interests of citizen children should play a role in determining whether removal of a non-citizen parent would be appropriate. It also deemed those interests to be “a primary consideration” in weighing the competing interests. Finally, the Court held that a determination of the best interests of the child should not lead inexorably to a disposition in line with those interests and that other considerations could conceivably outweigh the “best interests” of the child. Although there is much in the language of the Court’s decision to recommend this framework in the context of removing non-citizen parents with citizen children, the Court undercut its own statement of the rule in its application thereof.

\begin{footnotes}
\item 66. Id. at \[40\].
\item 67. Id. at \[40\]-\[41\] (opinion of Lord Hope).
\item 68. Id. at \[42\].
\item 69. Id. at \[44\].
\item 70. Id. at \[46\] (opinion of Lord Kerr).
\end{footnotes}
The Court forcefully asserted that the child’s citizenship will not act as a trump against removal, but it is hard to conceive of any alternative in the context of the Court’s holding in ZH’s case. ZH filed at least two fraudulent asylum claims. She gave birth to two children in the United Kingdom after the British Government rejected her claims to status and notwithstanding the reality that there were no avenues available to her through which to normalize her status. It would be hard to imagine a situation where the abuses of the applicant are greater than in the instant case. It is hard to believe that, aside from her children’s citizenship, her removal would not be proportionate. The health status of the father and his precarious living situation could explain why the children’s remaining in the United Kingdom would be an unreasonable prospect. The United Kingdom viewed him as unable provide the appropriate level of care given his own health issues. Yet that factor alone does little to undercut the likelihood of the children having to depart with their mother. In short, based on the Court’s decision in ZH, it seems that the child’s citizenship will be determinative, even though the Court went to great pains to refute or restrict that assertion.

To be sure, beyond this critique there are perhaps open issues left in the decision’s wake. First, what if both parents are removable as non-citizens? If citizenship is not to act as a trump and the family would remain intact after removal, it seems that other factors would outweigh the best interests of the citizen children. Early in its opinion, the Court noted a pending case presenting such facts. Accordingly, the courts may resolve this issue sooner than later. Second, what if the children are not citizens of the United Kingdom? This will lead to a fact intensive assessment of the children’s connections to the United Kingdom, among other considerations. Removing the importance of citizenship, however, even long-term residence would likely not halt deportation proceedings. In any event, these are questions that the UK courts will have to confront in the coming years. ZH provides, at the least, an appropriate framework in which to consider these additional issues.

III. REMOVING NON-CITIZEN PARENTS OF CITIZEN CHILDREN: THE US APPROACH

Whereas the previous section dealt with UK immigration law and policy, this section examines US immigration law and policy as it relates to the deportation of non-citizens. US immigration law is codified in the Immigration and Nationality Act of 1952 (“INA”), as amended. Non-citizens may be subject to deportation proceedings in the United States because they are either inadmissible or deportable under sections 212 and 237, respectively, of the INA. The for-
mer applies to aliens seeking lawful admission to the United States, whereas the latter applies to aliens who have been admitted or are otherwise already present in the United States.\textsuperscript{74} If the US Government successfully establishes that an alien meets the criteria for inadmissibility or deportability, the alien is then placed in removable status regardless of whether they may have US citizen children.\textsuperscript{75} In other words, the sole consideration at this stage is whether the alien falls within the inadmissibility or deportability criteria of the INA.

Nonetheless, the INA does provide that otherwise removable aliens may apply for certain waivers. Some of these waivers are available only for spouses and children of US citizens, thus excluding parents of US citizens. For instance, if an alien is inadmissible under INA section 212(a)(6)(C)(i) for seeking to procure or procuring a US visa, immigrant documentation, or admission “by fraud or willfully misrepresenting a material fact,”\textsuperscript{76} he may apply for a waiver under INA section 212(i)(1) if he is “the spouse, son, or daughter of a United States citizen or of an alien lawfully admitted for permanent residence” and “it is established to the satisfaction of the Attorney General that the refusal of admission . . . of such immigrant alien would result in extreme hardship to the citizen or lawfully resident spouse or parent[.]”\textsuperscript{77} Other waivers are available without regard to whether the non-citizen alien has a qualifying relationship with a US citizen. For example, INA sections 212(d)(11) and 237(a)(1)(E)(iii) provide for waivers if the Attorney General charges the alien with smuggling, provided that “the alien has encouraged, induced, assisted, abetted, or aided only an individual who at the time of such action was the alien’s spouse, parent, son, or daughter[.]”\textsuperscript{78} The imperative behind these waivers, as with other relief provisions of the INA, is family unity.\textsuperscript{79}

For other waivers, the fact that a non-citizen parent may have a US citizen child will be a factor in determining his eligibility for such a waiver. In such circumstances, the importance of this qualifying relationship may trump otherwise compelling interests in the removal of the non-citizen parent. For instance, under the INA, “[a]ny immigrant who is or has been a member of or affiliated with the Communist or any other totalitarian party ... domestic or foreign” is inadmissible.\textsuperscript{80} Nonetheless, the Attorney General may waive this provision if the immigrant “is the parent, spouse, son, daughter, brother, or sister of a citizen of the

\textsuperscript{74} 8 U.S.C. § 1227.
\textsuperscript{75} See id. § 1229a(a)(1), (c)(3); 8 C.F.R. § 1240.8 (2010).
\textsuperscript{76} 8 U.S.C. § 1182(a)(6)(C)(i).
\textsuperscript{77} Id. § 1182(i)(1).
\textsuperscript{78} Id. §§ 1182(d)(11), 1227(a)(1)(E)(iii); see also Batista v. Gonzales, 494 F.3d 67 (2d Cir. 2007) (addressing eligibility for a waiver under INA § 212(d)(i)).
\textsuperscript{79} See, for example, 8 U.S.C. § 1182(d)(11) (“The Attorney General may, in his discretion for humanitarian purposes, to assure family unity . . . .”).
\textsuperscript{80} Id. § 1182(a)(3)(D)(i).
United States . . . for humanitarian purposes, to assure family unity, or when it is otherwise in the public interest if the immigrant is not a threat to the security of the United States."81 The Attorney General may deem an alien inadmissible if they have "a communicable disease of public health significance."82 The Attorney General may waive this ground "in the case of an alien who . . . has a son or daughter who is a United States citizen[.]"83 Criminal convictions, involvement in criminal schemes, and intent to engage in illegal activities if admitted to the United States are encompassed by numerous grounds of inadmissibility and removability.84 The INA nonetheless provides certain waivers even in cases of criminality. The Attorney General, however, may waive certain criminal grounds of inadmissibility "in the case of an immigrant who is the spouse, parent, son, or daughter of a citizen of the United States . . . if it is established to the satisfaction of the Attorney General that the alien’s denial of admission would result in extreme hardship to the United States citizen . . . spouse, parent, son, or daughter."85 If the US Government deems that an alien within the United States is removable on account of being inadmissible at the time of his admission, the alien may obtain a waiver from the Attorney General if he or she is "the spouse, parent, son, or daughter of a citizen of the United States," assuming he or she meets other eligibility criteria.86

Beyond the statutory exceptions that the Attorney General must make in these types of cases, including discerning whether, in the appropriate cases, the non-citizen has a qualifying relationship with a US citizen, discretionary factors are the most important facet in determining whether waiver should be granted. The Department of Justice Board of Immigration Appeals has the ultimate discretion as to whether to grant or deny a waiver, but as part of that determination the Board must make an initial statutory eligibility determination about whether the removal of the alien will result in "extreme hardship" to a qualifying relative. The Board, however, does not define "extreme hardship" in the context of the preceding waivers. As the Board has written, "extreme hardship is not a definable term of fixed and inflexible meaning, and the elements to establish extreme hardship are dependent upon the facts and circumstances of each case."87 Nonetheless, in its decision in Matter of Cervantes-Gonzalez, the Board did enumerate certain factors that would be relevant to the determination:

The factors deemed relevant in determining extreme hardship to a qualifying relative include, but are not limited to, the following: the presence of lawful permanent residence or US citizen family ties to this country; the qualifying relative’s

81. Id. § 1182(a)(3)(D)(iv).
82. Id. § 1182(a)(1)(A)(i).
83. Id. § 1182(g)(1)(B).
84. See generally id. § 1182(a)(2).
85. Id. § 1182(h)(1)(B).
86. Id. § 1227(a)(1)(H).
family ties outside the United States; the conditions in the country or countries to which the qualifying relative would relocate and the extent of the qualifying relative’s ties to such countries; the financial impact of departure from this country; and, finally, significant conditions of health, particularly when tied to an unavailability of suitable medical care in the country to which the qualifying relative would relocate.\footnote{Id. at 565-66; see also 8 C.F.R. § 1240.58(b) (2010) (relating to the repealed relief of suspension of deportation) (“To establish extreme hardship, an applicant must demonstrate that deportation would result in a degree of hardship beyond that typically associated with deportation. Factors that may be considered in evaluating whether deportation would result in extreme hardship to the alien or to the alien’s qualified relative include, but are not limited to, the following: (1) The age of the alien, both at the time of entry to the United States and at the time of application for suspension of deportation; (2) The age, number, and immigration status of the alien’s children and their ability to speak the native language and to adjust to life in the country of return; (3) The health condition of the alien or the alien’s children, spouse, or parents and the availability of any required medical treatment in the country to which the alien would be returned; (4) The alien’s ability to obtain employment in the country to which the alien would be returned; (5) The length of residence in the United States; (6) The existence of other family members who are or will be legally residing in the United States; (7) The financial impact of the alien’s departure; (8) The impact of a disruption of educational opportunities; (9) The psychological impact of the alien’s deportation; (10) The current political and economic conditions in the country to which the alien would be returned; (11) Family and other ties to the country to which the alien would be returned; (12) Contributions to and ties to a community in the United States, including the degree of integration into society; (13) Immigration history, including authorized residence in the United States; and (14) The availability of other means of adjusting to permanent resident status.”).}

Discerning whether a US citizen child would face extreme hardship upon the removal of a non-citizen parent provides one basis by which to weigh the best interests of that child. In Matter of Cervantes-Gonzalez, the Board did not find extreme hardship, but that case did not involve a US citizen child.\footnote{In re Cervantes-Gonzalez, 22 I. & N. Dec. at 566.} The only qualifying relative in that case was the alien’s citizen spouse; they married during the course of removal proceedings, she was a native Mexican and could thus easily acclimate to life in Mexico, there were no financial ties to the United States, and the wife made no claim that she would suffer extreme hardship if her alien spouse was removed.\footnote{Id. at 566-68.} In Matter of Kao & Lin, the Board did find that the removal of non-citizen parents would result in extreme hardship to at least one of their US citizen children.\footnote{In re Kao & Lin, 23 I. & N. Dec. 45, 51 (B.I.A. 2001) (on application for suspension of deportation).} The husband and wife applicants in that case were Taiwanese, had lived in the United States for 19 and 17 years, respectively, and had five US citizen children who were not fluent in Chinese.\footnote{Id. at 46, 50.} The Board found “nothing in the record to indicate that [the children’s] language capabilities . . . [were] sufficient for an adequate transition to daily life in Taiwan.”\footnote{Id. at 50.} Moreover, the Board emphasized that the children “ha[d] lived their entire lives in the
United States and [were] completely integrated into their American lifestyles. Their needs for housing, food, clothing, education, and community support had been adequately met. As a result, the Board was "satisfied that to uproot the oldest daughter . . . at this stage in her education and social development and to require her to survive in a Chinese-only environment would [have been] a significant disruption that would constitute extreme hardship."

To establish eligibility for a waiver of inadmissibility or deportability, the alien must demonstrate not only statutory eligibility, including extreme hardship to a qualifying relative, but also that the waiver "is warranted in the exercise of discretion." Thus, even if statutory criteria are met, the Board may deny the alien's waiver application. In determining whether a waiver is justified, the adjudicator considers the adverse and positive factors in the specific case, thus providing a second opportunity to weigh the best interests of citizen children. Adverse factors include "the nature and underlying circumstances of the exclusion ground at issue, the presence of additional significant violations of this country's immigration laws, the existence of a criminal record and, if so, its nature, currency and seriousness, and the presence of other evidence indicative of an alien's bad character or undesirability as a permanent resident of this country." Favorable considerations include:

- family ties in the United States, residence of long duration in this country (particularly where the alien began his residency at a young age), evidence of hardship to the alien and his family if he is excluded and deported, service in this country's Armed Forces, a history of stable employment, the existence of property or business ties, evidence of value and service to the community, evidence of genuine rehabilitation if a criminal record exists, and other evidence attesting to the alien's good character.

In weighing these adverse and positive factors, the adjudicator must take into account the "underlying significance" of these factors:

[If the alien has relatives in the United States, the quality of their relationship must be considered in determining the weight to be awarded this equity. Further, the equity of a marriage and the weight given to any hardship to the spouse is diminished if the parties married after the commencement of deportation proceedings, with knowledge that the alien might be deported.

In Matter of Mendez-Moralez, the Board upheld the denial of a waiver of inadmissibility under INA section 212(h)(1)(B) despite the fact that Mendez-
Moralez had a citizen wife, three citizen children, two lawful permanent resident brothers, lawful permanent resident parents, and supported his family financially, because it found that adverse factors—the seriousness of his conviction for sexual assault of a minor and his lack of rehabilitation—outweighed these equities.102

Thus, although the Justice Department’s determination of removability of a non-citizen parent will not make reference to any US citizen children that the alien parent may have, the existence of such children may influence the Department’s ultimate inadmissibility or deportation determination. In making that decision, the INA permits the adjudicator to assess and weigh the best interests of the citizen child at two points, first at the threshold level of determining statutory eligibility, and second if and when the adjudicator exercises his or her discretion in granting the alien parent a waiver.

In addition to waivers of inadmissibility or deportability, the INA also provides various forms of relief, some of which are conditioned on a non-citizen possessing a qualifying relationship with a US citizen. This article focuses on two forms of relief: (i) status adjustment and (ii) cancellation of deportation proceedings for certain non-lawful permanent residents.

An alien non-lawful permanent resident in removal proceedings may seek cancellation of removal proceedings.103 To establish his eligibility for such relief, he must demonstrate that he has remained in the United States for a continuous period of ten years preceding his application, has good moral character, has not been convicted of certain disqualifying offenses, and “that removal would result in exceptional and extremely unusual hardship to the alien’s spouse, parent, or child, who is a citizen of the United States[]”.104 Even if the applicant meets the statutory eligibility criteria, he must still establish that cancellation of removal should be granted in the exercise of discretion.105 As with the various waivers discussed above, the focus will often be on if the alien can establish “exceptional and extremely unusual hardship” to a qualifying relative.106

The Board has held that the “exceptional and extremely unusual hardship” standard is much higher than the “extreme hardship” threshold required for INA section 212 inadmissibility waivers.107 The heightened requirement reflects Congress’s intent to limit cancellations of removal proceedings for non-permanent residents to “truly exceptional cases” where the applicant can demonstrate hardship “substantially’ beyond the ordinary hardship that would be ex-

102. Id. at 302-05.
105. See id. § 1229b(b)(1)(D).
106. Id.
107. See In re Monreal-Aguinaga, 23 I. & N. Dec. 56, 59 (B.L.A. 2001) (en banc) (citing Cortes-Castillo v. INS, 997 F.2d 1199 (7th Cir. 1993); Brown v. INS, 775 F.2d 383 (D.C. Cir. 1985)).
pected when a close family member leaves this country.” Nonetheless, the factors the agency considers in determining whether removal would result in exceptional and extremely unusual hardship are identical to those utilized in finding extreme hardship: the adjudicator simply weighs the evidence differently in light of the higher burden espoused by the alien seeking relief.

As with the extreme hardship determination, the adjudicator will not consider hardship factors relating to the applicant himself unless such factors also impact or affect the potential hardship to a qualifying relative. In evaluating the import given to qualifying relatives, the adjudicator should “consider the ages, health, and circumstances” of those relatives. For example, the Board in Matter of Monreal-Aguinaga opined “an applicant who has elderly parents in this country who are solely dependent upon him for support might well have a strong case. Another strong applicant might have a qualifying child with very serious health issues, or compelling special needs in school.” Economic issues may also be relevant: “A lower standard of living or adverse conditions in the country of return are factors to consider only insofar as they may affect a qualifyingrelative, but generally will be insufficient in themselves to support a finding of exceptional and extremely unusual hardship.”

The Board’s reasoning in its decisions illustrates how high a standard “exceptional and extremely unusual hardship” is in practice. In Matter of Monreal-Aguinaga, the Board found no exceptional and extremely unusual hardship in removing a non-citizen father with three US citizen children, one of whom had already returned to Mexico with the applicant’s non-citizen spouse. The Board found the case sympathetic, noting potential hardships to the school-aged children as well as less opportunities in Mexico for the family. Nonetheless, it pointed to mitigating factors that prevented a determination of “exceptional and extremely unusual hardship.” The Board reasoned that if the applicant was removed and the remaining two US citizen children followed, the nuclear family could be reunited, the applicant could work and support the family in Mexico, the children could understand and communicate in Spanish, and the applicant’s US resident parents could survive financially without his assistance.

Similarly, the Board found no exceptional and extremely unusual hardship

---

108. Id. at 62.
109. See id. at 63.
110. Id.
111. Id.
112. Id.
113. Id. at 63-64.
114. See id.
115. Id. at 64.
116. Id. at 63-64.
117. Id. at 58, 65.
Glen: The Removability of Non-Citizen Parents and the Best Interests of 2012] REMOVABILITY OF NON-CITIZEN PARENTS 19

in the case of Matter of Andazola-Rivas.118 There, the applicant was a single mother of two school-aged US citizen children, whose entire family lived within the United States. Despite the possibility of separating the alien parent from her family if the Board denied her waiver to halt removal proceedings, it found that the facts, including the health of the children, the possession of significant financial assets, and the diminished educational and economic opportunities available in Mexico were insufficient to demonstrate exceptional and extremely unusual hardship.119

The Board did, however, find exceptional and extremely unusual hardship in the case of Matter of Recinas.120 Recinas was a single mother with six children, four of whom were US citizens. Her four citizen children had never been to Mexico and did not speak, write, or understand Spanish well. The rest of her family, including two lawful resident parents and five lawful resident siblings, also lived in the United States, and Recinas also operated her own business in the United States. Recinas had no other means to lawfully immigrate to the United States in the near future, and, with no family in Mexico, the children would be entirely dependent on her for financial and emotional support.121 The Board concluded that these factors combined to render the hardship in this case “well beyond that which is normally experienced in most cases of removal.”122 In so holding, however, the Board did hold that Recinas’ case “present[ed] a close question” and that it “consider[ed] this case to be on the outer limit of the narrow spectrum of cases in which the exceptional and extremely unusual hardship standard will be met.”123 Thus, in the absence of facts as sympathetic and compelling as those in Matter of Recinas, the Board is unlikely to grant a waiver to cancel deportation proceedings.

Adjustment of status can also provide non-citizens with effective relief from removal. To adjust status, a US citizen or lawful permanent resident files a visa petition for an alien relative. So long as the petitioner has a qualifying relationship with the alien-beneficiary, the granting of this petition could then lead to the adjustment of status of that alien to a lawful permanent resident.124 Visas for qualifying alien relatives are, on the whole, allocated based on a hierarchy of preference categories, which establish when a visa will be available to the alien-beneficiary.125 The INA provides, however, that visas are immediately available

119. Id. at 320, 323-24.
121. Id. at 469-71.
122. Id. at 472.
123. Id. at 470.
125. See 8 U.S.C. §§ 1151(a)(1), (c), 1152(a), 1153(a); 8 C.F.R. § 204.1(a).
to aliens whose petitions are filed by US citizens, as long as the alien-beneficiary has an “immediate-relative” relationship with the petitioner. 126 This category, which is not subject to the direct numerical limitations on immigrant visas, encompasses the children, spouse, and parents of a US citizen. 127 Accordingly, a non-citizen parent could adjust status based on a petition filed by a US citizen child. However, if a parent is the beneficiary of the petition, the petitioner must be at least 21 years of age—that is, a child cannot petition for his parent(s) until he has attained that age. 128 Thus, although adjustment may provide an avenue by which a non-citizen parent could obtain lawful status in the United States, a potential beneficiary may have to wait for as long as 21 years before his or her child could file a petition, from the birth of the child to his obtaining the age of majority. 129

To conclude, the INA does provide avenues for a non-citizen to obtain relief from removal by way of parentage of a US citizen child. However, relief can be difficult to obtain, even when the case involves a parent-child relationship. Cancellation of removal involves onerous demonstrations of hardship to the US citizen child, and petitions by the child for the parent’s adjustment of status often involve prolonged delay. In short, contrary to the shouts of anchor-baby opponents, who seem to believe the mere fact of having a citizen child will insulate the non-citizen parent from removal, it can hardly be said that the INA will provide ready relief to non-citizen parents of US citizen children.

Furthermore, non-citizen parents have scant opportunity to appeal adverse outcomes. The administrative judgments of the Board and immigration judge are likely to be the last word on the matter, as the INA curtails the courts of appeals’ jurisdiction to review discretionary determinations. 130 The courts of appeals retain jurisdiction to review colorable legal and constitutional challenges, 131 but this is unlikely to provide a basis for review in most cases, as the substance of the challenge will likely revolve around the agency’s weighing of the evidence—an archetypal discretionary determination. 132

IV. COMPARING THE AMERICAN AND BRITISH APPROACHES

The United Kingdom and the United States differ in how they approach the

126. 8 U.S.C. § 1151(b).
127. Id. § 1151(b)(2)(A)(i).
128. See id.
129. See, for example, id. § 1182(a)(9)(A) (inadmissibility for a variable term of years for aliens previously ordered removed); INA § 212(a)(9)(C), id. § 1182(a)(9)(C) (inadmissibility for certain aliens unlawfully present in the United States after prior immigration violations).
130. See id. § 1252(a)(2)(B).
131. Id. § 1252(a)(2)(D).
132. See, for example, Solis v. Holder, 647 F.3d 831 (8th Cir. 2011).
removability of non-citizen parents in light of the “best interests” of citizen children. These differences are highlighted by three principal considerations: (1) the different weight each country gives to the child’s citizenship in removal determinations; (2) the characterization of how removal proceedings impact the child’s interests; and (3) the differing legal regimes and international obligations in each country.

First, the United Kingdom places greater importance on a child’s citizenship than does the United States in a parent’s removal proceedings. In ZH, the child’s British citizenship was overwhelmingly the most important factor in finding the removability of ZH disproportionate. Given the facts of ZH’s case, the children’s citizenship effectively functioned as a trump card despite the UK Supreme Court’s statement to the contrary. Even under the Court’s explicit analysis, the citizenship of children will be a very significant factor in determining the removability of the non-citizen parent, a fact in large part driven by that country’s treaty obligations.

The United States, in contrast, places less weight on a child’s citizenship in establishing both the removability of the non-citizen parent and whether that parent qualifies for a waiver or some other form of relief from removal. The US Government does not deem a parent’s removal as an infringement on any of the citizen child’s constitutional rights by virtue of their citizenship. As the US Court of Appeals for the First Circuit recently held, “[t]he circuits that have addressed the constitutional issue . . . have uniformly held that a parent’s otherwise valid deportation does not violate a child’s constitutional right.” In the words of the First Circuit:

If what were happening here was conscience shocking by contemporary American standards, the lack of precedent would not bar a new departure by a lower court; but deportations of parents are routine and do not of themselves dictate family separation. If there were such a right, it is difficult to see why children would not also have a constitutional right to object to a parent being sent to prison or, during periods when the draft laws are in effect, to the conscription of a parent for prolonged and dangerous military service.

Second, the two countries differ on how a parent’s removal will impact their children. The United Kingdom takes the more pragmatic view that removal proceedings will directly impact children who will likely follow the removed parents. As the UK Supreme Court stated in ZH, “if a non-citizen parent is compulsorily removed and agrees to take her children with her, the effect is that the

---

133. See supra Part I.
134. See supra Part I.
135. Payne-Barahona v. Gonzales, 474 F.3d 1, 2, n.1 (1st Cir. 2007); cf. Newton v. INS, 736 F.2d 336, 342 (6th Cir. 1984) (“[A] minor child who is fortuitously born here due to his parents’ decision to reside in this country, has not exercised a deliberate decision to make this country his home, and Congress did not give such a child the ability to confer immigration benefits on his parents”) (citation and internal quotation marks omitted).
136. Payne-Barahona, 474 F.3d at 3.
children have little or no choice in the matter.” 137 In the United States, courts characterize children’s interests as being only incidentally impacted by the removal of the parents. For example, in Cervantes v. INS, the Tenth Circuit Court of Appeals described “the enforcement of duly-enacted conditions on aliens’ entrance and residence” as having only an “incidental impact on aliens’ minor children.” 138 Five years later in 1980, the court reinforced that view: “This Court has repeatedly held that the incidental impact visited upon the children of deportable, illegal aliens does not raise constitutional problems.” 139 These statements also demonstrate how the evaluation of impact on children informs the treatment of citizenship, and vice versa. The court’s view that any impact is only incidental makes it easier to minimize any constitutional concerns tied to the children’s citizenship.

Third, differences in the legal frameworks of both countries are one of the root causes of the different approaches taken in removability determinations. The United Kingdom’s status as a state party to the European Convention on Human Rights (ECHR) and the Convention on the Rights of the Child (CRC) dictates the country’s approach. 140 The United Kingdom has implemented those international obligations in domestic law, and ZH’s legal action itself arose as a function of the country’s regional obligations under article 8 of the ECHR. 141 The United States, in contrast, does not have any binding regional obligations that color how it approaches the removal of non-citizen parents of citizen children. It has not ratified the CRC or otherwise implemented the Convention’s provisions in its domestic law. 142 Within the United States, authorities prioritize the statutory language of the INA, and the decision to grant waivers and relief is not influenced by the CRC. As such, there is no obligation that the best interest of the child be a primary consideration in determining the removability of a non-citizen parent. As the Second Circuit Court of Appeals has held, “Congress has enacted legislation defining the circumstances under which hardship to a child may appropriately be considered as a ground for granting relief from removal[.] . . . This statute, and not the CRC, necessarily determines the outcome of [the applicant’s] request for a hardship exception to removal.” 143

Therefore, though both the United Kingdom and the United States refer to consideration of the “best interests” of the child in making certain immigration-related decisions, the concept is not identical in both countries. Comparing the UK Supreme Court’s rationale in ZH with the US Ninth Circuit Court of Ap-

138. Cervantes v. INS, 510 F.2d 89, 92 (10th Cir. 1975).
139. Delgado v. INS, 637 F.2d 762, 764 (10th Cir. 1980).
140. See supra Part I.
141. See supra Part I.
143. Oliva v. U.S. Dep’t of Justice, 433 F.3d 229, 234 (2d Cir. 2005).
peals’ decision in Cabrera-Alvarez v. Gonzalez underscores the difference. 144

Cabrera-Alvarez, a Mexican citizen who had lived in the United States for a decade prior to his removal proceedings, argued that his removal would result in exceptional and extremely unusual hardship for his two US citizen children, and that authorities should interpret the statutory hardship standard in accordance with article 3 of the CRC. 145 Specifically, he contended that the hardship standard “should be interpreted consistently with the “best interests of the child” principle by giving the child’s best interests extra weight in balancing the factors relevant to evaluating hardship.” 146

The Court rejected that argument, opining that article 3, which pertains to actions directly affecting the child, does not have relevance to a removal proceeding as it only indirectly affects the child’s interests. 147 Moreover, the Court held that the CRC clearly anticipates separation of families on account of immigration-related decisions. Article 9 of the CRC, it stated, explicitly contemplates deportation of a parent and thus separation, while article 10 requires certain efforts to keep parents and children in contact once separated—a fact that implies permissible separation under the Convention. 148 At most, the Court concluded, the Convention requires the best interests of the child to be a primary consideration as they already are within the context of the cancellation statute. 149 The Court found no warrant under the Convention or statute for the “extra weight” argument that Cabrera-Alvarez advanced. 150 According to the Court:

When an alien parent seeks cancellation of removal because of exceptional and extremely unusual hardship to a qualifying child . . . that child’s ‘best interests’ are precisely the issue before the agency, in the sense that ‘best interests’ are merely the converse of ‘hardship.’ In other words, the agency’s entire inquiry focuses on the qualifying children, making their interests a ‘primary consideration’ in the cancellation-of-removal analysis. 151

This point, the Court held, is not undercut by the fact that the agency will weigh the competing interests rather than simply dispose of the application consistent with the finding of the “best interests” of the child. 152 As the Court stated, “[t]he fact that the agency also considers the relative weight of a child’s interests does not mean that the child’s interests are not a primary consideration.” 153 Within this comparative assessment, the interests of all affected parties are equally considered, and “[a]ny interpretation that required a

144. Cabrera-Alvarez v. Gonzalez, 423 F.3d 1006 (9th Cir. 2005).
145. Id. at 1007.
146. Id. at 1010.
147. Id. at 1010-11.
148. See id. at 1011.
149. Id. at 1011-12.
150. Id. at 1012.
151. Id.
152. Id.
153. Id.
child’s best interests to be weighted more heavily than the comparative assessment would be at odds with the text of the statute.”154 In Cabrera-Alvarez, the Court did not disturb the agency’s determination that his removal would not result in exceptional and extremely unusual hardship to his citizen children. Accordingly, his removal order stood.

Comparing the Ninth Circuit’s decision with the UK Supreme Court’s decision in ZH, two points of distinction are apparent.155 First, the United Kingdom uses the best interests of the child as a threshold consideration in determining the removability of a non-citizen parent. In the United States, authorities consider a citizen child’s best interests only after they deem the non-citizen parent removable, at which point the question shifts to whether countervailing considerations exist that should negate removability. Although the best interests of the child are a relevant consideration in both jurisdictions, they are an affirmative consideration in the United Kingdom but a negative one in the United States (or, in the words of the Ninth Circuit, a “converse consideration”).156 In the United Kingdom, authorities take the citizen child’s interests into account in their own regard to argue against deportation, whereas in the United States the consideration is whether removal would unduly impact those interests. Second, although both jurisdictions contemplate a weighing of interests, UK law elevates the best interests of the child to the initial consideration. Other evidence must then rebut a presumption of action consistent with those interests. US law, however, affords the best interests of the child equal consideration with those of other interested parties.

In short, the systems differ in their foundational approach to the question of how to weigh the child’s interests. The United Kingdom approaches the issue from how removal of the non-citizen parent will impact the citizen child, while weighing other factors in that light. The United States approaches the issue from the perspective of the removable parent who has violated US immigration law, while weighing other factors and considerations under that determination.

V. THE REMOVABILITY OF NON-CITIZEN PARENTS WITH CITIZEN CHILDREN: IS THERE A BEST APPROACH?

The United States and the United Kingdom employ different approaches for determining the removability of non-citizen parents with citizen children. Is one country’s approach generally preferable to the other? The short answer is no. The approaches employed by both have pros and cons, and neither represents a “best approach” to dealing with the issue raised by this Article.

The benefits of the United Kingdom’s approach include its more pragmatic
view regarding the direct impact on children that removal of a non-citizen parent will have, its strong support for the interests of the citizen children, and its promotion of the value of family unity. Nonetheless, its approach gives insufficient consideration to the negative equities, including potentially flagrant violations of its immigration law. Despite claims to the contrary in \textit{ZH}, the UK approach effectively elevates the citizenship of the child to the status of a trump card against the removal of non-citizen parents.\footnote{See supra Parts I, III.}

The benefits of the American approach include its holistic consideration of the interests at play, especially the violation of immigration law by the non-citizen parent, and its view that relief from removal should be discretionary and only granted in sufficiently compelling circumstances. However, there are shortcomings to this approach, including its view that the removal of non-citizen parents only indirectly impacts citizen children, its onerous standard for establishing eligibility for relief from removal, and the insufficient consideration given to the child’s citizenship.

A melding of the best elements in each approach would give rise to a more nuanced and complete framework within which to weigh the competing interests of the state, citizen children, and removable parents.

First, a threshold consideration that accounts for these competing interests must be set in order to orient a more nuanced and complete immigration system. This framework must operate as a discretionary mechanism where individuals who otherwise have no right to remain in a country can do so in consideration of this balance of equities. Discretion is necessary to avoid undercutting the state’s prerogative to oversee an effective and orderly immigration system. No individual should have an automatic right to remain simply because he has compiled equities during the course of his illegal stay in a country. Making removal a discretionary determination also serves to protect against the possibilities of rampant abuse of the system. By making resolution of removal depend not solely on objective eligibility questions, the system makes every case unique and every determination case-dependent. As the US system currently makes clear, the presence of citizen children will not effectively forestall the removal of non-citizen parents. This general point should survive in a best approach framework via the characterization of the state’s decision as discretionary.

Second, a new immigration framework must also determine how to weigh the fact of the child’s citizenship. It is necessarily difficult to account for the weight of a child’s citizenship in the “best interests” calculus. It should not be elevated to the point of dictating a disposition, nor should it be an ancillary interest, as it appears to be in the context of US law.

Perhaps the simplest way to reconcile these competing conceptions is to address the direct versus indirect impact dichotomy that differentiates US and UK approaches to the best interests of the child. It is unquestionably false to as-
sert that the removal proceedings of non-citizen parents have only an indirect impact on citizen children—especially if those children are minors. The impact is best conceived of as direct, insomuch as any determination reached in the removal proceeding will inevitably affect the citizen children. If the children depart with a parent, they are leaving their primary country of citizenship, whereas if they remain in their primary country of citizenship, removal proceedings will separate the citizen child from one or both of their parents. Yet how either decision impacts citizenship is not clear. It is not the case, as the UK Supreme Court asserted, that the child forfeits the benefits of citizenship if he departs. The main characterization of citizenship is the right to enter and remain in a country indefinitely, and citizen children do not sacrifice this right if they depart with their parents to a third country. To be sure, citizen children sacrifice certain benefits associated with actually residing in that country of citizenship, but that, again, does not impact the fact of citizenship or the ongoing legal relationship the children have with their country of primary citizenship.

The question of citizenship is more abstract than whether the children remain citizens and whether citizen children forfeit certain benefits of residence upon departure. The UK approach looks to the citizen children as innocent victims of their parents’ violation of immigration law. It is true that children who obtain citizenship via the jus soli principle cannot be held accountable for their parents’ violation of the law, where this violation enabled conferral of citizenship upon the children. Yet to order removal of non-citizen parents for their violation does not constitute any violation of the children’s rights as citizens, as US courts have held.

Although the children are innocent of misconduct in simply being born in the United States, that birth, especially when it is to two non-citizen parents, is properly characterized as a fortuitous accident. At core, then, the question is how forcefully the country of citizenship should ensure the full range of interests such children may have. The short answer is, as forcefully as is consistent within the range of other interests that the removal of the non-citizen parents raises. To order the removal of non-citizen parents is not to unduly infringe on any rights of citizen children if, on the balance of equities, the interests of removability

158. Here and elsewhere the phrase “country of primary citizenship” is used to denote, in the example of the United States, the country of citizenship stemming from application of the jus soli principle. Children born in the United States are citizens thereof, as a primary matter, but presumably they would also have an opportunity to be citizens from birth of the country of their parents’ nationality and citizenship, or have an opportunity to naturalize. Dual citizenship is permissible under United States law, see Vance v. Terrazas, 444 U.S. 252, 272 (1980) (Stevens, J., concurring in part, dissenting in part), and so as long as it is permissible under the laws of the second country, children could presumably be US citizens and citizens of their parents’ native countries.

159. See, for example, ZH v. Sec’y of State for the Home Dep’t, [2011] UKSC 4, [33].

160. See Delgado v. INS, 637 F.2d 762, 764 (10th Cir. 1980); Cervantes v. INS, 510 F.2d 89, 92 (10th Cir. 1975).

outweigh those of the citizen children. Likewise, when the state exercises its
discretion to cancel the removal of a non-citizen, either through a waiver of re-
movability or an affirmative grant of relief, it is not by that act vindicating
the interests of the citizen children, but determining that on the balance of the equi-
ties authorities should allow the non-citizen to remain in the United States de-
spite an unlawful initial entry. The citizenship of the child is simply one addi-
tional interest in any proper weighing of the relevant equities.

Having resolved the issue of how citizenship itself should factor into the
determination, this leaves only the question of where the best interests of the
children should fit within the general weighing of interests in determining
whether authorities should forestall the removal of a non-citizen parent. The UK
Supreme Court, as well as the Ninth Circuit in Cabrera-Alvarez, referencing the
Convention on the Rights of the Child, recognized that the best interests of chil-
dren should be a primary consideration for authorities in determining whether to
remove a non-citizen parent. In reaching its decision in ZH, the Supreme Court
specifically rejected the assertion that the best interests should be the primary
consideration. Yet its adoption of the Australian conception of “a primary con-
sideration” from the Federal Court’s decision in Wan effectively elevates the
best interests of children to that of the primary consideration, not merely one
consideration in a grouping of primary considerations. The iteration of the rule
in ZH is that authorities must consider the best interests of the child prior to ref-
encing any other considerations. The Court did not go as far as Lord Kerr in
his opinion, where he stated that “a child’s best interests should customarily dic-
tate the outcome of cases,” but by placing them at a level above other consid-
erations in the hierarchy of the decision-making process, the Court in effect ele-
vated it from the realm of merely a primary consideration. The proper approach,
if the best interests of the child are simply a primary consideration, is to consider
those interests on equal footing with the other considerations that will be rele-
vant to determining removability. These interests are only one factor that could
argue against removal, not the primary consideration in the authorities’ determi-
nation if non-citizens should be removed, and thus there is no warrant for a
framework that requires authorities to determine and address this issue before
any other consideration. It is the US approach in this regard that is the preferable
one, then, as it considers the interests of the child within a holistic assessment of
the equities arguing both for and against removal of the non-citizen parent.

What the relevant interests are in these cases is relatively consistent across
both the UK and US immigration systems. Gauging the best interests of the
child will encompass comparing opportunities between the country of removal

162. ZH, [2011] UKSC 4 at [25] (“[D]espite the looseness with which these terms are som-
times used, ‘a primary consideration’ is not the same as ‘the primary consideration’, still less as ‘the
paramount consideration.”’).
163. Id. at [26], [33].
164. Id. at [46].
and destination country, including those pertaining to education, economics, and general social well being. Also relevant will be the health of the children, how easily the children can acclimate to life in a new culture or country, whether they have been to the destination country previously, how strong their ties to the country of removal are, how long they have lived in the country from which their parents are being removed, what family or social organization exists in the destination country, and what, if any, difficulties they may encounter in that country.

The primary interest of the state is its right to control the terms of entry, residence, departure, and removal of non-citizens. Nonetheless, the exact parameters of the state’s interest may also involve questions pertaining to economic interests and public safety; for instance, there may be a greater interest in removing a criminal alien than an alien who is simply removable by virtue of illegal presence. Authorities also consider other interests in rendering the ultimate discretionary determination relative to the applicant himself, including any criminal history, employment record in the country of removal, other avenues for legalizing his status, financial means, and immigration history.

While there is consensus between the United States and the United Kingdom about what constitutes relevant interests, the relative weight each of these interests should receive is controversial. Proportionality, hardship, exceptional and extremely unusual hardship, and extreme hardship are standards authorities use to analyze these interests. The United Kingdom utilizes a proportionality framework that gives too little weight to the state’s discretionary interest. Proportionality shifts the burden of proof from the non-citizen alien to the government: the question posed in the proportionality framework is not whether the non-citizen can marshal equities that would negate his removal, but whether the government can establish that removal is proportionate.

Though a strict interpretation of proportionality under European law would suggest that the removable alien must demonstrate that removal would not be proportionate, the case of ZH and the ECtHR decisions cited therein suggest otherwise. These decisions apply proportionality by focusing on whether government action is proportionate rather than on whether the alien can demonstrate a lack of proportionality. For instance, in ZH, the UK Supreme Court did not focus on whether the applicant could marshal facts indicating that removal was a

167. ZH, [2011] UKSC 4 at [18].  
168. Id. at [18]; In re Mendez-Moralez, 21 I. & N. Dec. at 301.  
disproportionate response to her unlawful presence, but rather on whether the
government could demonstrate compelling interests in removal given the equi-
ties of the case.\textsuperscript{170} Likewise, the ECtHR did not require any particular showing
from the applicant in the \textit{Hoogkamer} case, instead focusing on whether the gov-
ernment’s interest could overcome the interest of the mother in remaining in the
Netherlands.\textsuperscript{171} This is inconsistent with what should be the discretionary nature
of negating a non-citizen’s removal. Given the importance of discretion, the
burden should rest with the alien to establish that on consideration of all equi-
ties, the positive factors in his or her case outweigh the negative. Because au-
thorities do not apply proportionality in this sense, it is not an appropriate stand-
ard for the instant framework.

The United States weighs the relevant interests based on varying degrees of
hardship. “Simple hardship” and “exceptional and extremely unusual hardship,”
the extremes of any potential hardship sliding scale, are inappropriate as the
governing standard in determining whether authorities should cancel removal
proceedings. Showing that the citizen children may face “some hardship,” a
standard not currently countenanced by the INA’s waiver and relief provisions,
should not negate removal. Any move inevitably will result in some hardship,
and such a low showing is inconsistent with a holistic weighing of the interests.
If simple hardship would cancel removal, it is hard to see how the state’s inter-
ests or any other negative equities would factor into the ultimate determination.

On the opposite end of the spectrum, requiring that the alien establish ex-
ceptional and extremely unusual hardship to a citizen child erects a standard so
onerous that the necessity to establish extraordinary positive equities subsumes
consideration of any negative equities. Citizen aliens rarely meet this standard,
as the Board’s jurisprudence on the issue of cancellation of removal makes
clear.\textsuperscript{172} If the facts of \textit{Matter of Recinas} establish the outer limit of the circum-
stances where authorities will be countenance removal proceedings.\textsuperscript{173} Then ap-
plying a standard of exceptional and extremely unusual hardship will not effec-
tuate the need to fairly balance the competing interests when cases involve
citizen children. Concededly, Congress desired that non-permanent residents
rarely should be able to cancel their removal.\textsuperscript{174} Nevertheless, it is fair to ques-
tion whether such cancellation should be as rare in reality as Congress intended
in amending the statute to require such a high standard.

The US standard of “extreme hardship” reaches the proper balance by es-
tablishing that citizen children who face hardship above and beyond that which is
typically associated with removal, separation of a family, or beginning a life

\textsuperscript{170} See supra Part I.
\textsuperscript{171} See supra Part I.
\textsuperscript{172} \textit{In re Recinas}, 23 I. & N. Dec., 467, 469-73 (B.I.A. 2002).
\textsuperscript{173} \textit{Id.}, 23 I. & N. Dec. at 469-73.
history).
in the destination country, should forestall the removal of a non-citizen parent. It is more important for the applicant to establish that those hardships would be atypical than to establish that any hardships faced would be “extreme,” as that term has been traditionally understood in the Board’s jurisprudence. Though it is not clear that any explicit iteration of a standard is necessary in weighing the interests involved, the standards noted by the Board governing the various hardship determinations will be relevant.175 Those cases that would fall within the range of the Board’s jurisprudence on “extreme hardship” would represent the threshold for determining that resulting hardship is atypical.176 Officials must weigh any hardship considerations against potentially negative equities, including criminal convictions, under what circumstances the citizen child or children are conceived and born, and prior immigration violations. Even if the hardship interests seem to outweigh the state’s interests and any countervailing considerations, the determination of whether authorities should move forward with removal proceedings remains discretionary. This discretionary use should enable more non-citizens to successfully contest their removal than Congress intended.177 However, it should not make such success commonplace. Such action by the state is a humanitarian gesture designed to ensure family unity in a range of circumstances, not a cure-all remedy to legalize the status of individuals who have conceded brokenly broken the law and then compiled positive equities during the course of their illegal presence. Adoption of the extreme hardship standard enables a more balanced weighing of the relevant interests than the proportionality, hardship, and exceptional and extreme hardship frameworks.

In the practical application of the extreme hardship standard, some general observations are foreseeable. One set of observations involves family unity. When both parents of a citizen child are removable non-citizens, the best interests of the citizen child rarely would operate to cancel the removal of the non-citizen parents. For instance, in Matter of Monreal-Aguinaga, the applicant’s non-citizen spouse already had departed the United States with one of the couple’s US citizen children.178 The fact that one spouse was no longer in the United States and a citizen child had already left influenced the finding of no qualifying hardship.179 Family unity is better preserved if the child departs with the parents, neither of which has any lawful right to remain. Although the child is separated from other family living in the country of removal, this separation and the possibility of leaving the child with the other family members should be insufficient to argue against removal.180 Conversely, a lack of family in the desti-

175. See supra Part II.
178. Id.
179. Id.
180. See Cabrera-Alvarez v. Gonzales, 423 F.3d 1006, 1008 (9th Cir. 2005) (“[I]f a candidate for cancellation of removal could simply by stating that he or she would choose to have the child or
nation country may operate as a significant factor arguing against removal. According to the dissent in Matter of Andazola-Rivas, the removal of a single mother with two citizen children from a country where the rest of her family legally resided would seem to present the type of circumstances where authorities would cancel removal proceedings.\(^{181}\) In the practical deliberations of relevant interests, observations regarding family unity are important.

Another set of observations involves an applicant’s interactions with the country’s government. History within the immigration system, including what forms of relief have been sought, granted, or denied, and whether there have been any fraudulent claims presented, is important. While the United Kingdom’s ultimate disposition of ZH’s article 8 application seemed to disregard the applicant’s history within the immigration system, the abuse of the immigration system evident in ZH’s conduct should weigh heavily against negating removal. She fraudulently sought asylum on two occasions, authorities denied lawful status under two concessions, and she had an initial article 8 application denied.\(^{182}\) Since the ultimate disposition of any claim will rest on a weighing of interests, it is arguable on the facts of ZH’s case that despite the enormous negative equities presented, a legitimate claim arose on account of the father’s precarious health and living situation.\(^{183}\) Nonetheless, histories of fraud and failure such as this should not culminate in the discretionary conferral of a benefit by the state. General criminal conduct also should be a relevant consideration. The more egregious or heinous the underlying criminal act, the weightier this consideration should be in the final analysis. For instance, the Board properly denied a waiver of inadmissibility in Matter of Mendez-Moralez on account of the heinousness of the underlying criminal conviction, sexual assault of a child, despite the presence of other equities.\(^{184}\) Though these are not bright line rules, authorities should take these general observations into account when the extreme hardship standard operates in practice.

While the adoption of the extreme hardship standard with its general observations is desirable, it is unlikely that it that the United Kingdom and the United States will adopt this standard. In the United Kingdom, the removal proceedings emerge from binding international obligations. Changing the UK system is not simply a matter of tweaking an administrative or judicial standard but requires establishing a new framework within which to determine the removability of a certain class of alien. Considering the nature of the United Kingdom’s obligations, such a reorientation seems profoundly unlikely. Because the United King-

---

183. Id. at 3.
dom’s systems are directly influenced in their operation by the obligations it shares as party to the CRC and ECHR, it is bound to continue applying its current framework absent a change in its conception of its obligations or a shift in the jurisprudence of the ECtHR.

In the United States, removal proceedings are statutorily based as well. However, reforms along the lines noted in this article could be possible, at least as a proposition, by adjusting administrative standards. Further, to effectuate many of the aims of this article in the context of immigration reform in the United States, simplification would be the watchword. Rather than having several different forms of relief dependent upon the status of the individual, such as whether authorities charge the alien with inadmissibility or deportability and on what specific ground of removal, one overarching type of relief should be available. The statutory eligibility questions might narrow the class of individual who could apply. For instance, the commission of an aggravated felony may render an alien ineligible for relief regardless of what other equities were in play. Yet having one form of relief for qualifying aliens and application of a uniform standard would streamline consideration of relief claims. A narrowing of the grounds of inadmissibility and deportability also would have a narrowing effect on the class of aliens subject to mandatory removal. If removability focused on criminal grounds, for instance, then many non-citizen parents would not be automatically removable by virtue of a statutory standard. Reform of this kind would move the issue away from whether authorities should grant aliens relief from removal and toward whether the government should exercise discretion in instituting proceedings against aliens residing in the United States with no lawful status. The government already has undertaken a similar approach, utilizing discretion governed by objective criteria to determine whether to institute removal proceedings against removable non-citizens.\(^{185}\) Among the relevant criteria for determining whether to institute removal proceedings is whether the alien has a citizen child.\(^{186}\)

In addition to simplifying the US system, authorities could make a form of temporary relief available on a discretionary basis, such as a nonimmigrant visa available solely for non-citizen parents of minor US citizen children. Granting of the visa would be discretionary, as all such decisions are, and the visa itself would be of only temporary nature, expiring when any minor child reached the age of majority. Even in citizen families, the family begins to separate consistent with the temporal terms of the visa, with children going to college or obtaining employment often in towns and cities away from where aliens live. Expiration


\(^{186}\) Id. at 4; see Damien Cave, Crossing Over, and Over, N.Y. TIMES, Oct. 3, 2011, at A1 (discussing the Obama administration’s shift to “surgical” deportations relying on the exercise of discretion in the institution of removal proceedings).
of the visa when the citizen child turns 18 would mean that aliens could not use their visa as a bridge to lawful permanent residency via the filing of a visa petition by a 21-year-old citizen child; there would still be a period of at least three years where the parent could not lawfully reside in the United States. Yet this type of visa could ensure family unity during the formative years of a citizen child’s development and its temporary nature would serve to guard against too rampant abuse of the system. Complications would arise with multiple children, but authorities could address such issues in terms of eligibility questions and the temporal duration of any status granted.

Providing affirmative relief to non-citizens who have citizen children still may encourage non-citizens to have babies in the United States. But again, authorities can always take into account the motives of non-citizens in granting the discretionary visa. If there are indications that the sole motivation for the US birth was to obtain status in this country, that could be a negative equity in the adjudication of the visa application. Overall, the streamlining of the US system and adoption of a temporary nonimmigrant visa would enable a balancing of the relevant factors more in line with the practical application of the extreme hardship standard.

While there are costs and benefits to both the UK and the US removal proceedings of non-citizen parents of citizen children, authorities can articulate the benefits of both systems in a particular policy. The United Kingdom and the United States both agree upon the relevant interests of the parents, the citizens, and the state. They disagree as to how to weigh the relevant factors. Using an extreme hardship standard would enable a fairer balancing of the interests than is currently in place in both nations. Though the feasibility of the adoption of this policy by the United Kingdom and the United States is limited, some steps towards this end are still possible.

VI.
CONCLUSION

The issues stemming from the removal of the non-citizen parents of US citizen children are likely to persist into the distant future. The illegal population in the US is not shrinking, and even if some aliens become legal via subsequent immigration reform—for example, through an amnesty—a broad swath of citizen children still will have parents without any right to remain permanently in the United States. Nonetheless, contrary to the arguments of many who assume that a citizen child effectively will forestall the removal of the non-citizen parent, the US system currently provides for few avenues for relief and those that do exist, such as cancellation of removal, often have onerous eligibility requirements. The problem of anchor babies may be more myth than reality, but there are many serious issues involved in the deportation of non-citizen parents who have citizen children. This article has proposed a general framework for weighing the competing interests of children caught up in the immigration process via
their parents’ illegality and the interests of the United States in a fairly functioning, efficient, and respected legal system. Such balancing inevitably will separate families, at times, and, at others, cancel the removal of aliens who have broken the laws of this country. No party is likely to be happy with the course of this balancing all the time, yet this is the reality of what is possible and what could prove acceptable to the relevant stakeholders.

Immigration reform may help to alleviate some of these issues by, for instance, legalizing status under certain conditions. This will shrink the population of citizen children whose non-citizen parents would be subject to removal. Reform also can more directly address this issue by focusing on the ends that US immigration law and policy should strive to obtain. If family unity is one such end, then there are compelling arguments that the relief provisions of the INA, as they currently stand, do not adequately address the factors that will be relevant to ensuring family unity—such as the best interests of US citizen children. In any event, this issue deserves serious consideration and engagement in the coming years as reforms are proposed, debated, and, hopefully implemented.
2012

Fictitious States, Effective Control, and the Use of Force Against Non-State Actors

Brian Finucane

Recommended Citation
Available at: http://scholarship.law.berkeley.edu/bjil/vol30/iss1/2

This Article is brought to you for free and open access by the Law Journals and Related Materials at Berkeley Law Scholarship Repository. It has been accepted for inclusion in Berkeley Journal of International Law by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.
Fictitious States, Effective Control, and the Use of Force Against Non-State Actors

By
Brian Finucane*

This Article examines how states respond to violent non-state actors operating from “fictitious” states. Fictitious states possess international legal personality but they lack effective control over their territories and populations. Examples of fictitious states include Pakistan, Yemen, and Somalia. These entities are not states that “failed” but territories where the paradigm of traditional statehood is inapplicable.

In contrast to much of the literature analyzing the use of force vis-à-vis contemporary threats, this Article contends that the global security problem posed by fictitious states is more fundamental than that posed by terrorism or failed states. The modern threat emanating from fictitious states is most vividly illustrated by Al Qa’ida; however, violent non-state actors have long exploited territories beyond the writ of any central government. Often, when threatened by these non-state actors, victim states have responded with transborder force.

This Article discusses key incidents over the past two centuries that elucidate the extensive and under-appreciated history of a state’s right to exercise defensive force against non-state actors in ungoverned territory. Such incidents include US intervention in Spanish Florida, British intervention in New York State, and Russian intervention in Mongolia. These incidents show that: (1) there is a well established customary right of self-defense of victim states that is not contingent upon the consent of host states; and (2) this customary right was preserved by the United Nations (UN) Charter.

The challenge for the international order is to identify the conditions under

* Yale Law School, J.D. 2010; Oxford University, DPhil, 2008; Cornell University, A.B. 2003. The author is an attorney-adviser at the U.S. Department of State. This Article was written prior to the author’s employment at the Department of State and any views expressed herein are solely those of the author, and are not necessarily those of the United States or the Department of State. The author is grateful to Sigrun Kahl, Oona Hathaway, Sean Murphy, and Ashley Deeks for their comments on earlier drafts.
which a victim state may act in self-defense against a non-state actor. The state practice and legal claims investigated in this Article demonstrate that the principles governing such action are the same as those that apply to state-to-state self-defense: necessity and proportionality. This Article explains how these principles structure the use of force against non-state actors and delineates the battlefield in a conflict between a victim state and a non-state actor, such as the United States' conflict with Al Qa’ida.

INTRODUCTION

Targeted killing of Al Qa’ida fighters in the ungoverned regions of Pakistan, Yemen, and Somalia exposes a lacuna in the treaty-based regime governing the use of force.1 If, as the International Court of Justice claims, the UN Charter’s prohibition on the use of force is the “cornerstone” of the modern international system, this cornerstone rests on a foundation of sand.2 The Charter embodies a security framework intended to regulate relations between effective states—states possessing effective control over people and territory. Taken by itself, such a state-centric regime is incomplete. However, this treaty-based regime is supplemented by a body of customary law regulating the use of force against non-state actors operating from ungoverned territory. This Article examines this body of customary law, focusing on the principles that can be drawn from state practices and legal claims over the past two centuries. These customary principles define, in part, the effective state’s inherent right of self-defense.

Effective states exercising control over their nominal territories and populations, i.e., “positive sovereignty”, are not the natural form of political organization.3 The ideal of the Westphalian state-system posits adjacent territorial entities, each ruled by central authorities exercising effective control over populations and territories within defined borders. This exists, most of the time, in much of northern Eurasia and in neo-European settlement colonies, such as the United States, Australia, and Canada.

However, this model is the exception rather than the rule in much of the

---

1. U.N. Charter art. 2, para. 4 (“All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations”).


3. GEORG SCHWARZENBERGER, A MANUAL OF INTERNATIONAL LAW 38 (1950) (describing positive sovereignty as the “claim to have control over persons, things and territory”); ROBERT JACKSON, QUASI-STATES, SOVEREIGNTY, INTERNATIONAL RELATIONS AND THE THIRD WORLD 11, 29 (1990) (discussing the concept of positive sovereignty and its relationship with negative sovereignty).
rest of the world. Fictitious states” lack central authority capable of exercising effective control over a substantial fraction of the territory and population within their internationally recognized boundaries, making their sovereignty a legal fiction. Nonetheless, fictitious states are entitled to the legal right to non-interference, or “negative sovereignty.” Fictitious states include many of the post-colonial entities of sub-Saharan Africa (e.g., the Democratic Republic of Congo), as well as countries in Latin America (e.g., Peru), and South and Central Asia (e.g., Pakistan). In extreme cases, fictitious states, such as Somalia, may control a seat at the United Nations but virtually none of their own territory.

In contrast to much of the literature analyzing the use of force vis-à-vis contemporary threats, this Article contends that the global security problem is more fundamental than terrorism or failed states. To analyze the security dilemma facing governments as one of “terrorism” is a misdiagnosis. The global security problem is not limited to a specific tactic, whether employed by state or non-state actors. Nor is the problem limited to the exceptional cases of completely “failed” states, such as Somalia. Indeed, the notion of a “failed” state presupposes, usually incorrectly, that an effective state existed or should have existed in the first place. Instead, the fundamental challenge to international order is that the constitutive unit of the international system is a fiction in many parts of the world. Perhaps a majority of the entities with seats at the United Nations do not control some or all of their territory and population. The effective state and the monopoly of the state over internationally significant violence cannot be taken for granted.

The disjunction between positive and negative sovereignty that characterizes fictitious statehood represents a persistent threat to the international order. The US 2002 National Security Strategy emphasized that “America is now

---


5. This term overlaps with Jackson’s term “quasi-state” but is broader. See JACKSON, supra note 3, at 1. I use the term fictitious state to emphasize that their reality is a legal fiction.

6. SCHWARZENBERGER, supra note 3, at 38 (“In its negative aspect State sovereignty means independence from outside interference”).


9. See JACKSON, supra note 3; HERBST, supra note 4; Clapham, supra note 4; CENTENO, supra note 4.
threatened less by conquering states than we are by failing ones." Secretary of Defense Robert Gates reiterates this conclusion by noting that “[i]n the decades to come, the most lethal threats to the United States’ safety and security . . . are likely to emanate from states that cannot adequately govern themselves or secure their own territory. Dealing with such fractured or failing states is, in many ways, the main security challenge of our time.”

This threat has only increased with the political upheaval that has accompanied the Arab Spring. The fall of former Egyptian President Hosni Mubarak’s regime has led to lawlessness in the Sinai Peninsula that anti-Israeli fighters have exploited. Yemeni militants have taken advantage of the weakening of the Saleh regime to expand their territorial control. Most worrying for the United States, the fall of Colonel Muammar Gaddafi and the failure of the Libyan rebels to adequately exercise control of the country has led to the dispersal of an unknown quantity of Man Portable Air Defense Systems (colloquially known as MANPADs) from Libyan arms depots.

The collision between the legal fiction and the reality of statehood reveals that, taken by itself, the UN Charter’s regime governing the use of force is incomplete. Such a framework does not adequately promote global security because it does not account for fictitious states and non-state actors. Fictitious


13. Laura Kasinov, In Yemen One Islamist Dead, But the Battle Goes On, N.Y. TIMES, Oct. 2, 2011, available at http://www.nytimes.com/2011/10/03/world/middleeast/in-yemen-one-islamist-dead-many-more-in-arms.html?_r=1&scp=12&sq=yemen%20+%20saleh%20+%20control&st=cse (noting that Islamic militants “control large areas of territory in the country’s restive south”); see also ANGEL RABASA, ET AL., UNGOVERNED TERRITORIES: UNDERSTANDING AND REDUCING TERRORISM RISKS (2007) (a RAND Corporation analysis of the security threats present in several ungoverned areas for the United States Air Force); A Lonely Master of a Divided House, THE ECONOMIST, Apr. 24, 2010, at 45 (describing Yemen in the context of Al Qaeda activity and multiple insurgencies as “famously hard to govern. Yet even if the power of Mr. Saleh’s state has seldom extended beyond Yemen’s main towns, roads and oilfields, it is remarkable he has maintained even a semblance of control”).

states are incapable of binding through international agreements the territory and people over whom they lack effective control. There is a need for a framework or residual mechanism that accounts for the existence of ungoverned territory. As this Article explains, such an effective, complementary regime already exists in customary international law.

This Article proceeds in four parts. Part I briefly analyzes the state-centric use of force principles embodied in the UN Charter. In particular, it examines the concepts of statehood and sovereignty, and, especially, the effective control of people and territory upon which the Charter rests.

Part II explains why a strict state-centric regime has never been viable. Drawing upon empirical insights from comparative politics and anthropology, it demonstrates that the traditional statehood of international law is nothing more than a legal fiction in much of the world. Part II argues that the disjunction between the state-centric security regime and the current threat environment has been exacerbated by the redefinition of statehood during decolonization, when statehood became an entitlement rather than a factual condition. This Part concludes with an analysis of the International Court of Justice’s decision in Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), 2005 I.C.J. 168 (Dec. 19). Armed Activities serves as a case study of the inadequacy of a state-centric security paradigm in a world of fictitious states and non-state actors.

Part III places contemporary security threats in historical and legal perspective by examining the extensive and under-appreciated history of cross-border defensive measures premised upon state ineffectiveness. This Part shows that, notwithstanding the state-centric regime of the UN Charter, many interventions against non-state actors were accepted as lawful by the governments of major powers even before 9/11. It examines in detail the legal claims made by governments and demonstrates that these claims both explicitly and implicitly invoke state weakness as a basis for intervention.

Part IV proposes a framework for governing the international use of force that balances the danger posed by interstate conflict and opportunist intervention with the growing threat posed by non-state actors and ungoverned territories. Drawing upon pre-Charter state practice and opinio juris, Part IV details how customary principles of necessity and proportionality should structure the use of force against non-state actors in the future. It explains how these principles establish geographic restrictions upon the recourse to force and why differentiation between state and non-state actors is necessary. It concludes with a reconsideration of the armed attack requirement of Article 51 of the UN Charter.
Berkeley Journal of International Law, Vol. 30, Iss. 1 [2012], Art. 2

BERKELEY JOURNAL OF INTERNATIONAL LAW

[Vol. 30:1

I. THE ASSUMPTION: THE STATE-CENTRIC REGIME

The victorious Allies created the United Nations to “maintain international peace and security,” that is, peace and security between states. To this end, the UN Charter imposes a general prohibition in Article 2(4) on the use of force by Members against other states.

Both explicitly and implicitly, the Charters of the United Nations and NATO embody a state-centered use of force paradigm. Drafted in the final months of the interstate conflict of the Second World War, the UN Charter reflects international efforts to prevent a repeat of the then ongoing war. The interstate focus of the Charter is clear in its Preamble, which emphasizes the horrors of the World Wars. The peace and security at stake was international (used here to mean between two or more states).

Article 51, however explicitly preserves a narrow exception to the prohibition on Members’ unilateral use of force against other states: “Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations….” The scope of this exception has been vigorously contested ever since, especially as concerns its application to non-state actors.

This Article argues that the customary right of self-defense preserved by Article 51 unequivocally encompasses defensive action taken against non-state actors. These customary principles of self-defense supplement and complement the state-centric regime embodied in the UN Charter.

However, before turning to the scope of the exception in Article 51, it is

15. U.N. Charter art. 1, para 1 (“We the peoples of the United Nations determined to save succeeding generations from the scourge of war, which twice in our lifetime has brought untold sorrow to mankind . . .”).

16. U.N. Charter art. 2, para. 4 (“All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations”).


18. U.N. Charter, pmbl.; Ann-Marie Slaughter & Bill Burke-White, An International Constitu-
tional Moment, 43 HARV. INT’L L. J. 1,1 (2002) (“The framers of the U.N. Charter were responding to two worlds wars, countless interstate wars, and indeed centuries in which the primary threat to international peace and security was the aggressive use of force by one state against another”).

19. IAN BROWNLIE, INTERNATIONAL LAW AND THE USE OF FORCE BY STATES 279 (1963) (arguing that even if a non-state actor could mount an armed attack, “[t]he incursions of armed bands can be countered by measures of defense which do not involve military operations across frontiers”).

20. U.N. Charter art. 51 (emphasis added).

A. Effectiveness: The Essence of Statehood in International Law

Underlying the Charter’s state-centric security framework is the assumption that, within each internationally recognized state exists some central authority in control of all internationally significant armed forces within that territory. This assumption stems from the traditional conception of statehood. During the late nineteenth and early twentieth centuries, statehood was a matter of fact. The critical quality of a state from the standpoint of internal and external security was effectiveness.

State effectiveness is defined by: 1) effective control of territory, and 2) effective control of people. The effective control of territory and the exclusion of external private armed groups is primarily a function of a state’s coercive capability. Such capability is measured by the strength of its police and military forces vis-à-vis non-state actors, the latter seeking a safe haven within the state’s territory for themselves. The effective control of people implies “social control” and “effective authority” and in part a function of a state’s coercive capability. However, control in all but the most authoritarian entities is also contingent upon legitimacy and shared identity. A state’s effective control of people prevents the indigenous development of independent armed groups. Effective control of people and territory are mutually reinforcing as a state’s legitimacy enhances its coercive capabilities and vice versa.

22. J.L. BRIERLY, THE LAW OF NATIONS 137 (6th ed. 1963) ("Whether or not a new state has actually begun to exist is a pure question fact").

23. JAMES CRAWFORD, THE CREATION OF STATES IN INTERNATIONAL LAW 77 (1979) ("The traditional criteria for statehood were based almost entirely on the principle of effectiveness. The proposition that statehood is a question of fact derives strong support from this equation of effectiveness and statehood. In other words, although it is admitted that effectiveness in this context is a legal requirement, it is denied that there can exist legal criteria for statehood not based on effectiveness").

24. Id. at 42 (Explaining that traditionally “international law defines ‘territory’ not by adopting private law analogies of real property, but by reference to the extent of governmental power exercised, or capable of being exercised, with respect to some area and population”).

25. JOEL MIGDAL, WEAK STATES, STRONG SOCIETIES 22-23 (1988) ("State social control involves the successful subordination of peoples own inclinations of social behavior or behavior sought by other social organizations in favor of the behavior prescribed by state rules . . . . Getting the population to obey the rules of the state rather than the rules of the local manor, clan or other organization").

26. H. Lauterpacht, Recognition of States in International Law, 53 YALE L. J. 385, 410 (1943-1944) ("The second essential requirement of statehood is a sufficient degree of internal stability as expressed through the functioning of a government enjoying the habitual obedience of the bulk of the population").

The significance of effectiveness as the signature element of statehood is readily apparent in the words and deeds of the statesmen, jurists and legal scholars of the late nineteenth and early twentieth century. During the nineteenth century, governments contemplating the recognition of aspirant entities emphasized effectiveness as a crucial empirical prerequisite for statehood. In determining whether newly independent Mexico qualified as a state, British Foreign Secretary Canning focused upon two aspects of effectiveness. First, he questioned whether the Mexican government was “in military possession of the country, and also whether it was in a respectable condition of military defense against any probable attack from Europe” (e.g., territorial control). Second, he questioned whether the Mexican government had “acquired a reasonable degree of consistency” and enjoyed “the confidence and goodwill of the several orders of the people” (e.g., control over people). US President Grant also emphasized control over the population as a prerequisite to statehood when he refused to recognize Cuba’s independence until there was “some known and defined form of government, acknowledged by those subject thereto, in which the functions of government are administered by usual methods, competent to mete out justice to citizens and strangers, to afford remedies for public and for private wrongs, and able to assume the correlative international obligations and capable of performing the corresponding international duties resulting from its acquisition of the rights of sovereignty.”

Pre-UN international organizations also explicitly and implicitly understood statehood in terms of effectiveness. In assessing whether a mandatory territory qualified as a state, the League of Nations identified the following criteria in 1931: 1) a settled government and an administration capable of maintaining the regular operation of essential government services, 2) capacity to maintain its territorial integrity and political independence, 3) capacity to maintain peace throughout the territory. Article 1 of the Montevideo Convention of 1933 provides the standard requirements for statehood and personality in international law: “a) a permanent population, b) a defined territory, c) government, and d) capacity to enter into relations with other states.”

28. Although this Article focuses on the existence of states as subjects of international law, rather than the recognition of new states, the recognition of a new state presupposes that a new state exists. Thus, changes in the recognition of states reflect changes in the understanding of statehood.


32. Convention on Rights and Duties of States, Dec. 26, 1933, 165 L.N.T.S. 19; see also RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES § 201 (1987) (defining a state as “an entity that has a defined territory and a permanent population, under the control of its own government, and that engages in, or has the capacity to engage in, formal relations with other such entities”); CRAWFORD, supra note 23, at 36 (“It is a characteristic of these criteria . . . that they are based on the effectiveness among territorial units”).
Jurists of the early twentieth century considered effectiveness the *sine qua non* of statehood as well. Following Finland’s independence from Russia in 1917, a Commission of Jurists appointed by the League of Nations observed that Finland did not achieve statehood immediately upon independence:

> [F]or a considerable time, the conditions required for the formation of a sovereign State did not exist. In the midst of revolution and anarchy, certain elements essential to the existence of a State, even some elements of fact, were lacking for a fairly considerable period. Political and social life was disorganized; the authorities were not strong enough to assert themselves . . . the Government has been chased from the capital and forcibly prevented from carrying out its duties . . . . It is therefore difficult to say at what exact date the Finnish Republic, in the legal sense of the term, actually became a definitely constituted sovereign State. This certainly did not take place until a stable political organization had been created, and until the public authorities had become strong enough to assert themselves throughout the territories of the State without the assistance of foreign troops.33

Leading treatises of international law from the nineteenth and early twentieth century enumerate similar criteria. Phillimore defines a state as “a people permanently occupying a fixed territory, bound together by common laws, habits and customs in one body politic, exercising, through the medium of an organized Government, independent sovereignty and control over all persons and things within its boundaries.”34 According to Wheaton, the “legal idea of a State necessarily implies that of the habitual obedience of its members to those persons in whom the superiority is vested.”35 Lauterpacht states that “the requirements of statehood as laid down by international law and as uniformly expressed in text-books, [are] namely, the existence of an independent government exercising effective authority within a defined area.”36 According to Lawrence, a state is “political community, the members of which are bound together by the ties of a common subjection to some central authority, whose commands the bulk of them habitually obey.”37 Noting that the “sovereign state is the typical subject of international law,”38 Schwarzenberger states that “it has become customary to assume that a subject of international law must have a stable government, which does not recognize any outside superior authority, [and] that it must rule supreme within a territory.”39

33. *Aaland Island Dispute*, LEAGUE OF NATIONS O. J., Special Supplement, No. 4, 8-9 (1920) (emphasis added).
34. 1 ROBERT PHILLIMORE, COMMENTARIES UPON INTERNATIONAL LAW 94 (1854) (emphasis added).
35. HENRY WHEATON, ELEMENTS OF INTERNATIONAL LAW 26 (8th ed. 1866).
36. Lauterpacht, supra note 26, at 408.
38. SCHWARZENBERGER, supra note 3, at 122 (lists the essential characteristics of the state as “an organized government, a defined territory, and such a degree of independence of control by any other state as to be capable of conducting its own international relations”).
39. Id. at 31 (emphasis added); See also HANS KELSEN, PRINCIPLES OF INTERNATIONAL LAW 100-102, 108 (1952) (describing the state as a centralized coercive legal order possessing a monopo-
Two additional points should be noted about the role of effectiveness in statehood. First, the temporary loss of effective control due to insurgency or beligerent occupation did not necessarily compromise statehood in the eyes of other governments. As Part III illustrates, temporary lapses of control were common and other governments tolerated these lapses. However, chronic ineffectiveness was a different matter. When effective control was the exception rather than the rule within a territory the existence of a state could not be taken for granted. Second, though statesmen may have promoted policy aims by conditioning the recognition of states upon additional normative considerations, such as the slave trade, religion, the degree of civilization, or the existence of a democratic government, these criteria were supplements to—not substitutes for—the factual prerequisite of effective control.

In sum, from the standpoint of international law, the key criteria of effective, factual statehood were not the existence of an impersonal technocratic bureaucracy, the character of the state’s legal order, or other components of Max Weber’s classic formulation per se. Rather, the most important criteria were the control over violence within a defined territory. These other features are relevant only to the extent that they enable a state to exercise effective control over people and territory. As summarized by Fukiyama, “[t]he essence of stateness is, in other words, enforcement: the ultimate ability to send someone with a uniform and a gun to force people to comply with the state’s laws.”

B. Sovereignty: Positive vs. Negative

The *sine qua non* of traditional effective statehood in international law is effective control or positive sovereignty. An entity’s positive sovereignty is a function of the “capabilities which enable governments to be their own masters: it is a substantive rather than a formal condition.” Secretary of State Lansing considered such positive sovereignty, to have the following attributes.

1) Sovereignty is *real* (or *actual*) only when the possessor can compel the obedient—


42. See JACKSON, supra note 3, at 29 (describing the related concept of empirical statehood). I employ the term effective state to emphasize the key attribute of statehood, effective control of people and territory.

43. JACKSON, supra note 3, at 29; THEODORE DWIGHT WOOLSEY, INTERNATIONAL LAW 35 (6th ed., 1897) (“By sovereignty we intend the uncontrolled exclusive exercise of powers of the state; that is . . . the power of governing its own subjects”); see also STEPHEN D. KRASNER, SOVEREIGNTY: ORGANIZED HYPOCRISY (1999) 3-4 (distinguishing between different forms of sovereignty); CRAWFORD, supra note 23, at 42 (Explaining that traditionally “international law defines ‘territory’ not by adopting private law analogies of real property, but by reference to the extent of governmental power exercised, or capable of being exercised, with respect to some area and population. Territorial sovereignty is not ownership of, but governing power with respect to, territory”).
ence to the sovereign will of every individual composing the political state and within the territorial state.
2) Such complete power to compel obedience necessarily arises from the possession of physical force superior to any other such force in the state.
3) The exercise of sovereignty in a state does not involve reasonableness, justice or morality, but is simply the application or the menace of brute force. Positive sovereignty is a political fact. It exists in contradistinction to negative sovereignty, which is a legal right to non-intervention and non-interference.

The state-centric security paradigm embodied in the UN Charter came into existence at a time when normative principles such as non-intervention flowed from factual precedents. “Independence and territorial as well as personal supremacy are not rights, but recognized and therefore protected qualities of states as International Persons.” Negative sovereignty is the “[r]espect for the inviolability of the territory of a State and rests on the theory that it possess the power and will to exercise control therein.” Thus the internal might of positive sovereignty entitled one set of governing elites to the external right of non-intervention vis-à-vis external governing elites.

C. The One Army Rule of Statehood

From the traditional indicia of sovereign statehood I derive what I term the “one army rule” of sovereign statehood. A single, supreme military force within a defined territory characterizes a sovereign state. A state cannot enjoy a monopoly on force if there is another, independent armed force within its territory. If a state voluntarily or involuntarily shares its territory with an independent armed group, its positive sovereignty is compromised.

The logic of the UN Charter’s use of force regime is contingent upon the one-army rule. A state-centric regime is tenable only as long as the putative state authorities exercise effective control over their nominal territory and any military forces within this territory. A state-centric framework is viable if governments can bind all the significant armed groups in a regime of non-aggression and non-intervention vis-à-vis other states. Violation of the one-army rule not only compromises an entity’s sovereignty, but also undermines a state-centric

44. Robert Lansing, Notes on Sovereignty in a State, 1 AM. J. INT’L L. 105, 110 (1907) (emphasis added).
45. JACKSON, supra note 3, at 1, 50-53.
47. 1 CHARLES CHENEY HYDE, INTERNATIONAL LAW, 646 (1945).
48. See W. Michael Reisman, Private Armies in a Global War System, in LAW AND CIVIL WAR IN THE MODERN WORLD 252, 256-69 (John Norton Moore ed., 1974). Professor Reisman discusses the related concept of the “private army rule,” that is, the traditional intolerance by the international community for private armed groups and the strict attribution of their violence to their territorial host. I use of the term “one army rule” to highlight the implications for sovereignty and the ideal of statehood posed by the presence of non-state armed groups.
49. This is a necessary but not sufficient condition for statehood.
security regime.

As described in the next Part, the ideal of sovereign statehood is rarely realized and the one army rule is often violated in practice. This gap between the ideal and reality of statehood is a longstanding and persistent fact. Rather than a world divided between contiguous states, each containing a single armed force, we have long lived in a world of states, fictitious states and violent non-state actors. In this world, the internationally recognized territorial ruler is often not the ruler of the territory.

II. A GROWING PROBLEM: FICTITIOUS STATEHOOD

Taken by itself, a state-centric security regime is inadequate to regulate the use of force and preserve minimal public order because the two assumptions upon which the regime is premised are invalid. First, weak states are the norm rather than the exception in much of the world. Second, violent non-state actors possess significant military capabilities often comparable to those of state militaries and thus pose internationally significant security threats. The large number of conflicts involving weak states and transnational non-state actors alone demonstrates the inadequacy of the state-centric regime.

This Part explains how and why the predicate conditions for an exclusively state-centric security regime have never existed, and why state-centrism became even less tenable in the wake of decolonization. I then use the case study of the Democratic Republic of the Congo (DRC) to illustrate both the security threat resulting from the absence of effective central authority and the response of one major international body, the International Court of Justice, to the problem of fictitious statehood.

A. Effective Statehood is the Exception

From a global perspective, states exercising effective control over their territory and the people within that territory are anomalous. This fact, long appreciated by social scientists, is often neglected by legal scholars and policymakers.

50. See Jackson, supra note 3; Herbst, supra note 4; Clapham, supra note 4; Centeno, supra note 4.


52. See, e.g., David Nugent, Building the State and Making the Nation: The Bases and Limits of State Centralization in Modern Peru, 96 AM. ANTHROPOLOGIST 333, 335 (1994) (reviewing the social science literature discussing differences between contemporary Western and non-Western states); Thomas Blom Hansen & Finn Stepputat, Introduction, in STATES OF IMAGINATION: ETHNOGRAPHIC EXPLORATION OF THE POSTCOLONIAL STATE 2 (Thomas Blom Hansen & Finn Stepputat, eds., 2001) (noting that the “myth of the state seems to persist in the face of everyday experiences of the often profoundly violent and ineffective practices of government or outright collapse of states. It persists because the state, or institutional sovereign government, remains pivotal in
Finucane: Fictitious States, Effective Control, and the Use of Force Againsts

Although there is awareness among policymakers of the problem that “failed states” pose, the term “failed state” itself presupposes the prior existence of a state. It thus reflects a fundamental misunderstanding of the realities of political power around the globe. For the most part, the regions beyond the writ of any central government are not states that failed. They are instead regions where effective states never existed.

This is not to say that states do not exist outside of Europe and its settlement colonies. Rather, where states exist, they often do not exercise effective control over the territory within their borders. As in Afghanistan, the governments of many countries have employed the “Swiss cheese approach” to governance. The state controls only the most populated and economically valuable areas and leaves autonomous the populations of the economically/environmentally marginal and difficult to control regions, such as mountains and deserts, so long as they do not challenge the central authorities. Put differently, the writ of the state is not congruent with the international borders of the country it occupies. Sometimes the states within such territories are little more than city-states.

Examples of such fictitious states abound. “Many other regions of the world share the African experience of having significant outlying territories that are difficult for the state to control because of relatively low population densities and difficult physical geographies.” In Latin America,

The state’s capacity to maintain monopoly over the use of violence or territoriality has also always been suspect. With a couple of exceptions, few national capitals could be said to have ruled the hinterlands of the nineteenth or even early twentieth century. Even today, Peru, Ecuador, and Bolivia still lack the ability to control the Sierra; Mexico continues to fight rebels in at least two provinces; Brazil cannot enforce federal policies on regions; and Colombia is quickly disintegrating.

Fictitious states span a broad range of capabilities. Fictitious states in-
clude entities such as Somalia, which lack any central authority. However, even territorial entities such as India that contain strong states are fictitious to the extent that their borders are not coterminous with the writ of their central governments. The extent of an entity’s statehood is fictitious to some degree whenever the central government is unwilling or unable to establish control over some portion of its territory.

A state may possess formidable military capabilities yet be unable to exercise effective control over its own territory. In Sudan, the al-Bashir government in Khartoum is capable of committing atrocities in Darfur but incapable of compelling compliance with its commands throughout much of Sudanese territory. The quintessential example is Pakistan, which possesses a nuclear arsenal but cannot control its Federally Administered Tribal Areas. The ability of territorial elites to destroy is not the ability to control. The central authorities of fictitious states may be dangerous internally and externally, yet they may still be ineffective. Prior to the twentieth century even the United States faced difficulty in suppressing violent non-state incursions from its territory into Canada, Mexico, and Central America.

The effective state is neither the default form of government nor a natural function of territory. They are a form of political organization and therefore aspects of human culture. The state is a set of relationships and understandings relating to authority, compliance, loyalty, and identity.

The political relationships characterizing the state are historically contingent. There is no reason to assume that the complex historical processes lead-
Finucane: Fictitious States, Effective Control, and the Use of Force Again

ing to the formation of the effective state, as understood by Charter-era interna-
tional law, played out the same everywhere in the world. Extensive scholarship
in the social sciences indicates that the environmental as well as the resulting prehistoric and historic processes leading to the in situ development of effective states in parts of Eurasia did not occur in all re-
gions of the world, particularly Latin America and Africa. The development of
the effective state was far from universal. Moreover, future development to-
wards the Euro-centric model of effective statehood is unlikely. The differing
conditions and developmental trajectories of different areas of the world result
in different forms of political organization, not all of which can be characterized
as effective states.

To summarize, many of the territorial communities recognized as states are
legal fictions to some degree. Effective states exercising direct, relatively ho-

gogenous effective control throughout their internationally-recognized borders
are anomalous in much of the world. The next section traces how the primarily
European phenomenon of effective statehood came to be universalized into a
principle of justice.

B. Statehood as Norm, Not Fact: Reification of Fictitious Statehood

The reality of statehood is not universal. However, the ideal of—indeed the
right to—statehood gained widespread acceptance in the mid-twentieth century.
This Section examines the transformation of statehood from an empirical fact
into a principle of justice.

1. The Delegitimization of Formal Political Inequality

The principles of self-determination and human equality gained ascendance
following the Second World War. Western statesmen generalized liberalism

Prior to that time human populations in Europe where the modern state was invented and elsewhere
organized themselves politically along rather different institutional lines . . . ”).

64. See JARED DIAMOND, GUNS, GERMS AND STEEL (1997); BARFIELD, supra note 54;
AFGHANISTAN: A CULTURAL AND POLITICAL HISTORY 67-70 (2010); SCOTT, supra note 52.
65. Id.
66. TILLY, supra note 63 (describing the factors that led to the development of the “national state”, analogous to the effective state, in Europe but not in other regions of the world).
67. CENTENO, supra note 4; HERBST, supra note 4.
68. Clifford Geertz, What is a State If It is Not a Sovereign: Reflections of the Politics of Com-
plicated Places, 45 CURRENT ANTHROPOLOGY 577, 578 (2004) (“[S]o far as state formation . . . is
concerned, whatever has already happened in supposedly better-organized places is less prologue
than chapters in a different sort of story not to be reenacted. Whatever directions what is called . . .
“nation building” may take in Africa, the Middle East, Asia, or Latin America, a mere retracing
without the bloodshed of earlier cases—England, France, or Germany, Russia, the United States, or
Japan—is not in the cards . . . ”).
69. JACKSON, supra note 3, at 16 (“[D]omestic ideologies promoting enfranchisement of racial
and ethnic minorities in Western states” reinforced such international equality norms.)
among individuals to liberalism among societies.\textsuperscript{70} If all peoples are equal, then they must all have the same capabilities with respect to political organization.

2. Decolonization

The transformation of statehood and sovereignty from empirical realities into international norms was largely a product of decolonization. Major powers recognized that colonial peoples were entitled to statehood as a right.\textsuperscript{71} The international community adhered to and enforced principles of “equal rights and self determination of peoples,”\textsuperscript{72} as well as territorial integrity\textsuperscript{73}, \textit{uti possidetis}, and non-intervention. The result of decolonization was the creation of entities with international legal personalities irrespective of their internal capabilities.

During decolonization, indigenous elites, particularly those educated in the metropolis, embraced the ideal of European-style statehood because it was “modern,”\textsuperscript{74} and also because it was the only means by which they could enjoy formal equality \textit{vis-à-vis} European elites. If states were the principal entities of the world community, then indigenous elites needed states to stand on equal footing with European colonial powers. Through the process of decolonization, the European Westphalian state system became globalized.

However, most of these soon-to-be independent colonial entities had never been states prior to colonization, or at least not states defined by their external colonial frontiers. European colonial rule did not usually result in the formation of effective states either. The exogenous creation of effective states, even if possible, had not been the program of most European colonial powers.\textsuperscript{75} Yet, during decolonization many governments apparently “assumed that the new, young countries would ultimately develop into carbon copies of the European and North American states.”\textsuperscript{76} This assumption has proven to be generally unfounded.\textsuperscript{77} Herbst’s observation relating to Africa is germane to many other areas of the world, particularly the Middle East and South and Central Asia, from which

\begin{thebibliography}{9}
\bibitem{70} \textit{Id.} at 14 (“The constitutional leveling that occurred within Western domestic societies has taken place internationally and for most of the same reasons which have to do with the doctrine of equal rights and equal dignity of all mankind”).
\bibitem{71} \textit{Id.}
\bibitem{72} U.N. Charter art. 1, para 1.
\bibitem{73} \textit{Id.} at art. 2, para 4.
\bibitem{74} \textit{Herbst, supra} note 4, at 99-101.
\bibitem{75} \textit{See generally, Crawford Young, The African Colonial State in Contemporary Perspective} (1994).
\bibitem{77} \textit{Id.} at 79 (“The universalization of the territorial state format does not mean that all states share the same characteristics. In particular, artificial states—the creation of colonial authorities and international organizations—are in many ways fundamentally different from states that grew slowly through organic processes involving wars, administrative centralization, the provision of welfare environments, and the development of national identities and sentiments”).
\end{thebibliography}
many contemporary threats emanate. 78 “[I]nternational society, by dint of the granting of sovereignty, still assumes that all African countries are able to control all of the territory within their boundaries. The gap between how power is exercised in Africa and international assumptions is significant and, in some cases, growing.” 79

In the societies traditionally characterized by effective states, statehood and positive sovereignty were taken for granted. Many Western policymakers have blithely assumed that if all peoples are equal, then that equality must be on Western terms. 80 In addition, the governments of European colonial empires tired of their colonial projects in the face of rising costs, growing condemnation at home and abroad, and, in some cases, successful indigenous rebellions. Finally, the leaders of the major Communist powers (and to a lesser extent the United States) sought Cold War advantage in backing independence and statehood ambitions of Europe’s overseas colonies.

Thus, the leaders of post-colonial entities became entitled to negative sovereignty, even when they could not demonstrate positive sovereignty. Like statehood more generally, sovereignty has been transformed from a fact into a norm. Traditionally, the non-intervention norm of negative sovereignty flowed from the empirical reality of positive sovereignty (e.g., ultimate control over some delimited territory). From the mid-twentieth century, negative sovereignty became a right of all states, irrespective of their capabilities to exercise effective control over their territory. This normative shift redefined statehood.

3. 1960 and the Redefinition of Statehood

The year 1960 stands as the watershed year for the acceptance of fictitious statehood in international law. Many governments and international institutions unequivocally rejected the traditional criteria for statehood in words and deeds. The United Nations’ Declaration on the Granting of Independence to Colonial Countries and Peoples (“Declaration”) proclaimed that “[a]ll peoples have the right to self-determination; by virtue of that right they freely determine their political status and freely pursue their economic, social and cultural development.” 81 The General Assembly rejected effectiveness as a precondition for statehood: “Inadequacy of political, economic, social or educational preparedness should never serve as a pretext for delaying independence.” 82 The Declaration called on all states to observe its provisions “on the basis of equality, non-interference in the internal affairs of all States, and respect for the sovereign

78. See generally RABASA, supra note 13.
79. HERBST, supra note 4, at 3.
80. JACKSON, supra note 3, at 15-17.
82. Id. ¶ 2 (original emphasis).
rights of all peoples and their territorial integrity. Irrespective of their capabilities, colonial peoples were now entitled to independent statehood and the right to negative sovereignty that accompanied it. The central authorities of these post-colonial entities were entitled to exclude others from their nominal territory regardless of whether they themselves exercised effective control over it.

As the archetypal fictitious state, the status of the DRC in international law illustrates the changed understanding of statehood, as well as the ramifications of this redefinition for global security. In 1960, the same year as the Declaration, the DRC became a member of the United Nations. Yet following Belgium’s withdrawal as the colonial power, the DRC had descended into anarchy. The DRC lacked any semblance of a government exercising control over the area or population of its nominal territory. Thus, the DRC failed the crucial traditional test for statehood under international law. Notwithstanding the ostensible limitation of UN membership to states, the United Nations admitted the geographical expression of the DRC without dissent.

Despite the absence of a government exercising anything approaching effective control, there was and continues to be a formal commitment by the international community to the sanctity of the DRC’s territorial integrity and negative sovereignty. Thus, the UN Security Council condemned as illegal the attempted secession of the DRC’s Katanga province. UN forces intervened on behalf of the authorities in Kinshasa and were instrumental in suppressing Katanga’s attempted independence. A slew of Security Council resolutions have reaffirmed the commitment of its member governments “to the sovereignty, territorial integrity and political independence of the Democratic Republic of the Congo” and even expressed an unfounded expectation that the “Government of the Democratic Republic of the Congo [is capable of] ensuring security in its territory and protecting its civilians with respect for the rule of law, human rights and international humanitarian law.” The International Court of Justice...
Finucane: Fictitious States, Effective Control, and the Use of Force Againsts

FICTITIOUS STATES, EFFECTIVE CONTROL

2012

4. Suspending Disbelief in Fictitious States

As the DRC’s persistence reveals, once born, the fictitious states of the post-colonial era retain their international legal personality irrespective of their actual internal capabilities. Governments have been loath to fully and expressly acknowledge the gulf between the theory and reality of statehood in much of the world. Such acknowledgement would have resulted in the “derecognition” of completely fictitious states such as the DRC and Somalia. Major powers continue to commit to the existence of fictitious states for a number reasons.

Foremost among them is a “domino theory” of derecognition. For example, the derecognition of a completely fictitious state such as Somalia might destabilize other marginal entities such as Chad. Second, the international community simply does not know what to do with stateless territories. A lack of imaginati


96. Id. at 149-150.

97. Id.
tion, resources, interest, and long-term commitment all militate against exploring alternative forms of political organization such as protectorates or trusteeships. Third, the abolition of a territory’s de jure statehood, and hence claims to negative sovereignty, leaves such territories vulnerable to predatory interventions and territorial competition by other states. The current international system exists in large measure to prevent such wars of territorial conquest. Fourth, derecognition would render the territory’s population stateless and thus deprive it of both status and protection under international law. Thus, governments have significant reasons to maintain the current statehood charade and to continue suspending disbelief. Despite these policy considerations speaking against derecognition, the ubiquity of fictitious states and the rarity of effective statehood has serious implications for a global security regime premised upon a system of states. The UN Charter’s state-centric use of force regime is predicated upon states exercising a monopoly over force within their territorial boundaries.

However, because weak states are unable to control their populations and territories, weak states cannot be relied upon to fulfill their international obligations. Samuel Huntington’s observation that “[t]he most important political distinction among countries concerns not their form of government, but their degree of government” is particularly germane to international security and the regime governing the use of force. Whereas in the nineteenth and early twentieth century the gaps between the theory and fact of statehood were generally exceptions, after the mid-twentieth century the existence of such gaps between legal identity and political reality became the norm throughout much of the world. Because decolonization transformed the mismatch between the ideal and reality of statehood from a marginal and usually temporary condition into a permanent state of affairs, it exacerbated the challenges posed by fictitious statehood.

The gap between the ideal of effective statehood and the reality of fictitious statehood provides a habitat for violent non-state actors to thrive. The next section briefly examines some of the characteristics of the non-state entities, which occupy the lacuna between the fact and fiction of statehood and threaten international order.

98. Id.
99. Id.
100. Frontier Dispute (Burk. Faso v. Mali), 1986 I.C.J. 554, ¶ 20 (Dec. 22) (noting that the purpose of uti possidetis in the context of decolonization of Spanish America and Africa was “to prevent the independence and stability of new States being endangered by fratricidal struggles provoked by the challenging of frontiers”).
101. Delahunty & Yoo, supra note 95, at 149-150.
102. SAMUEL HUNTINGTON, POLITICAL ORDER IN CHANGING SOCIETIES 1 (1968) (emphasis added).
103. See RABASA, supra note 13 (a RAND Corporation analysis for the United States Air Force of the security threats present in several ungoverned areas).
C. Violent Non-State Actors

Under a variety of labels — terrorists, marauders, mafias, filibusters, pirates, armed bands, warlords, militias, mercenaries, private armies, insurgents, bandits, free-companies, narco-trafficantes, and the militaries of de facto states — violent non-state actors have been a longstanding fact of international relations. Indeed, not only have they continuously existed alongside modern sovereign states, they precede the modern state. The persistence of independent non-state actors is the corollary to the anomalous nature of effective statehood. Groups such as Hizbollah and the Somali pirates serve as reminders that the ideal of statehood is often unrealized in reality. When such non-state actors exercise independent decision-making and control of their members, their existence violates the “one army rule” of sovereign statehood. Yet, the major powers downplayed or outright ignored the significance and strength of independent non-state actors during decolonization. The international legal regime that developed to support fictitious states could not formally accommodate the existence of violent non-state actors. Explicit acknowledgment would have exposed the gulf between the reality and ideal of statehood in many of the world community’s newest members.

This Article focuses on independent violent non-state actors. Such actors neither satisfy the ICJ’s “effective control” test for state responsibility, nor do they function as de facto agents of state authorities. Although these independent actors may receive safe-harbor, and financial and logistical support from states, the relationships between states and non-state actors are alliances, not forms of agency. The relation of independent non-state actors to state authorities is horizontal, not vertical.

The fact that the internal organizational structure of these non-state actors varies, ranging from a quasi-military hierarchy to a diffuse network of cells or individuals, is irrelevant for the legal framework governing the recourse to force. Although organizational structure may be relevant to the applicability of some provisions of the law of armed conflict (jus in bello), it is irrelevant to

---

105. Fearon & Laitin, supra note 51, at 88 (finding that state weakness permitting the development of insurgency is a strong predictor of civil war).
106. Reisman, Private Armies in a Global War System, supra note 48, at 258 (noting that “the traditional private-army rule seems to have been explicitly rejected by Communist states in association with a number of nations in the Third World”), at 259 (“If the private-army rule of international law were strictly applied and reprisals were undertaken, these nominal states might crumble”).
108. Geneva Convention (III) Relative to the Treatment of Prisoners of War, art. 4(2), Aug. 12,
the rules with respect to recourse to force (*jus ad bellum*). The only relevant factors for *jus ad bellum* are: 1) that the non-state actor operates across an international frontier, and 2) the magnitude of the security threat posed by the non-state actor. 109 Non-state actors lacking the traditional military-style command structures may nonetheless pose serious threats to international security.

Although areas such as Somaliland, Kurdistan, Hamas-controlled Gaza, or Hezbollah-occupied southern Lebanon host non-state actors that may function as *de facto* states, many significant violent non-state actors organize themselves according to principles that are radically different from sovereign statehood. Such actors include those whom the classical Islamic political thinker Muhammad Ibn Khaldu’n called the “desert peoples,” i.e., tribal societies occupying environmentally marginal regions and resisting control by central authorities. 110 Although predatory armed groups are less immediately relevant to American security than Al Qaeda, these groups pose a greater threat to overall human security, particularly in central Africa. Militias such as the Mai Mai, the Forces Démocratiques pour la Libération du Rwanda and the Congrès National pour la Défense du Peuple are among the rotating cast of dozens of independent armed groups that kill, rape and pillage in the DRC and neighboring countries. The threat posed by such groups is emphasized by the recent dispatch of US Special Forces to aid Ugandan efforts against the Lord’s Resistance Army. 111 These groups have played key roles in a complex series of interconnected regional armed conflicts in central Africa that have killed millions, all the while paying little heed to such niceties as international borders.

Independent violent non-state actors are the complement to the fictitious states discussed in the previous section. Such groups flourish where states have failed to consolidate control over society and to co-opt domestic competitors, and where people have relied on alternative models of political organization. 112

---

109. *See WETTBERG*, supra note 104, at 65-66 (also arguing that the organizational structure is irrelevant to the issue of recourse to force against non-state actors).


The presence of such independent armed forces on the territory of a state compromises its positive sovereignty because it violates the “one army rule” of effective statehood. Whether by consent or ineptitude, the state has failed to maintain its monopoly on the use of force.

The prevalence of non-state actors exposes the inadequacy of a strictly state-centric security regime. These armed groups also represent an international constitutional challenge. They belie the formal leveling that accompanied decolonization insofar as they demonstrate that formal equality has not resulted in substantive equality, even with respect to the most fundamental attributes of statehood. Therefore, an international system premised upon the formal equality of states fails doubly; it fails to account for fictitious statehood and it fails to respond to the threat posed by non-state actors. The opinions of the ICJ in Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), 2005 I.C.J. 168 (Dec. 19) illustrate this failure.

**D. Armed Activities: The State-Centric Paradigm vs. Reality**

The International Court of Justice’s opinion in Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), 2005 I.C.J. 168 (Dec. 19), epitomizes the commitment to fictitious statehood and adherence to a rigidly state-centric interpretation of the right of self-defense. The hallmarks of this paradigm are: 1) the primacy of the central authorities, no matter how feckless they are; and 2) the dispositive consent of these authorities with respect to lawful military intervention.

Armed Activities arose out of the complex, multi-party conflict in the territory of DRC. There, “rebel groups were able to operate ‘unimpeded’ in the border region between the DRC and Uganda ‘because of its mountainous terrain, its remoteness from Kinshasa (more than 1,500 km), and almost complete absence of central government presence or authority in the region during President Mobutu’s 32-year term in office’”\(^\text{114}\). After Mobuto Sese Seko’s fall, the security situation deteriorated even further. By 2001, six states and a number of militias were embroiled in a fluid, multisided conflict characterized by the opportunistic exploitation of the DRC’s natural resources and widespread atrocities against civilians, including the deaths of millions.\(^\text{115}\)

In this context, Uganda claimed it could lawfully intervene to counter the threat posed by one militia, the Allied Democratic Forces (“ADF”), which was operating from Congolese territory. Uganda maintained that intervention was

---

113. See JACKSON, supra note 3, at 14.
lawful on the basis of both: 1) the consent of the authorities in Kinshasa, and 2) self-defense.116 As a matter of self-defense, Uganda justified its entry into the Congolese conflagration on the grounds that its own “security situation had become untenable” because “the successive governments of the DRC had not been in effective control of all the territory of the DRC”117 and because of the “political and administrative vacuum” in the eastern DRC.118 Uganda observed that “[t]he fissiparous tendencies of the dysfunctional Congolese state inherited from President Mobutu, which President Kabila’s government had barely papered over, were set loose by the rebellion, and the central government soon lost effective control over the eastern half of the country.”119 In light of the lawlessness in the eastern DRC, Uganda argued that it was necessary to resort to the cross-border use of force in order to prevent further attacks by ADF fighters, who had previously attacked civilian targets inside Uganda. Although Uganda invoked the DRC’s lack of effective control, it premised it right to self-defense upon the fact that the DRC had “not only a duty to refrain from providing any support to groups carrying out subversive or terrorist activities against another State, but also a duty of vigilance to ensure that such activities are not tolerated.”120 Invoking Corfu Channel, (U.K. v. Albania), 1949 I.C.J. 4 (Apr. 9),121 Uganda contended that the DRC’s failure to discharge this duty rendered the DRC responsible for the attacks of the ADF. Thus, Uganda relied upon the DRC’s lack of effective control as evidence of the DRC’s negligence and hence, wrongfulness, rather than the DRC’s compromised sovereignty.

Having established that certain aspects of Uganda’s presence were non-consensual and that Uganda had violated the DRC’s negative sovereignty, the Court considered whether such a violation could be justified on the grounds of self-defense. It rejected Uganda’s claim of self-defense and, in so doing, implicitly rejected a reading of Article 51 that would accommodate the right to use cross-border force against non-state actors in self-defense. The Court observed that “in order to ascertain whether Uganda was entitled to engage in military action on Congolese territory in self-defense, it is first necessary to examine” the connection between the militias and the Congolese state (such as it was).122 In the Court’s view, Uganda’s right to self-defense was contingent upon the responsibility of the DRC for the actions of the non-state actors. Absent such a

117. Id. at ¶ 109.
119. Id. at ¶ 45.
Finucane: Fictitious States, Effective Control, and the Use of Force Againsts

FICTITIOUS STATES, EFFECTIVE CONTROL  59

connection, “the Court [found] that the legal and factual circumstances for the exercise of a right of self-defense by Uganda against the DRC were not present.”

Thus, the Court only addressed the issue of Uganda’s right to self-defense vis-à-vis the DRC. It refused to consider whether Uganda enjoyed a right to self-defense vis-à-vis the ADF or other armed groups operating independently in the anarchic environment of the eastern DRC. However, the judges pointed to the limitations of the state-centric regime governing the use of force when they stated that they were “uncomfortable being confronted with certain questions of utmost importance in contemporary international relations.”

Though a majority of the Court rigidly adhered to a state-centric paradigm for the use of force, two judges criticized the Court’s unwillingness to confront reality. The separate opinions of Judges Kooijmans and Simma reflect an appreciation for the reality of violent non-state actors and the role of state practice and opinio juris in shaping international law. Judge Kooijmans observed that:

The Parties to the present dispute share a hapless post-decolonization history . . . . In this respect the Parties shared the plight which seems to have become endemic in much of the African continent: régimes under constant threat from armed movements often operating from the territory of neighboring States, whose governments sometimes support such movements but often merely tolerate them since they do not have the means to control or repel them. The latter case is one where a government lacks power and consequently fails to exercise effectively its territorial authority; in short, there is a partial failure of State authority and such failure is badly concealed by the formal performance of State functions on the international level. Commitments entered into by governments unable to implement them are unworthy of reliance from the very start and hardly contribute to the creation of more stability.

Judge Kooijmans observed that the Judgment of the Court “inadequately reflects the structural instability and insecurity in the region, the overall pattern of lawlessness and disorder.” He noted that:

Article 51 merely conditions the exercise of the inherent right of self-defense on a previous armed attack without saying that this armed attack must come from an-

123.  Id. at ¶ 147.
124.  See Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), 2005 I.C.J. 168, ¶ 26 (Dec. 19) (separate opinion of Judge Kooijmans) (“The Court only deals with the question whether Uganda was entitled to act in self-defense against the DRC and replies in the negative since the activities of the rebel movements could not be attributed to the DRC. By doing so, the Court does not answer the question as to the kind of action a victim State is entitled to take if the armed operation by irregulars, because of its scale and effects, would have been classified as an armed attack rather than as a mere frontier incident had it been carried out by regular armed forces but no involvement of the host government can be proved”) (internal quotations omitted).
127.  Id. at ¶ 14.
other State even if this has been the generally accepted interpretation for more than 50 years. I also observed that this [state-centric] interpretation no longer seems to be shared by the Security Council, since in resolutions 1368 (2001) and 1373 (2001) it recognizes the inherent right of individual or collective self-defense without making any reference to an armed attack by a State. In these resolutions the Council called acts of international terrorism, without any further qualification and without ascribing them to a particular State, a threat to international peace and security.128

Judge Kooijmans further opined that the case arose in a context “which in present-day international relations has unfortunately become as familiar as terrorism, viz. the almost complete absence of government authority in the whole or part of the territory of a State. If armed attacks are carried out by irregular bands from such territory against a neighboring State, they are still armed attacks even if they cannot be attributed to the territorial State.”129 These non-state actors may pose a security threat equal to that of states. Therefore,

The lawfulness of the conduct of the attacked State must be put to the same test as that applied in the case of a claim of self-defense against a State: does the armed action by the irregulars amount to an armed attack and, if so, is the armed action by the attacked State in conformity with the requirements of necessity and proportionality.”130

Judge Simma also criticized the restrictive reading of Article 51 as inconsistent with the current state of international law.

Such a restrictive reading of Article 51 might well have reflected the state, or rather the prevailing interpretation, of the international law on self-defense for a long time. However, in the light of more recent developments not only in State practice but also with regard to accompanying opinio juris, it ought urgently to be reconsidered, also by the Court. As is well known, these developments were triggered by the terrorist attacks of September 11, in the wake of which claims that Article 51 also covers defensive measures against terrorist groups have been received far more favorably by the international community than other extensive readings of the relevant Charter provisions, particularly the “Bush doctrine” justifying the pre-emptive use of force. Security Council resolutions 1368 (2001) and 1373 (2001) cannot but be read as affirmations of the view that large-scale attacks by non-State actors can qualify as “armed attacks” within the meaning of Article 51.131

Thus, Judges Kooijmans and Simma acknowledged the reality of a world of fictitious states and non-state actors and interpreted the Charter accordingly. The judges recognized that the right of self-defense applied to actions taken against non-state and state actors.

However, the Judges erred by suggesting that 9/11 expanded the scope of this customary right. As the next Part illustrates, the customary right of self-

128. Id. at ¶ 28.
129. Id. at ¶ 30 (internal quotations omitted).
130. Id. at ¶ 31.
defense predates 9/11 and has encompassed actions against non-state actors for two centuries. Violent non-state actors operating from fictitious states have long threatened the security of states, which have often responded to this threat through the use of military force across international borders. The next Part examines international incidents involving the use of force against non-state actors in ungoverned territory and analyzes the legal claims made by states to justify such actions. These incidents place current security threats posed by groups like Al Qa’ida in Yemen or Pakistan in their historical context. These legal claims also reveal a consistent set of principles that justify military action.

III.

STATE RESPONSE: THE OTHER FACE OF EFFECTIVE CONTROL

Customary international law has long accommodated the need to respond to threats from violent non-state actors through a right of self-defense. Historically, governments have been unwilling to tolerate security threats posed by independent violent non-state actors who exploit gaps in state control. Consequently, there is a long history of the use of force against non-state actors and a correspondingly long history of justifications premised upon self-defense in the face of state ineffectiveness.

This body of state practice, 
opinio juris,
and the reactions of other states indicate that self-defense against non-state actors was lawful at the time of the Charter’s drafting. Legal justifications of the pre- and post-Charter eras demonstrate substantial continuity. This history shows that governments were mindful of practical deficiencies in state authority and willing to use self-help to enforce the international legal obligations of other states who were unwilling or unable to do so themselves.

Regardless of whether a state is unwilling or unable to evict violent non-state actors, the presence of non-state actors nevertheless compromises state sovereignty. A state cannot claim a monopoly on violence if an independent armed force is present within its nominal territory. Forceful intervention by the defending state amounts to “extraterritorial law enforcement” because the defending state is enforcing the international obligations of the fictitious state.

This Section examines state practice, self-defense justifications, and, where possible, the responses of other governments to these claims of self-defense. It reviews incidents in which the defending state implicitly or explicitly invokes state ineffectiveness or otherwise emphasizes deficiencies in state control. This tacit acknowledgement of fictitious statehood by governments generally serves one of two purposes. Governments either raise state incapacity or unwillingness to justify the use of defensive force as a necessity, or governments cite host state ineffectiveness as a waiver of the host state’s negative sovereignty.

Such claims premised upon state ineffectiveness invert the ICJ’s “effective control” test for state responsibility. Rather than justifying military action based on the effective control of a state over a violent non-state actor, states justify the defensive use of force based on the host state’s lack of effective control over its nominal territory. From the post-colonial perspective, these claims are radical because they reassert the connection between positive and negative sovereignty. To claim their right to exclude, states must demonstrate their ability to control.

During the nineteenth and early twentieth centuries, the United States repeatedly resorted to defensive force in response to threats posed by violent non-state actors operating from lawless regions across its borders. Whether justified as defensive, preventative, precautionary, or punitive, these actions all had the same stated end: abating this threat. These interventions contributed to the establishment of an international historical pattern and a legal framework for state response to lawlessness and international terrorism that persists today.

A. United States and Spanish Florida

The modern template for intervention against non-state actors operating from ungoverned territory dates to 1817, when US military forces under the command of Andrew Jackson launched an expedition into Spanish Florida. US forces fought against armed bands of Seminole Indians that had been instigated by British military officers. Commenting on the legal basis for the United States’ use of force on Spanish territory, Jackson observed to Secretary of War John C. Calhoun that “[t]he Spanish Government is bound by treaty . . . to keep her Indians at peace with us. They have acknowledged their incompetency to do this, and are consequently bound, by the law of nations, to yield us all facilities to reduce them.” Jackson’s invocation of a longstanding legal framework demonstrates that the right of self-defense against non-state actors was already an established component of international law by the early nineteenth century. The 1817 incident merely re-asserted those pre-existing norms and it applied them to relations between emerging modern states, non-state territorial entities, and violent non-state actors.

Further illuminating this reasoning, Secretary of State John Quincy Adams later described the legal basis for US action in a similar fashion:

He [General Jackson] took possession, therefore, of [Spanish territory] . . . not in a spirit of hostility to Spain, but as a necessary measure of self-defense; giving notice that they should be restored whenever Spain should place commanders and a force there able and willing to fulfill the engagements of Spain towards the United States, or of restraining by force the Florida Indians from hostilities...
against their citizens.136

The legal claims advanced by the United States are notable because they emphasize the following: 1) Spanish “incompetence,” 2) the distinction between the armed bands and Spain, and 3) a good-faith willingness to cooperate with Spanish authorities to restore Spain’s effective control over the affected area. Spain implicitly accepted this reasoning and did not contest the lawfulness of the incursion itself. Instead, Spain’s protests to the United States focused on the plundering and destruction of Spanish property that accompanied the intervention.137 In short, Spain contested the *jus in bello* rather than the *jus ad bellum* of the American action.

In 1817, American forces also launched an incursion into Spanish territory off the eastern coast of Florida. Amelia Island had fallen under the control of Gregor McGregor and a band of what President James Monroe described as “adventurers from different countries, with very few, if any of the native inhabitants of the Spanish colonies.”138 Secretary of State Adams justified the United States’ use of force against Amelia Island, arguing:

> When an island is occupied by a nest of pirates, harassing the commerce of the United States, they may be pursued and driven from it, by authority of the United States, even though such island were nominally under the jurisdiction of Spain, Spain not exercising over it any control.139

Once again the United States emphasized the threat it was responding to and the necessity of using force. In this case, however, Spain’s lack of control over Amelia Island appears to have mitigated any violation of its negative sovereignty by the United States. Spain, as well as the putative representative of Spain’s then-renegade colonies, protested US action.140 However, diplomatic correspondence reveals that the United Kingdom and other major powers expressed “no dissatisfaction” with US occupation of the island.141

**B. The Caroline Incident: The United States of Ineffectiveness**

The United States has not always been the defending state in international incidents involving military response to violent non-state actors. The *Caroline* incident described below reveals that the United States does not invoke principles it is unwilling to apply to itself. The United States has itself, at times, lost effective control over its territory, thereby permitting the presence of independent armed groups.

During the British suppression of an 1837 insurrection in southern Canada,

136. *Id.* at 405-06.
137. *Id.* at 404.
138. *Id.* at 408
139. *Id.*
140. *Id.*
141. *Id.*; see also Reisman, *Private Armies in a Global War System*, supra note 48, at 252, 256.
Canadian rebels sought refuge across the border in New York State. The rebels recruited American sympathizers to gather private arms and pillage the armory at Batavia, New York, making off with several hundred weapons, including two field artillery pieces. In a letter to the President, the mayor of Buffalo complained that “[t]he civil authorities have no adequate force to control these men, and unless the General Government should interfere, there is no way to prevent serious disturbances.”

The “General Government” did not interfere and the Canadian-American insurrection went on to occupy Navy Island, a Canadian island in the Niagara River. From the island these forces repeatedly engaged in “Acts of Warlike aggression on the Canadian shore, and also on British Boats passing the Island.” The insurrectionists also manned a steamer, the Caroline, to provision (especially with munitions) the forces encamped on the island. British authorities alerted the government of New York of the threat posed to British territory and subjects by the lawlessness in New York, but they received no reply.

With the local authorities unable to control their territory and population and the state and federal authorities uninterested in asserting control, Britain resorted to unilateral force to prevent future attacks. On the night of December 29, 1837, while the Caroline was docked on the American bank of the river, British forces assaulted the vessel. They killed and wounded several of the Americans onboard and sent the Caroline over Niagara Falls in flames.

During the diplomatic row that ensued, the governments of the United States and Britain established what would become the defining principles of lawful self-defense against state and non-state actors in customary international law. The British justified their incursion into American territory on the basis of: 1) the threat posed by the Caroline, 2) the insurrectionists’ status as an independent armed force, and 3) the United States’ lack of effective control over its own territory and populace. Henry Stephen Fox, the British Minister in Washington, replied to a complaint by the US Secretary of State by citing both the “piratical character of the steam boat ‘Caroline,’” and the necessity of self-defense and self-preservation. Moreover, Fox noted that:

At the time when the event happened, the ordinary laws of the United States were not enforced within the frontier district of the State of New York. The authority

143. Id. at 83 (citing H. Ex. Doc. No. 74, 25th Cong., 2d Sess.).
144. Id. at 86 n.12 (citing H. Ex. Doc. No. 74, 25th Cong., 2d Sess.).
145. Id. at 83.
146. Id. (citing Law Officer’s Report, Feb. 21, 1838).
147. Id.
148. Id. at 84.
149. MOORE, supra note 30, at Vol. 2, 409.
150. Jennings, supra note 142, at 85.
of the law was overborne, publicly, by piratical violence . . . . This extraordinary state of things appears, naturally and necessarily, to have impelled them [Canadian armed forces] to consult their own security, by pursuing and destroying the vessel of their piratical enemy, wheresoever they might find her.”

Thus, the United States’ lack of effective control necessitated Britain’s resort to defensive force.

In a report to the British Foreign Minister, the legal officers of Britain’s Foreign Office emphasized both the necessity for intervention in the Caroline incident and the precautionary motive behind Britain’s actions. The report stated: “We feel bound to suggest to your Lordship that the grounds on which we consider the conduct of the British Authorities to be justified is that it was absolutely necessary as a measure of precaution for the future and not as a measure of retaliation for the past. What has been done previously is only important as affording irresistible evidence of what would occur afterwards.” In other words, the lawfulness of British action hinged upon the fact that it was intended to prevent future harm, and not to serve as revenge for prior injury.

In a letter to Lord Ashburton, US Secretary of State Daniel Webster acknowledged the potential lawfulness of cross-border defensive measures against non-state actors. However, in his famous formulation, Webster established a high bar for the legitimacy of such action: hostile acts “within the territory of a party at Peace” could only be justified by “clear and absolute necessity.” Moreover, Webster wrote that the “necessity of [that] self-defense . . . [is] instant, overwhelming, leaving no choice of means, and no moment for deliberation.” Even if such necessity authorized British authorities to enter US territory, Webster argued, their actions were only lawful if they “did nothing unreasonable or excessive; since the act, justified by the necessity of self-defense, must be limited by that necessity, and kept clearly within it.”

In his reply, Ashburton accepted Webster’s framework for lawful self-defense. He argued that the necessity prong was satisfied due to the lack of effective control by the American authorities. Given this lack of control, Ashburton reasoned, appeal to the American authorities would have been futile. Ashburton wrote, “I might safely put it to any candid man, acquainted with the existing state of things, to say whether the military commander in Canada had the remotest reason . . . to expect to be relieved from this state of suffering by
the protective intervention of any American authority[.]. He argued that the lack of alternative means necessitated the use of military force and that the British nighttime raid and destruction of the vessel was a proportional response to the threat. Furthermore, he maintained that the timing of the raid and the loosing of the flaming vessel over the falls were proportional because they would minimize both the loss of life and damage to other property.  

Webster noted that this narrow exception to the “inviolable character of the territory of independent states . . . [grew] out of the great law of self-defense.” He further observed that the governments of the United States and the United Kingdom agreed on the legal standard for self-defense, but disagreed as to whether the “facts in the case of the Caroline make out a case of such necessity for the purposes of self-defense.”

The Caroline incident is particularly significant for the law of self-defense against non-state actors. The British did not attempt to justify intervention on the basis of American responsibility for the Caroline but New York State’s (admitted) lack of effective control over its own territory. Both Britain and the United States agreed in principle to the existence of a right to self-defense against non-state actors. They also agreed that such a right could justify the violation of another state’s negative sovereignty. Moreover, a prior armed attack was merely additional evidence of the existence of a security threat that necessitated the use of force. Thus, such self-defense was forward looking insofar as its purpose was aimed to future harm, and not to retaliate for past wrongs.

C. United States and Mexico: Ungoverned Space on the Border

During the nineteenth century, the United States repeatedly resorted to cross-border incursion into Mexico in response to threats posed by “predatory Indians and other marauders.” In 1836, Secretary of State John Forsyth wrote to his Mexican counterpart explaining that any US military incursion into Mexico was entirely on the basis of self-defense and emphasized that force was not directed towards the Mexican state.

Should the [American] troops, in the performance of their duty, be advanced beyond the point Mexico might suppose was within the territory of the United States, the occupation of the position was not to be taken as an indication of any hostile feeling, or of a desire to establish a possession or claim not justified by the treaty of limits, [but only as] precautionary and provisional, [to be] abandoned whenever the disturbance in that region should cease, they being the only motive for it.  

160.  Id. at 90 (citing 61 Parl. Deb., H.C. (3d ser.) (1843) (U.K.); 30 B.S.P. 195).
161.  Id.
163.  Id.
164.  Id. at 418.
165.  Id. at 418.
Forsyth elaborated upon the United States’ legal position in a letter to the American ambassador to Mexico. The letter emphasized that intervention into Mexico rested “upon principles of the law of nations, entirely distinct from those on which war is justified—upon the immutable principles of self-defense—upon the principles which justify decisive measures of precaution to prevent irreparable evil to our own or to a neighboring people.”

The United States’ right to self-defense against the non-state threat entailed preemptive action. In justifying proactive action, Forsyth argued that:

Our fellow-citizens . . . are to be exposed to massacre . . . and the whole frontier to be laid waste by those savages Mexico was bound to control. Until these evils happen, on Mr. Gorostiza’s theory, we have no right to take a position which will enable us to act with effect; and before we do act . . . after the frontier has been desolated, we must demand redress of Mexico, wait for it to be refused, and then make war upon Mexico. We are quietly to suffer injuries we might prevent in the expectation of redress—redress from irreparable injuries from Mexico, who did not inflict them, but who was, from circumstance, without the power to prevent . . . To make war upon Mexico for this involuntary failure to comply with her obligations, would be equivalent to an attempt to convert her misfortunes into crimes—her inability into guilt.

The United States’ legal claim: 1) differentiated between the Mexican state and non-state actors, 2) premised self-defense upon the necessity arising from Mexico’s lack of effective control, and 3) emphasized that American action was preventative or precautionary.

Mexico did not receive the American legal claim well. In light of Texas’ subsequent admission into the United States and the Mexican Cession of 1848 following the Mexican-American War, Mexico may have had real concerns over the sincerity of any American disavowal of territorial designs. In short, Mexico rejected the United States’ legal claims because it deemed them pretextual, and thus the United States’ recourse to force as unnecessary.

Twenty years later, a company of Texas Rangers pursued into Mexico a band of Mexican natives that had launched a raid into Texas. In a letter to his Mexican counterpart, Secretary of State William Marcy asserted that international law allowed such hot pursuit in self-defense and that the United States would not complain if Mexico launched, out of necessity, similar proportionate incursions into the United States.

If Indians whom the United States are bound to restrain shall, under the same circumstances, make a hostile incursion into Mexico, this Government will not complain if the Mexican forces who may be sent to repel them shall cross to this side of the line for that purpose, provided that in so doing they abstain from injuring the persons and property of the citizens of the United States.

166. Id. at 420 (emphasis added).
167. Id. at 420-421.
168. Id. at 418-420.
169. Id. at 421.
It is thus clear that the United States was not claiming for itself a right it was unwilling to grant other states. As subsequent Secretary of State Hamilton Fish observed to the Secretary of War William Belknap, “[a]n incursion into the territory of Mexico for the purpose of dispersing a band of Indian marauders is, if necessary, not a violation of the law of nations.”\(^\text{170}\)

President Buchanan went a step further and proposed the establishment of a temporary protectorate in areas of “anarchy and violence” of Northern Mexico.\(^\text{171}\) The government would withdraw American forces “as soon as local governments shall be established in these Mexican States capable of performing their duties to the United States, restraining the lawless, and preserving peace along the border.”\(^\text{172}\) Again, the United States based this claim of necessity upon Mexico’s lack of effective control over its nominal territory.

Lawlessness in Mexico was also the basis for the United States’ most famous incursion into Mexico: John Pershing’s 1916 pursuit of Francisco “Pancho” Villa, following Villa’s attack upon Columbus, New Mexico. The United States initially justified Pershing’s expedition on the basis of both defensive reprisal and the consent of the Mexican state. President Woodrow Wilson publicly announced that, “the expedition into Mexico was ordered under an agreement with the de facto government of Mexico for the single purpose of taking the bandit Villa . . . and is in no sense intended as an invasion of that republic or as an infringement of its sovereignty.”\(^\text{173}\) Wilson also asserted that “[t]he expedition is simply a necessary punitive measure, aimed solely at the elimination of the marauders who raided Columbus and who infest an unprotected district near the border which they use as a base in making attacks upon the lives and property of our citizens with our own territory.”\(^\text{174}\) Wilson’s reasoning clearly indicates that his use of “punitive” is best understood as precautionary, and not as retaliatory: the purpose of the “punitive measure” is “solely . . . the elimination of the marauders.” The general consent of the de facto government of Mexico for the US military expedition was based on the prior reciprocal agreement between the two countries permitting hot pursuit of “bandits” across the international border into the United States.\(^\text{175}\) The United States renewed its commitment to this agreement hoping to “suppress this state of lawlessness . . . in the regions contiguous to the boundary between the two Republics.”\(^\text{176}\) In a parallel to the current US involvement in Pakistan, the Mexican state subsequently publicly denied consenting to the presence of the US military in its territory. The United States replied by “vigorously defending its action in protecting

\(^{170}\) Id.

\(^{171}\) Id.

\(^{172}\) Id.

\(^{173}\) George Finch, Mexico and the United States, 11 Am. J. Int’l L. 399, 399-400 (1917).

\(^{174}\) Id. at 400.

\(^{175}\) Green Haywood Hackworth, Digest of International Law 292-293 (1940).

\(^{176}\) Id. at 292.
its border by patrolling a portion of Mexico” where the Mexican state “was obviously unable to exert any semblance of authority”, and the United States refused to withdraw until Mexico “gave evidence of some ability to fulfill his international obligations to his neighbor on the north.”

Writing to his Mexican counterpart, US Secretary of State Robert Lansing decried the persistent inability of the Mexican state to repress the “marauding attacks” across the international border. Such ineffectiveness “may excuse the failure to check the outrages complained of, but it only makes stronger the duty of the United States to prevent them.”

A bilateral commission eventually negotiated the withdrawal of American forces, contingent on their immediate replacement by Mexican troops and the “occupation and protection of territory evacuated by the American forces.”

The American members of the commission also insisted that the United States reserved “the right to pursue marauders coming from Mexico into the United States so long as conditions in northern Mexico are in their present abnormal condition. Such pursuit is not however, to be regarded by Mexico as in any way hostile to the Carranza Government, for these marauders are our common enemies.”

Despite Mexican authorities’ poor reception of the agreement, US forces withdrew by early 1917.

Pershing’s expedition is notable because the United States premised its legal claim upon the necessity of self-defense arising from Mexico’s lack of effective control over the territory of Northern Mexico. In such a power vacuum the United States was entitled to fulfill the international legal obligations neglected by Mexico. Nonetheless, the United States took pains to distinguish between Pancho Villa and the Mexican state and did not justify the expedition on the grounds of Mexican responsibility. Like his predecessor, Secretary of State Forsyth, Lansing distinguished between Mexican wrongfulness (failure to control its nominal territory) and the threat posed by “bandits.”

**D. Soviet Union: Civil War Spillover**

Further confirming the existence of a customary law right, defensive actions premised upon state weakness were not merely the preserve of western, capitalist states. The Soviet Union repeatedly resorted to cross-border force against White Guard forces in order to prevent “acts of aggression.” In a

---

177. Finch, supra note 173, at 401.
178. HACKWORTH, supra note 175, at 297.
179. Id.
180. Finch, supra note 175, at 403.
181. Id. at 404.
182. Id. at 405.
183. HACKWORTH, supra note 175, at 297.
184. Note from Chicherin and Rakovsky to the Rumanian Foreign Minister on Anti-Soviet Or-
1921 note to the Romanian Foreign Minister, the Soviet Government wrote that:

In its desire to assist the Rumanian authorities to disperse the bands organized in Bessarabia and Rumania for the purposes of carrying out acts of aggression against the Soviet Republics, the allied Soviet Government consider it necessary that if such bands, when pursued by Soviet troops, should cross into territory occupied by the Rumanian authorities, they should be followed into this latter territory, the Rumanian authorities being informed in time so that the operations . . . shall not be interpreted as acts directed . . . against the Rumanian Government and people.”

Thus, the Soviet Government implied that recourse to force was necessary because Romania was incapable of dispersing the White Guards on its territory. As with the earlier US incursions into Mexico, the Soviet Union took pains to differentiate between the target non-state actor and the Romanian state and population and to emphasize that force was directed against the former and not the latter.

In 1921, the Red Army also attacked White Russian forces based in Outer Mongolia, which was under the nominal sovereignty of the post-Qing Chinese Republic. The anti-Bolshevik forces had established themselves in Mongolia at a time when central authority in China was exceedingly weak and warlords held local power. In a 1925 note to the Chinese Foreign Minister, the Soviet Government argued, in tones reminiscent of the earlier diplomatic exchanges involving the United States, Britain, and Mexico, that it had resorted to force only after exhausting all other options.

Repeated requests addressed to the Chinese Government for the liquidation of the White Guard bands of Semenov, Ungern, and others, freely operating and organizing on the territory of Mongolia, led to no positive results, as the Chinese Government was indifferent to these urgent appeals by the Soviet Government. In view of this in the interest of the safety of its frontiers the Soviet Government was constrained to conduct part of the Red Army into Mongolian territory and liquidate all the White Guard bands and organizations which, organized and supported by foreign imperialism, were preparing to invade the Soviet Republic once more from Mongolia. After the liquidation of the White Guard armies part of the Red Army remained in Mongolia in the interest of the preservation of order and for the purpose of preventing the organization of White bands as a new menace to the safety of the USSR.

As with the United States and Britain, the Soviet Union resorted to force when its neighbors were unwilling or unable to control their own territories and it defended such actions on the basis of self-defense.

185. Id.


Notwithstanding the reification of fictitious statehood during decolonization, fictitious states did not disappear as a result of these good intentions. Nor did violent non-state actors disappear, bred as they were by these newly minted states. As a result, states continued to resort to force against non-state actors, and they continued to emphasize a state’s lack of control as a key factor necessitating self-help. The post-Charter justifications for such uses of force echo many of the legal claims made in the pre-Charter era. These claims indicate that, though the governments of defending states may have been willing to tolerate fictional statehood and negative sovereignty in general, they often justified actual defensive measures by citing a lack of effective control.

At least in principle, many major powers accepted such recourses to defensive force as lawful. This acceptance—or at least acquiescence—reflects the survival of the customary right of self-defense against non-state actors in the post-Charter era. However, some governments, particularly those of the former colonies and their Communist backers, generally rejected any claim of self-defense against non-state actors, whether or not premised upon state ineffectiveness.

For example, as a perennial target of violent non-state actors, Israel has advanced a disproportionate number of claims of self-defense. Many former colonies and Communist states considered Israel, like South Africa and Rhodesia, to be a pariah. As a result, many states regarded Israel’s claims of self-defense as per se invalid.188 Given this perceived a priori illegitimacy of any Israeli military action, an incident-based analysis of international reaction to Israel muddies the waters regarding the law of self-defense.

1. Israel

Israel has repeatedly invoked the absence of positive sovereignty as a justification for violating a host state’s negative sovereignty.189 In 1976, Israel staged a successful commando raid on Entebbe, Uganda, to rescue Israeli hostages held by Palestinian hijackers. In the aftermath of the assault, Israel’s representative wrote to the UN Secretary General justifying its use of force as self-defense on behalf of its citizens.190 Israel argued that its incursion into Ugandan

---

188. Elimination of All Forms of Racial Discrimination, G.A. Res. 3379 (XXX), U.N. Doc. A/RES/3379 (Nov. 10, 1975) (stating “that the racist regime in occupied Palestine and the racist regime in Zimbabwe and South Africa have a common imperialist origin, forming a whole and having the same racist structure and being organically linked in their policy aimed at repression of the dignity and integrity of the human being”).
189. The majority of instances of Israeli recourse to force in self-defense against non-state actors fall outside the scope of this article. For a more complete study, see WETTBERG, supra note 104, at 73-122, 164-165.
territory was lawful because Uganda “does not exercise sovereignty over its territory and was incapable of dealing with half a dozen terrorists.”\textsuperscript{191} Thus, Israel’s violation of Uganda’s sovereignty was justified because Uganda did not in fact exercise sovereignty, and also because such state weakness posed a real and immediate threat to Israeli citizens. The reaction to Israel’s use of force was mixed, but ultimately efforts to condemn Israel in the UN Security Council failed and Uganda made no attempt to convene the General Assembly under the “Uniting for Peace” procedure.\textsuperscript{192}

In 1981, Israel again justified its actions by reference to its right to self-defense, “a right also preserved under Article 51 of the Charter of the United Nations.”\textsuperscript{193} Israel cited the weakness of the Lebanese state as necessitating its 1981 airstrikes against Palestinian Liberation Organization (PLO) targets in Lebanon, emphasizing that “PLO domination over large parts of Lebanon and the anarchy it has created there.”\textsuperscript{194} Israel also highlighted the fact that the Lebanese state lacked “effective authority” over “large parts of its territory controlled by foreign elements.”\textsuperscript{195} Israel explicitly invoked both the Caroline incident and the Pershing’s expedition into Mexico as legal precedents for Israel’s defensive use of force against the PLO.\textsuperscript{196} According to Israeli authorities, the inability of the Lebanese government to control its territory necessitated Israel’s recourse to force. However, members of the Security Council states rejected Israel’s justification of self-defense because they deemed its military measures to be preemptive and disproportionate.\textsuperscript{197}

Israel also cited the gap between positive and negative sovereignty when it justified its 1982 invasion of Lebanon. It first cited the necessity of defending itself against attacks by the PLO.\textsuperscript{198} Israel argued that the invasion of Lebanese territory was a lawful defensive response because Lebanon had “lost much of its sovereignty over its own territory to the terrorist PLO.”\textsuperscript{199}

The United Nations, however, deemed Israel’s 1982 invasion unlawful. The UN Security Council unanimously adopted a resolution demanding that “Israel withdraw its forces forthwith.”\textsuperscript{200} The General Assembly, convened under the

\begin{itemize}
\item \textsuperscript{192} THOMAS FRANCK, RECOURSE TO FORCE 83-85 (2002).
\item \textsuperscript{194} Id. at 5.
\item \textsuperscript{195} Id. at 6.
\item \textsuperscript{196} William Claiborne, \textit{Begin Widens Targeting to PLO Sites in Cities}, WASH. POST, Jul. 18, 1981, at A1.
\item \textsuperscript{197} U.N. SCOR, 36th Sess., 2293rd mtg. at 4, U.N. Doc. S/PV.2293 (July 21, 1981) (France criticizing Israel’s action as preemptive); Id. at 5 (the United Kingdom criticizing the scale of Israeli action); Id. at 7 (Egypt criticizing any action that “fails to be proportionate”).
\item \textsuperscript{198} U.N. SCOR, 37th Sess., 2375th mtg. at 3, U.N. Doc. S/PV.2375 (Jun. 6, 1982).
\item \textsuperscript{199} Id. at 5.
\item \textsuperscript{200} S.C. Res. 509, ¶ 1, U.N. Doc. S/RES 509 (June 6, 1982).
\end{itemize}
Finucane: Fictitious States, Effective Control, and the Use of Force Againsts

“Uniting for Peace” procedure, expressed alarm in a vote of 127 to 2 (with Israel and the United States casting the opposing votes) at “Israel’s acts of aggression.”\(^{201}\) The resolution also reaffirmed the fundamental principles of Lebanese “sovereignty, territorial integrity, unity and political independence.”\(^{202}\) Even US President Ronald Reagan deplored Israel’s actions as disproportionate, a sentiment shared by many leaders.\(^{203}\)

Israel continued to argue that state weakness necessitated self-defense in subsequent military actions. In 1988, Israel launched airstrikes against PLO targets in the Lebanese city of Sidon in response to rocket attacks by the PLO from southern Lebanon. Citing self-defense as the basis for its actions, Israel argued that its action was necessary given Lebanon’s lack of effective control over its territory. “[I]n the absence of a Lebanese Government capable of assuming its responsibilities, we have no option but to take necessary measures for our security.”\(^{204}\) Moreover, Israel contended that its “ongoing measures for self-defense . . . are restrained, they are temporary, but they are necessary.”\(^{205}\) The United States vetoed a proposed Security Council resolution condemning Israel.\(^{206}\)

Israel’s claim of necessity for its 1996 “Operation Grapes of Wrath” campaign against Hezbollah targets in southern Lebanon was also based upon the fecklessness of the “so-called sovereign Government of the State of Lebanon.”\(^{207}\) “The Lebanese Government does not have the ability — or the will — to control Hezbollah activities. Therefore, Israel must defend the security of its north by all necessary measures . . . . The Lebanese Government was told time and again: control the Hezbollah. If you are, as you claim, the sovereign Government of Lebanon, then this is your obligation.”\(^{208}\) Thus once again, Israel invoked state weakness as necessitating its recourse to force.

Israel also highlighted the hypocrisy of sovereignty. “It was very strange to hear from the Prime Minister of Lebanon, just last night, that ‘It is not within our ability to do this.’ Please decide: either his is the sovereign Government, or


\(^{202}\) Id.


\(^{205}\) Id.


\(^{208}\) Id. at 6.
it is not within its ability.”\textsuperscript{209} Both Germany and Russia criticized Israel’s actions as disproportionate.\textsuperscript{210} Egypt rejected Israel’s claim of self-defense and in doing so explicitly invoked the Caroline formulation.\textsuperscript{211} Israel, Egypt contended, had not exhausted other means of achieving its security objectives, and its actions were therefore unnecessary.\textsuperscript{212} Moreover, Egypt claimed that Israel’s response to the threat was disproportionate.\textsuperscript{213}

The contrasting international reactions to the Entebbe incident, the 1982 invasion of Lebanon, and the 1996 Grapes of Wrath operation can in part be understood in terms of proportionality.\textsuperscript{214} Many key states considered Israel’s surgical and brief incursion into Uganda lawful because it was a proportionate defensive response. However, UN member states regarded Israel’s full-scale invasion of Lebanon to be a disproportionate response to the PLO’s cross-border raids. In that instance, Israel invaded territory beyond that from which the PLO had launched its attacks. Israel resorted to force not only in the ungoverned territory of Lebanon’s south, but also in territory under the effective control of the Lebanese state. Thus, Israel violated both Lebanon’s positive and negative sovereignty.

2. Russia-Afghanistan

Foreshadowing the events and claims of 2001, Russia cited self-defense in 1993 when it moved against militants in Afghanistan. Russian and Tajik forces repulsed an incursion by “armed formations of irregular forces of the Tajik opposition and Afghan Mujahidin” launched from Afghan territory into Tajikistan.\textsuperscript{215} Russia alleged that “responsibility for this act of banditry lies squarely with those extremist groups in Afghanistan,” but it did not attribute the incursion to the Afghan state.\textsuperscript{216} Indeed, the Afghan government denied any involvement in the border incident.\textsuperscript{217} Russia and Tajikistan justified their subsequent artillery assault against targets within Afghan territory on the basis of the right of collective self-defense under Article 51 of the Charter.\textsuperscript{218} Russia concluded by

\begin{itemize}
\item \textsuperscript{209} Id.
\item \textsuperscript{210} Id. at 9-10.
\item \textsuperscript{211} Id. at 14-15.
\item \textsuperscript{212} Id.
\item \textsuperscript{213} Id.
\item \textsuperscript{214} As well as the fact that Uganda’s then dictator, Idi Amin, was himself an international pariah.
\item \textsuperscript{216} Id.
\item \textsuperscript{218} Permanent Rep. of the Russian Federation to the United Nations, Letter dated Jul. 15, 1993
\end{itemize}
noting that, “the moment has come when the international community as a whole must adopt a responsible and realistic approach towards those forces which flagrantly violate the norms of international law.”\textsuperscript{219} Russia thus deemed its right to self-defense valid independent of any responsibility of the Afghan state.

The major powers’ response to these defensive actions was tepid. None of the UN political organs specifically objected to the joint Russian-Tajik military operations on Afghan territory. The Security Council called “for the cessation of all hostile actions on the Tajik-Afghan border,” but it also reaffirmed the “necessity to respect the sovereignty and territorial integrity of Tajikistan.”\textsuperscript{220} The European Community issued a statement calling for “moderation” by all parties.\textsuperscript{221} The general acquiescence of key states to the Russian-Tajik action contrasts sharply with the international response to the Soviet Union’s 1979 invasion of Afghanistan. It suggests that the 1993 action was lawful both in principle and practice.

3. Turkey-Iran-Iraq

Major powers and Israel have not been the only nations to premise self-defense upon state weakness. Turkey and Iran have also claimed the right in response to the threats posed by non-state actors. The authorities in Baghdad have rarely, if ever, exercised effective control over the mountainous, Kurdish-occupied territory of northern Iraq, despite their repeated and brutal military campaigns in the region.\textsuperscript{222} Armed Kurdish nationalist groups, in particular the Kurdish Workers Party (“PKK”), have exploited this absence of control to establish bases of operation from which to launch attacks into Turkey and Iran. Prior to the 1991 Gulf War, the government of Saddam Hussein tacitly admitted its lack of control over northern Iraq and the threat this posed to its neighbors by creating a six-mile-wide joint Iraqi-Turkish security zone straddling the border and by consenting to repeated incursions by the Turkish military.\textsuperscript{223}

\begin{flushleft}
\textsuperscript{219} FICTITIOUS STATES, EFFECTIVE CONTROL
\end{flushleft}
Following the Gulf War, Iraq withdrew its consent to external intervention and reasserted negative sovereignty. This claim of sovereignty was particularly striking given that Baghdad’s control over the region had become even weaker following the establishment of the “no-fly” zone after the first Gulf War. Despite lack of consent by the authorities in Baghdad, Turkey and Iran continued to resort to force against Kurdish targets in northern Iraq. Both states invoked the Baghdad government’s absence of control as a basis for their interventions in northern Iraq.

In 1995, Turkey launched Operation Steel, its most comprehensive military assault upon PKK targets located in the “no-fly” zone of northern Iraq. Turkey argued that its actions constituted legitimate measures of self-defense intended to prevent the use of [Iraqi] territory for the staging of terrorist acts against Turkey. Turkey’s resort to force was necessary and hence lawful because Iraq had “not been able to exercise its authority over the northern part of its country since 1991.” Moreover, Turkey emphasized that its operations were “of limited time and scope” and thus proportional. Despite Iraqi complaints of Turkish aggression to the Secretary-General and the Security Council, neither the General Assembly nor the Security Council met to discuss the issue, much less to take action on it. The United States expressed understanding for Turkish self-defense measures, though the United Kingdom and Germany questioned the proportionality of Turkish military actions and called for Turkey to “practice restraint.”

In 1996, Iran advanced a similar justification for its own attacks on Kurdish targets in northern Iraq. In fulfilling its Article 51 reporting requirement, Iran argued that its use of force on Iraqi soil was “[i]n response to these encroachments by terrorist armed groups and in accordance with its inherent right of self-

---


226. Id.

227. Id. at 1.

228. FRANCK, supra note 192, at 63; Christine Gray, International Law and the Use of Force 117 (2008).


defense enshrined in Article 51 of the Charter.”231 Iran noted that the “Government of Iraq is not in a position to exercise effective control over its territory in the northern part of that country.”232 As a consequence, “transborder armed attacks and sabotage operations by terrorist groups against Iranian border towns, originating from Iraqi territory, have been intensified and escalated.”233 Thus, Iraq’s lack of control over its territory necessitated Iran’s “immediate and proportional measures.”234

As with the Turkish incursions, the United Nations took no action against Iran, despite Iraq’s protests.235 The silence of the UN organs and other critical actors suggests continued acceptance of transnational force against non-state actors.

**F. Post 9/11: Continuity and Formalization**

1. **United States-Afghanistan**

Al Qa’ida’s September 11, 2001, attacks upon the United States and the subsequent US invasion of Afghanistan confirmed the post-Charter vitality of an internationally recognized right to self-defense against non-state actors. Although some have argued that the attacks and their aftermath represent an “international constitutional moment”236 with respect to the use of force against non-state actors, 9/11 merely marked a more explicit recognition of a longstanding right. Previously, governments had often acquiesced to defensive action against non-state actors after the fact, but in the aftermath of 9/11 the international community legitimized defensive action in advance. The UN Security Council immediately condemned the atrocities as “attacks” which represented a “threat to international peace and security” and recognized the “inherent right of individual or collective self-defense in accordance with the Charter.”237 Notably, the Security Council did not attribute the threat to international peace and security to any state, evidenced in part by the fact that the resolution recognized a right to self-defense without reference to any state.238

---

232. Id.
233. Id.
234. Id.
236. Slaughter & Burke-White, supra note 18, at 1.
238. See Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), 2005 I.C.J. at ¶ 28 (separate opinion of Judge Kooijmans) (“I also observed that this interpretation no longer seems to be shared by the Security Council, since in resolutions 1368 (2001) and 1373 (2001) it recognizes the inherent right of individual or collective self-defense without making any reference to an armed attack by a State. In these resolutions the Council called acts of international terrorism,
In reporting its subsequent military response to the Security Council, the United States invoked its “inherent right . . . to self-defense” against both “organizations and states.” Although the United States alleged that Al Qa’ida was supported by the “Taliban regime,” the United States did not claim that the Afghan state (if it even existed) was responsible for the 9/11 attacks. Instead the United States took action against the “Al-Qaeda terrorist training camps and military installations of the Taliban regime in Afghanistan” but not against the Afghan state as such. Faced with the lethal threat of Al Qa’ida, the fact or fiction of the Afghan state and its responsibility were simply irrelevant to the issue of self-defense. Subsequent US military action in Afghanistan met with near universal approval.

2. Russia-Georgia

Russia’s 2002 actions in Georgia illustrate that while major powers may agree in principle to the right to self-defense against non-state actors, they may differ on the application of this principle. Invoking the 9/11 attacks, Russia argued before the Security Council in 2002 that the UN Charter must be applied in such a way as to accommodate the existence of ungoverned territories and transnational terrorists. In the context of its ongoing campaign against Chechen separatist and allied Islamist fighters, Russia argued that “[t]he continued existence in separate parts of the world of territorial enclaves outside the control of national governments, which, owing to the most diverse circumstances, are unable or unwilling to counteract the terrorist threat is one of the reasons that complicate efforts to combat terrorism effectively.” Russia identified Georgia’s Pankisi Gorge as one such ungoverned territory and maintained that it could lawfully use force against non-state actors operating from this region: “If the Georgian leadership is unable to establish a security zone in the area of the Georgian-Russian border . . . and does not put an end to the bandit sorties and attacks on adjoining areas in the Russian Federation, we reserve the right to act

without any further qualification and without ascribing them to a particular State, a threat to international peace and security”).


240. Id.


242. It should be noted that over a year before the attacks of September 11, 2001, Russia warned the Taliban it was willing to take “preventative measures if necessary” if the Taliban did not cease sheltering and supporting Islamic rebels from Chechnya and the former Soviet Republics of Central Asia. See Michael R. Gordon, Russia warns Afghanistan not to aid rebel groups, N.Y. TIMES, May 25, 2000, at A7.

in accordance with Article 51 of the Charter of the United Nations, which lays
down every Member State’s inalienable right of individual or collective self-
defense.”244 Georgia’s fecklessness necessitated Russia’s use of force.

The United States responded by condemning the Russian incursions into
George but it did not reject the legal framework invoked by Russia. Rather, the
United States argued that Russia’s claim of self-defense was merely a pretext for
pursuing other goals. The United States acknowledged that the Georgian gov-
ernment’s lack of effective control over the Pankisi Gorge had permitted the
presence of “foreign militants and international terrorists” who represented a
threat to regional security.245 Nonetheless, unilateral Russian action was not jus-
tified because Georgia was making (with US military assistance) good faith ef-
forts to “root out Chechen fighters and criminal and international terrorist ele-
ments. These efforts signal Georgia’s commitment to restoring Georgian
authority in the Pankisi Gorge and dealing seriously with international terrorists
linked to al-Qa’ida.”246 In contrast to Georgia’s good faith efforts to “reassert
control in the Pankisi Gorge”, the United States suggested that Russia had acted
in bad faith. The United States contended that the intervention was prompted in
part by “Russia’s displeasure with Georgia’s commitment to the East-West en-
ergy transportation corridor” as well as “the fact that some in Russia viscerally
dislike [then Georgian President Eduard] Shevardnadze.”247 Russia’s prior pat-
tern of conduct towards Georgia also undermined Russia’s legal claims. This
pattern included “Russia’s periodic cutting off of Georgia’s winter gas supply,
Russia’s stalling of negotiations on political settlement in the break-away Geor-
gian region of Abkhazia, and its delaying of negotiations to meet CFE Istanbul
commitments for the withdrawal of Russian military forces still on Georgian ter-
ritory.”248 Thus, the United States and Russia appear to agree on the legal prin-
ciple but disagree on its application to the facts.

3. Israel-Lebanon

The legal arguments following Israel’s 2006 invasion of southern Lebanon
suggest further formal recognition of the security threat posed by state weakness
and non-state actors as well as an acceptance of a right to self-defense. During
July and August 2006, Israel engaged in a sustained aerial and ground assault
upon targets in southern Lebanon in response to a transborder raid launched by
Hezbollah. In complying with its Article 51 reporting requirement, Israel at-

244. Id.
245. Statement by B. Lynn Pascoe, Dept. Sec. State European and Eurasian Affairs, State De-
partment (Sept. 27, 2002), available at http://www.america.gov/st/washfile-
english/2002/September/20020927155334benner@pd.state.gov7.893008E-02.html.
246. Id.
247. Id.
248. Id.
tributed responsibility for Hezbollah’s raid to the Government of Lebanon.249 However, Lebanon’s responsibility arose not from any affirmative act on its part. Rather, Israel attributed Lebanon’s responsibility to the fact that the “ineptitude and inaction of the Government of Lebanon has led to a situation in which it has not exercised jurisdiction over its own territory for many years.”250 Therefore, Israel maintained that it had “the right to act in accordance with Article 51 of the Charter of the United Nations and exercise its right of self-defense when an armed attack is launched against a Member of the United Nations.”251

In response, Lebanon conceded that it lacked effective control over the territory in southern Lebanon in which Hezbollah was based, but suggested that such lack of control absolved it of responsibility. “The Lebanese Government was not aware of the events that occurred and are occurring on the international Lebanese border. The Lebanese Government is not responsible for these events and does not endorse them.”252 The Lebanese Government called upon the Security Council to take up the situation.

The resolution that the Security Council ultimately issued in response to this conflict is significant because it recognized Lebanon’s ineffective control over its territory and implicitly legitimized Israel’s claim of self-defense. More specifically, the Security Council acknowledged Israel’s jus belli claim, accepting the basis for hostilities by emphasizing the “need to address urgently the causes that have given rise to the current crisis,”253 including the weakness of the Lebanese state and the strength of Hezbollah. The Security Council therefore called for the “Government of Lebanon . . . to extend its authority over its territory, through its own legitimate armed forces, such that there will be no weapons without the consent of the Government of Lebanon and no authority other than that of the Government of Lebanon.”254 The Council emphasized the need for “the extension of the control of the Government of Lebanon over all Lebanese territory . . . [and] for it to exercise its full sovereignty, so that there will be no weapons without the consent of the Government of Lebanon and no authority other than that of the Government of Lebanon.”255 The resolution used similar language, calling for “the disarmament of all armed groups in Lebanon, so that . . . there will be no weapons or authority in Lebanon other than that of the Leba-

250. Id.
251. Id.
254. Id.
255. Id.
nese State.” Although the Security Council ultimately decided that the scope of Israel’s military action was disproportionate, most of the key players (Israel, Lebanon and the Security Council) acknowledged the significant threat to security posed by state weakness and independent non-state actors such as Hezbollah, who exploit the lacuna between positive and negative sovereignty.

4. Colombia-Ecuador

As with the debate over Israel’s strikes against Hezbollah, the international reaction to Colombia’s raid upon targets in Ecuador is another case of major powers tacitly recognizing a right to cross-border defensive action against independent non-state actors. Colombia resorted to force in the context of a once internal armed conflict that had spilled over the country’s borders as the government’s non-state adversaries sought refuge in neighboring countries. As with the Soviet Union’s incursion into Mongolia and Britain’s raid into the United States, Colombia’s raid into Ecuador targeted insurgent forces exploiting the ungoverned territory of a neighbor to mount attacks against a neighboring state. In this case, the guerillas were the Revolutionary Armed Forces of Colombia (FARC). On March 1, 2008, Colombian military forces responded by launching an aerial and ground assault upon a FARC base located on the Ecuadorian side of the border. The assault resulted in the death of Raul Reyes, a FARC commander, among others.

Colombia justified its incursion into Ecuadorian territory on the grounds that Ecuador “violated international norms which prohibit countries from harboring terrorists.” Colombian officials stressed that “the real issue is not that ‘Raul Reyes’ was in his pajamas and not in his military fatigues when he was killed; the issue is that he felt comfortable enough in his Ecuadorean base to feel that he could sleep in his pajamas.” Ecuadorian officials conceded that their

256. Id.
258. RABASA, supra note 13 (describing the FARC’s refuges in border regions).
260. STAFF OF S. COMM. ON FOREIGN RELATIONS, 110TH CONG., PLAYING WITH FIRE:
state was incapable of controlling its border region to exclude the FARC.\textsuperscript{261} Colombia later released information (substantiated by Interpol) indicating that FARC presence on Ecuadorian territory was not only the result of the lack of state capacity but also of active collusion. A document captured during the military raid contained an offer by an Ecuadorian official “to transfer police and army commanders in the area who proved hostile to the FARC.”\textsuperscript{262} Despite outrage by Ecuador and its ally Venezuela and condemnation by the Organization of American States, neither the Security Council nor the General Assembly took any action regarding the attack.

G. Summary of State Responses and International Reactions

States have accepted, if not always endorsed, defensive actions against non-state actors. They have deemed such interventions lawful when transborder uses of force are proportional and necessary.\textsuperscript{263} It appears to be irrelevant whether a host state is unwilling or unable to exclude violent transnational actors. Rather than lowering the bar for state attribution to negligence or strict liability standards (e.g., responsibility for harboring), many states recognize a right to self-defense \textit{vis-à-vis} non-state actors irrespective of host state responsibility.

The hostility of many states during decolonization to such defensive uses of force, particularly by the post-colonial entities and Communist states supporting non-state actors in the context of “wars of national liberations,”\textsuperscript{264} has waned as both these post-colonial entities and Russia have themselves become the targets of independent armed groups acting across international borders. This shift in the perceived security interest of post-colonial entities has occasioned a changed understanding of the use of force, as illustrated by Uganda’s arguments in \textit{Armed Activities on the Territory of the Congo} (Dem. Rep. Congo v. Uganda), 2005 I.C.J. 168 (Dec. 19).

IV. APPLYING CUSTOMARY PRINCIPLES

A narrow, state-centric interpretation of the law of self-defense fails as a matter of policy and law. A state-centric regime does not strike the proper bal-

\textsuperscript{261} COLOMBIA, ECUADOR, AND VENEZUELA, 5 (Comm. Print 2008).
\textsuperscript{263} The ICJ has recognized that necessity and proportionality are customary limitations upon the use of defensive force. \textit{Military and Paramilitary Activities (Nicar. v. U.S.)}, 1986 I.C.J. at ¶ 194.
\textsuperscript{264} Reisman, \textit{Private Armies in a Global War System}, supra note 48, at 252, 258 (discussing the embrace by Communist powers and Third World states of “private armies” as instruments of policy in wars of national liberation).
Finucane: Fictitious States, Effective Control, and the Use of Force Against

FICTITIOUS STATES, EFFECTIVE CONTROL

ance between responding to legitimate security threats and restricting pretextual, bad faith uses of force.

Nor is a state-centric regime international law. In practice, many governments recognize power and threats as they actually exist. As demonstrated in Part III, governments have for centuries faced security threats posed by non-state actors exploiting the lacuna between positive and negative sovereignty. In order to respond to these threats, states long ago developed principles regulating the transborder defensive use of force against non-state actors. The principles that emerged from the Caroline and similar incidents, namely necessity and proportionality, strike the proper balance between responding to legitimate security threats and restraining illegitimate and excessive force.265 State practice and opinio juris demonstrate that the right of self-defense against non-state actors exists in customary law irrespective of state involvement.

The following section details how the principles derived from these historic incidents should be operationalized to regulate future uses of force against non-state actors in order to vindicate the central policy aim of minimizing violence and promoting stability.

The key parameters of this framework are:
1) (Ir)relevance of consent
2) Establishing necessity
3) Geographic restrictions upon resort to force against a non-state actor
4) Differentiation between state and non-state actors
5) Temporal scope of recourse to force against non-state actors

A. Consent: Desirable, Not Dispositive

As a matter of policy, the consent of a host state may be desirable prior to the recourse to force against a non-state actor on its territory. As a matter of law, however, such consent is not dispositive as to whether a state may resort to defensive force.

1. Policy

From a policy perspective, the consent of the host state to defensive measures against non-state actors on its territory is desirable. The US targeted killing program in Pakistan illustrates the benefits of such consensual intervention.266 The host state may provide logistical and intelligence support for defen-

266. Although Pakistan has not publicly consented to United States use of force on its territory, there is compelling evidence that lethal operations are undertaken on Pakistani territory with consent from the highest levels of the Pakistani government. In the fall of 2008, the New York Times reported that an Obama administration official said that “. . . [n]o tacit agreement had been reached to allow increased Predator [drone] strikes [in Pakistan] in exchange for a backing off from additional American ground raids” referring to a September 3, 2009, raid by United States Special Operations
sive action. For example, Pakistan allows the United States to dispatch unmanned aerial vehicles (UAVs) from its Shamsi airbase in Baluchistan against targets in the Federally Administered Tribal Areas. Yemen and Pakistan also provide the CIA with human intelligence that drives at least some of the targeting by these UAVs.

Further, consensual intervention avoids the possible escalation of a conflict between the defending state and the non-state actor into an international armed conflict between the defending state and the host state.

As desirable as consent may be, the consent of the host state may not always be forthcoming or may be restricted in scope. The above example of Pakistan is also illustrative of the problematic nature of host state consent. Specifically, the scope of Pakistan’s consent is limited insofar as the interests of US and Pakistani authorities diverge, as described in the paragraph below.

First, the foreign military intervention on Pakistani territory undermines the domestic legitimacy of the regime in Islamabad. The most obvious consequence of this legitimacy concern is the failure of Islamabad to formally acknowledge that UAV strikes are conducted with its consent. Instead, Pakistani authorities continue to issue increasingly implausible denials and pro forma protests over violations of Pakistani “sovereignty.” Paki-269 Pakistani concern over domestic legitimacy limits the scope of US action by restricting UAV strikes to “boxes” that include the Tribal Areas but exclude cities such as Quetta, Taliban leader Mullah Mohhammed Omar’s base in Pakistan. Concerns over plausible deniability personnel into Pakistan. See Mark Mazzetti & Eric Schmitt, U.S. Takes to Air to Hit Militants Inside Pakistan, N.Y. TIMES, Oct. 27, 2008, available at https://www.nytimes.com/2008/10/27/washington/27intel.html [hereinafter U.S. Takes to Air]. According to the story, “[P]akistani officials have made clear in public statements that they regard the drone attacks as a less objectionable violation of Pakistani sovereignty.” Id. Subsequent media reports claim that many of the American drone strikes originate from a base in Pakistan and that their increased frequency is the result of a March 2009 deal between the United States and Pakistan, which provides Pakistanis greater control of the targets. See James Kiffield, Wanted: Dead, NATIONAL JOURNAL, Jan. 9, 2010, at 6 (“[f]or domestic political consumption, Pakistan’s leaders promote the image of CIA agents flying drones from its American headquarters, but the program clearly involves a high degree of involvement by Americans inside Pakistan, and by the Pakistani government”).


270. Greg Miller, U.S. Wants to Widen Area in Pakistan Where It Can Operate Drones, WASH.
also appear to underlie Pakistan’s refusal to allow actions by American Special Operations forces against Al Qa’ida fighters on its territory, such as the SEAL raid against Osama Bin Laden in Abbottabad.271

Second, as a result of Pakistan’s divergent strategic interests from the United States, Pakistan considers certain non-state actors to be assets, whereas the United States regards these same actors to be threats. For example, Pakistan considers both the Quetta Shura and the Haqqani Network to be assets for achieving strategic depth in Afghanistan vis-à-vis its archrival India, notwithstanding their ties to Al Qa’ida and their continued attacks upon US and Afghan forces.272 In light of these broader interests, Pakistani cooperation with US efforts in Afghanistan has been less than complete.

2. Law

As a matter of international law, the consent of a host government, whether Pakistani or Yemeni, is not dispositive. The United States may, under customary international law, resort to force against Al Qa’ida and associated non-state actors on Pakistani or Yemeni territory, with or without the consent of these governments.

Two of the incidents analyzed in Part III illustrate that under customary international law, states enjoy a right of self-defense vis-à-vis non-state actors irrespective of the consent of the host state.273 Both the United States’ 1916 expedition into Mexico against Pancho Villa and Turkey’s 1995 intervention against Kurdish militias in northern Iraq were initially premised upon the consent of the host state. However when the host states withdrew their consent, both the United States and Turkey continued to take defensive measures against the non-state actors and justified these as self-defense.274 In light of the international response, the lawfulness of Turkey’s defensive measures in Iraq is particularly clear. Despite Iraq’s protests, international reaction ranged from understanding to muted concern over proportionality.

271. U.S. Takes to Air, supra note 266. According to the story, “[P]akistani officials have made clear in public statements that they regard the drone attacks as a less objectionable violation of Pakistani sovereignty.”


273. Finch, supra note 173; Olson, supra note 222.

274. Olson, supra note 222.
In the context of the United States’ use of force in Pakistan, the lawfulness of the UAV strikes, Special Operations missions, or other measures against Al Qaeda or associated groups is not contingent upon either the existence or scope of the Pakistani government’s consent. The United States may lawfully resort to force against non-state actors in Pakistan—or Yemen or Somalia—on the independent basis of self-defense.

B. Establishing Necessity

Having established that consent is not required for lawful self-defense, what is required for lawful self-defense? This threshold inquiry regarding the unilateral use of force against a non-state actor is one of necessity. Such force is lawful only when it is necessary, leaving “no choice of means.”\(^{275}\) Necessity exists when alternative means of neutralizing a security threat are inadequate.

The necessity of self-defense turns on the reality of the threat posed by non-state actors, as well as the ability and willingness of their host states to take action against them. The host state’s ability to abate the threat hinges upon the host state’s effective control of its territory. In a fictitious state, where the central authorities have relinquished or simply failed to consolidate their monopoly over violence, recourse to transborder force may be the only means of abating the threat. As the British authorities observed in the Caroline incident, recourse to feckless or indifferent authorities in such circumstances would be futile.\(^{276}\)

Thus, when the territorial state hosting a violent non-state actor is willing and able to adequately address the threat, the unilateral use of force will be unnecessary and hence unlawful. The availability of alternative means such as diplomatic entreaty or extradition would thus render self-defense unlawful. If the host state is unable or unwilling to address the threat, the use of defensive force is necessary and hence lawful.\(^{277}\)

The intervening state carries the burden of establishing that alternative remedies such as diplomatic entreaties and law enforcement measures are insufficient to abate the threat posed by non-state actors operating from the territory of another state. Absent such a showing, the unilateral recourse to transborder force is presumptively unlawful.

A few examples illustrate this point. The Russian incursions into Georgia’s Pankisi Gorge in 2002 were unnecessary and hence unlawful. As described above, at the time the US military was assisting Georgian authorities in countering the same Chechen threat to which Moscow claimed to be reacting. Notwithstanding Russian declarations to the contrary, Georgia was in fact confronting

\(^{275}\) Jennings, supra note 142, at 82 (citing H. Ex. Doc. No. 74, 25th Cong., 2nd Sess.).

\(^{276}\) Id. at 90.

the non-state actors within its territory.

In contrast, necessity exists when entreaties to the authorities of the host state would be counterproductive to the suppression of the security threat. For example, Colombia presented evidence of collusion between Ecuadorian authorities and the FARC in the case of Colombia’s raid against Raul Reyes’ camp inside Ecuador. Colombia argued that requests to the Ecuadorian authorities only would have afforded FARC advance warning.

The United States adopted a similar position with respect to the SEAL raid upon Abbottabad. The United States did not to notify the authorities in Islamabad prior to the raid, much less seek their consent, because of concerns over collusion between elements of the Pakistani security services and Al Qa’ida.

C. The Geographic Scope of the Use of Force Against a Non-State Actor

Necessity establishes not only when a defending state may resort to force against a non-state actor, but also where the defending state may act. Put differently, necessity defines the legal limits of the battlefield in a non-international armed conflict. Establishing the boundaries of this battlefield is especially crucial in conflicts involving diffuse and loosely organized entities such as Al Qa’ida or highly mobile actors such as Lord’s Resistance Army, which operate in the ungoverned territories of multiple countries. As illustrated by the controversy over the “Global War on Terror” and especially the United States’ ongoing program of targeted killings via missile strikes from UAVs, such boundaries are needed.

Armed conflict is not football. Both with respect to fact and law, the physical boundaries of the playing field in an armed conflict are not permanently fixed prior to the onset of action. Contrary to the assertion of the UN Special Rapporteur on Extrajudicial Killings, the boundaries of an armed conflict are not determined by the law of armed conflict or the prior occurrence of hostilities in a given region.

Rather, the boundaries of where a state may resort to force are defined by long standing principles of jus ad bellum and the location of hostilities. Although not usually couched in these terms, what defines the lawful battlefield is the principle of necessity. The region of conflict is that area where a state may lawfully resort to force against a state or non-state actors. Only in the region of

278. FARC Documents, supra note 262.

279. Alan Cowell, Pakistan Sees Shared Intelligence Lapse, N.Y. TIMES, May 4, 2011, available at https://www.nytimes.com/2011/05/05/world/asia/05react.html (noting that Pakistani officials are angry about CIA director Leon Panetta’s claim that the United States did not inform Pakistani officials in advance of the raid because it might have leaked, allowing Bin Laden to escape).


conflict may a state lawfully violate the sovereignty of other states. Defining the region of conflict on the basis of necessity comports with the traditional rules regulating the belligerency or neutrality of states during interstate conflict. During interstate armed conflict, neutral states enjoy a right of territorial inviolability vis-à-vis belligerent states. However, respect by the belligerents of the neutral’s negative sovereignty is contingent upon the fulfillment of corresponding duties. Neutral states must not allow the movement of belligerent troops or military materiel across its territory. Should belligerent troops enter the territory of a neutral state, that state is required to intern the troops. Neutral states are also under an obligation to prevent the formation or recruitment of belligerent military units upon its territory.

The logic of these well-established rules of interstate conflict applies equally to conflicts involving non-state actors. The lawfulness of a belligerent state’s violation of a neutral state’s territory hinges upon the reality of the threat and whether the neutral state has or will take adequate measures to abate this threat. When due to incompetence or connivance a neutral state fails to incapacitate a violent non-state actor on its territory, it jeopardizes its negative sovereignty. The neutral state’s territory may become part of the lawful region of the conflict.

Applying this rule to the United States’ current conflict with Al Qa’ida reflexively imposes limits upon the scope of the “Global War on Terror.” With respect to the US targeted killing program, killings within the territory of a host state are lawful where the central authorities are on notice and have failed to intern Al Qa’ida fighters. Therefore, the use of force in violation of the host state’s negative sovereignty would be lawful in Pakistan’s Federally Administered Tribal Areas, Somalia, or the Yemeni hinterland. In contrast, a missile strike from a UAV against an Al Qa’ida member on the streets of London (even with zero collateral casualties) would be unlawful. Such action would be an unnecessary violation of the United Kingdom’s negative sovereignty because the British

283. Id. at arts 2, 5.
284. Id. at art. 11.
285. Id. at arts. 4, 5.
286. See Staff of S. Comm. Relations, 111th Cong., Al Qaeda in Yemen and Somalia: A Ticking Time Bomb, 3, 8, (Comm. Print 2010) (“Both Yemen and Somalia have weak central governments that exercise little or no control over vast swaths of their own territory and forbidding, harsh terrains that would make it virtually impossible for U.S. forces to operate freely. They have abundant weapons and experience using them on the battlefield . . . . As Al Qaeda members continue to resist U.S. and Pakistani forces along the Afghanistan-Pakistan border, some of their comrades appear to be moving to Yemen and Somalia, where the political climate allows them to seek safe haven, recruit new members, and train for future operations. . . . There are parallels between Pakistan and Yemen, according to U.S. counter-terrorism officials, military leaders, and policymakers. Both have become havens for significant numbers of Al Qaeda fighters formerly active in Afghanistan. Both have weak central governments that have difficulty controlling vast swaths of their own territory and populations that are often hostile to the United States”).
authorities are willing and able to incapacitate Al Qa’ida fighters on their soil. The Tribal Areas and Yemen are within the region of conflict, but London is not.

In summary, necessity defines the outer geographical boundaries of the region of conflict. If a defending state can counter a non-state threat in a given area by means other than unilateral force, such as diplomatic entreaty, then that area lies outside the region of conflict. Thus, for example, necessity would establish the geographic limits of the war on terror.

D. Differentiation vs. Attenuated State Responsibility

The principles of necessity and proportionality dictate not only whether and where a state may resort to force against non-state actors, but also how. These principles favor differentiating between non-state actors and the personnel and infrastructure of their host states. Necessity and proportionality militate against premising the defensive use of force upon an expanded scope of host state responsibility for the acts of private armed groups. Expanded responsibility holds the host state to a standard of negligence or strict liability. Basing self-defense upon expanded host state responsibility reflects a state-centric view that fails to account for the reality of how power and control actually exist. While a fictitious state may act wrongfully and breach an international duty for which it is liable for restitution, as a matter of law and policy a distinction should be drawn between wrongful behavior and dangerous conduct.

When a truly independent non-state actor threatens a state, the necessity of the use of force relates solely to the non-state actor, not to the territorial host state. Targeting the infrastructure or personnel of the host state is unlawful because it is unnecessary and disproportionate.

Failure to differentiate between the state and the non-state actor will usually render unilateral defensive action unlawful. For example, in its 2006 military campaign in Lebanon, Israel failed to differentiate between its non-state adversary, Hizbollah, and the Lebanese state, as evidenced by the Israeli destruction of the Beirut airport and extensive civilian casualties. This disproportionate use of force, unrelated to the security threat posed by Hizbollah, rendered the


288. See National Security Strategy, supra note 10, at 5 (“We make no distinction between terrorists and those who knowingly harbor or provide aid to them”).


Israeli invasion unlawful as a matter of *jus ad bellum*.  

There should be a strong presumption that necessity and proportionality require differentiation between state and non-state personnel and infrastructure, and that the occupation of territory is presumptively unlawful. Nonetheless, abating some persistent security threats may require expansive intervention and territorial occupation. In order to deny violent non-state actors a sanctuary, a fundamental reordering of local conditions may be necessary where these actors represent a chronic threat. Such interventions could take the form of a temporary protectorate, such as President Buchanan proposed for northern Mexico. In extreme cases, changing the regime in the capital of the fictitious state may be necessary to prevent the return of the threatening non-state actor. Thus, the installation of the Northern Alliance in Kabul after 9/11 was necessary to establish a local ally in Afghanistan willing to suppress the persistent threat from Al Qaeda. The dictates of necessity and proportionality are highly fact-specific and “can ultimately be subjected only to that most comprehensive and fundamental test of law, reasonableness in a particular context.”

Differentiation based on necessity and proportionality, in contrast to attenuated state responsibility, also serves to promote global stability more generally. Although it is not the default form of political organization, the effective state is still the best entity to provide basic public goods, especially domestic security. Whereas fictitious states may occasionally imperil their neighbors through spill-over, state weakness poses a much greater threat to the inhabitants of the fictitious state’s own territory. The DRC illustrates the problem, as millions of Congolese civilians have died in the violence and disruption resulting from their state’s weakness. By distinguishing between whatever weak, inchoate, or embryonic state may exist within a territory and the threatening non-state actor, defending states can preserve incipient order within a territory. Differentiation between the infrastructure and personnel of the nominal state and the non-state actor preserves order while real security threats are addressed. Such differentiation would precede the distinction between lawful and unlawful targets undertaken in accord with principles of *jus in bello*.

Underlying differentiation is an analytical distinction between wrongful conduct and dangerousness. Although a fictitious state may act wrongfully and

---

291. Qatar characterized Israel’s use of force as beyond its stated objective, given the civilian nature of the Israeli targets. U.N. SCOR, 61st Sess., 5493d mtg. at 14, U.N. Doc. S/PV.5493 (Jul. 21, 2006). Lebanon clearly rejected Israel’s contention to be acting primarily against non-State terrorist targets. It has been very clear from the beginning that it was not Hezbollah that was the target. It was Lebanon that was the target. Infrastructure was targeted and hundreds of civilians were killed before Israel even took up any campaign against Hezbollah and its positions. *Id.* at 6. The Secretary General also characterized Israel’s use of force as collective punishment of the Lebanese people. *See* U.N. SCOR, 61st Sess., 5492nd mtg. at 3, U.N. Doc. S/PV.5492 (Jul. 20, 2006). Argentina qualified Israel’s use of force as collective punishment. U.N. SCOR, 61st Sess., 5493d mtg. at 9, U.N. Doc. S/PV.5493 (Jul. 21, 2006).

breach a duty under international law by failing to suppress independent armed forces on its territory, it is not the state itself that poses the threat. Prior to the UN Charter, a victim state could have resorted to war to redress wrongs it had suffered as a consequence of the fictitious state’s breach.293 Yet such retributive war was distinct from self-defense, which focused exclusively upon the threat posed by the non-state actor. This distinction is evident in the nineteenth century diplomatic correspondence of the United States. As Secretary of State Forsyth observed, American military intervention against marauders in Mexico, “[r]ests upon principles of the law of nations, entirely distinct from those on which war is justified—upon the immutable principles of self-defense—upon the principles which justify decisive measures of precaution to prevent irreparable evil to our own or to neighboring people.”294 Forsyth argued that such “decisive measures of precaution” that solely targeted the marauders were preferable over a war against Mexico over Mexico’s irresponsibility.295

The great contribution of the UN Charter regime to world public order was the proscription of casus belli other than self-defense. Recourses to force premised upon attenuated state responsibility are responses to wrongful conduct and not necessarily an imminent threat. The use of force premised upon the breach of an international duty, rather than in response to a direct threat, erodes the restriction of jus ad bellum to self-defense.

E. Temporal Scope: Armed Attack as Evidentiary Standard

The principle of necessity should also inform the much debated “armed attack” requirement of Article 51. The armed attack requirement (if it exists) has particular significance with respect to violent non-state actors, such as Al Qaeda, who are not susceptible to the traditional logic of deterrence.296 Precisely what uses of force constitute an “armed attack” and whether or not such an attack is the sine qua non of self-defense are hotly contested, in particular whether the right to self-defense applies against independent non-state actors

293. OPPENHEIM, supra note 46, at 67 (stressing recourse to war to vindicate rights: “A State may be driven into war because it cannot otherwise get reparations for an international delinquency, and may then maintain that it exercised by war nothing else than legally recognized self-help”); Brownlie, supra note 19 at 20-22 (describing war as a “means of obtaining redress in the absence of a system of international justice and sanctions”).


295. Id. at 420-421; see also OPPENHEIM, supra note 46, at 67 (stressing recourse to war to vindicate rights: “A State may be driven into war because it cannot otherwise get reparations for an international delinquency, and may then maintain that it exercised by war nothing else than legally recognized self-help”).

296. Some groups, such as Hezbollah, which function as unrecognized de facto states may be more amenable to the territorial logic of deterrence. See Nadim Ladiki, Hezbollah Gives Israel More Clues in Strategy, Reuters, Nov. 5, 2007, available at http://www.reuters.com/article/idUSL05623345 (noting that Hezbollah leader Hassan Nasrallah “said shortly after the [2006] war that he would not have ordered the attack had he known the Israeli retaliation would be so fierce”).
and whether actions variously termed preventative, precautionary, preemptive, anticipatory, or interceptive self-defense are lawful. 297

Article 51 should be interpreted in the context of the motivating purpose of the UN Charter, in a way as to reconcile it with the prior and subsequent customary law of self-defense, and that restrains the use of force while allowing for legitimate security measures. Read in this fashion, the armed attack requirement of Article 51 provides an objective basis for the determining validity of a claim of self-defense. 298

However, because wrongfulness is no longer a causus belli, an armed attack should not be read as evidence of a requirement of prior trespass, but rather as evidence of a threat. As noted in 1838 by the legal officers of the British Foreign Office with respect to the Caroline incident, the defensive use of force in response to an attack is justified not by the attack’s wrongfulness, but by the threat it evinces. Such a response is “necessary as a measure of precaution for the future and not as a measure of retaliation for the past. What has been done previously is only important as affording irrefutable evidence of what would occur afterwards.” 299

A focus on the necessity of defense against future attacks rather than the wrongfulness of prior injury suggests a general standard that fulfills the evidentiary function of an armed attack. 300 An objective reasonableness standard premised on propensity and capability balances the dangers of false positives and false negatives. Such a standard provides for good-faith security measures while restricting pretextual intervention. 301 Yet, its workability applied to either state or non-state actors is completely contingent upon the quality of intelligence regarding the intentions and capabilities of violent actors, a fact painfully illustrated by the 2003 invasion of Iraq.

V. CONCLUSION

This Article has described the reality of fictitious statehood, the persistent security threat posed by non-state actors exploiting this fiction, and the forceful
response of states to this threat. In describing a number of episodes of self-defense involving weak states and non-state actors, this Article has situated current US military operations against Al Qa’ida in their proper historic context. The pre-Charter principles illustrated by these incidents remain relevant and can be used to structure future defensive measures.

However, the challenges for international order posed by the gap between the theory and practice of statehood are much broader than the aspect explored in this Article. Consequently, there is a larger body of international law that bridges the gap between the theory and reality of statehood by accommodating the fact of state weakness. This body of law will continue to develop despite the persistent enthusiasm of some for nation building. There are no ready “fixes” for fictitious states.
The Evolving Definition of the Refugee in Contemporary International Law

William Thomas Worster
The Evolving Definition of the Refugee
In Contemporary International Law

By
William Thomas Worster*

I.
INTRODUCTION

The Refugee Convention is one of the cornerstones of the larger human rights system for protecting vulnerable persons and yet it is also a very narrow instrument, protecting a very specific group of persons. This duality is reflected in refugee protection generally where, on the one hand, states appear to believe in a moral, humanitarian imperative to protect individuals seeking refuge, yet, on the other hand, they are reluctant to permit entry to all those persons falling under their responsibility. When we consider the contemporary definition of refugee, and how customary international law may supplement the definition of refugee, we see this same division of interests. If we were motivated strictly by human-centered interests, we would find a broadening of the definition, although perhaps with limited state compliance. If we were motivated strictly by state-centered interests, we might find a narrowing of the definition, although perhaps abandoning desperate individuals truly in need.

This Article will attempt to navigate between these perspectives to look, first, at how the definition may be broadening under customary international law, and second, at how the definition may be narrowing. Section II will address the evolving interpretation of the definition of refugee under the Refugee Convention, especially the evolving means for interpreting the Convention, to determine whether the conventional definition has outgrown its conventional shell. Section III will turn to customary international law proper, analyzing state practice and opinio juris on point. In particular, this Section will reflect on the role

* Lecturer, International Law, The Hague University of Applied Sciences, The Hague, The Netherlands; LL. M. (Adv.) in Public International Law, cum laude, Leiden University, Faculty of Law, Leiden, The Netherlands; J.D., Chicago-Kent College of Law, Illinois Institute of Technology, Chicago, Illinois; B.A., Modern European History, University of Kansas, Lawrence, Kansas. The author wishes to thank Maximilian L. Garrê for his assistance in locating many of the sources used herein.
of specially interested or specially affected states in the formation of customary international law and the growth of “subsidiary” protection. Also this Section will consider the contribution of the practice and opinio juris of international organizations in the frame of contemporary international law’s understanding of the contribution international organizations can make.

Section IV will turn to look at the opposite side of the coin: the ways in which customary international law may have narrowed the definition beyond the terms of the Refugee Convention. If we are open to considering the broadening role of customary international law, then we must equally be open to the narrowing role. Some of the provisions examined in this Section are perhaps not correctly considered aspects of the definition, for example, safe third country option and diplomatic assurances. These might be better understood as exceptions to the non-refoulement obligation. However, they will be considered nonetheless within the broader notion of the definition of those deserving refuge similar to the “exclusion” or public danger clauses in the Refugee Convention, which interact with the application of the definition in important ways, such as in the case of child soldiers. These additional concerns are developing under customary international law to increasingly restrict the availability of refuge. Section V will conclude by summing up the findings of the analysis.

II. INTERPRETATION OF THE DEFINITION IN THE REFUGEE CONVENTION

The beginning for any inquiry into the definition of a refugee is the Refugee Convention and its protocol. The Refugee Convention specifies that a person qualifies as a refugee if (1) the person has already been considered a refugee under prior treaty arrangements or (2) the person is outside the country of his nationality (or not having a nationality) and is unable or unwilling to avail himself of the protection of that country due to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership in a particular social group or political opinion. The latter are commonly referred to as the “includ...
Worster: The Evolving Definition of the Refugee in Contemporary Internatio

96 BERKELEY JOURNAL OF INTERNATIONAL LAW [Vol. 30:1

sion” clauses. Failure to qualify under the former does not defeat the possi-

This Article will not address in depth the requirements of persecution or

social group membership, other than to note in passing that the classification of

social group membership appears to be broadening to consider cultural chang-
es. Also one of the most significant developments in interpretation is that some

Convention [Refugee Convention] in Article 1A(2), does not confer that status. The first matter to be

established under the Article is that the claimant is outside the country of his nationality owing to a

well-founded fear of persecution. That well-founded fear must exist at the time his claim for refugee

status is to be determined; it is not sufficient as a matter of the ordinary meaning of the words of the


4. Refugee Convention, art. 1A(1) (“Decisions of non-eligibility taken by the International Refugee Organization during the period of its activities shall not prevent the status of refugee being accorded to persons who fulfil the conditions of paragraph 2 of this section”).

tion Act, (2005) (recognizing gender-related persecution); Bundesverwaltungsgericht [Fed. Admin. Ct], Jan. 18, 1994, Case 9 C 48/92 (Ger.); BVerwG, Mar. 15, 1988, Case 9 C 278/86 (Ger.); Verwaltungsgerichtshof [Higher Administrative Court] (13th Sen.), Hessen, Nov. 14, 1988, Case 13 TH 1094/87 (Ger.); VGH (10th Sen.), Hessen, Aug. 21, 1986, Case 10 OE 69/83 (Ger.); Verwaltungsgericht [Admin. Ct.], Dresden, Feb. 1, 2005, Case A 7 K 31131/03 (Ger.) (recognizing gender persecution); VG, Frankfurt am Main, Feb. 19, 2005, Case 5 E 7021/03.A(3) (Ger.) (same); VG, Frankfurt am Main, Mar. 29, 1999, Case 9 E 30919/97.A(2) (Ger.); VG, Frankfurt am Main, Oct. 23, 1996, Case 5E 35532/94.A (3) (Ger.); VG, Magdeburg, June 20, 1996, Case 1 A 185.95
states have begun to recognize non-state actors as potential sources of persecution, rather than only states. These two developments suggest that at least the applicable inclusion provisions within the definition of refugee appear to be interpreted dynamically with a focus on the object and purpose of the Refugee Convention, and the contemporary meaning of the treaty text.

However, there are also several provisions that defeat refugee status for otherwise qualified individuals. Individuals who do qualify under (1) or (2) above may fall outside the definition if they have voluntarily re-availed themselves of the protection of their country of nationality; have voluntarily re-acquired the nationality of their state; or voluntarily re-established themselves in the state which they left or remained outside of owing to fear of persecution. In the case of multiple nationals, the individual must qualify as a refugee as per all the states of nationality. This latter possibility has been interpreted so far as to refuse refugee sta-

---

6. See Arthur C. Helton & Pamela Birchenough, Forced Migration in Europe, 20 Fletcher F. World Affairs 1996 (reporting on Austria, France, Sweden, and Switzerland; although Switzerland does often permit these individuals to remain on other humanitarian grounds).

7. Refugee Convention, art. 1C(1).

8. Id. at art. 1C(2). See also Cessation of Refugee Status (No. 2001/01/0499) [2003] VGH (Austl.).

9. Cessation of Refugee Status at art. 1C(3).

10. Id. at art. 1C(4).

tus to persons who have the potential nationality (perhaps even ethnicity) of an-
other state. However, this dual nationality (or residence) provision is not always strictly applied under municipal law, evidencing a purposive application of the terms of the convention. For example, North Koreans who avail themselves of South Korean nationality are considered firmly resettled in South Korea by the United States, but if they do not move to acquire South Korean nationality, they cannot be returned to the Korean peninsula on the theory that they could be considered South Korean nationals.

Persons will also not qualify if the persecuting circumstances within the

12. See e.g., Case No. MIG 2007: 33 ll, UM837-06 [Supreme Migrations Court] 2007-06-15 (Swed.); Rolleiv Solholm, Amnesty Accept Expulsions, NORWAY POST (Sept. 14, 2010), http://www.norwaypost.no/news/amnesty-accept-expulsions.html (holding that Serbs from Kosovo may be returned to their “homeland” in Serbia, not Kosovo; also reporting that Amnesty International has accepted the practice). Similarly, individuals do not qualify if they are recognized by the state of residence as having the rights and obligations of nationality of that state, even though not formally holding nationality. See U.N.H.C.R. PROTOCOL ON POLICY AND LEGAL ADV. SEC., DEPT. INT’L PROTOCOL, Background Note on the Application of the Exclusion Clauses: Article 1F of the 1951 Convention Relating to the Status of Refugees ¶ 9, reprinted in 17 INT’L J. OF REFUGEE L. 293 (2003) [hereinafter UNHCR, Background Note on the Exclusion Clauses]. Under Article 1E, the 1951 [Refugee] Convention does not “apply to a person who is recognized by the competent authorities of the country in which he [or she] has taken residence as having the rights and obligations which are attached to the possession of the nationality of that country.” The object and purpose of this Article can be seen as excluding from refugee status those persons who do not require refugee protection because they already enjoy greater protection than that provided under the 1951 [Refugee] Convention in another country apart from the country of origin where they have regular or permanent residence and where they enjoy a status that is in effect akin to citizenship.


relevant state have ceased to exist under the so-called “cessation” clauses. According to the Executive Committee of the United Nations High Commissioner for Refugees (UNHCR), the cessation clauses are to be applied sparingly. Further, in order to invoke a cessation clause:

There must have been a change in the refugee’s country of origin, which is fundamental, durable, and effective. Fundamental changes are considered as effective only when they remove the basis of the fear of persecution.

This provision is usually interpreted more broadly than even the UNHCR appears to advise and refugee status usually will not be considered to have ceased as long as the situation remains one of general danger or instability. In

---

15. Refugee Convention, art. 1C(5) & (6); 98/01/0502 v. Indep. Fed. Asylum Bd. [hereinafter UBAS], (VwGH [Admin. Ct.], Aust., Sept. 8, 1999); S.S. (Czech Rep.) v. Swiss Fed. Office for Refugees (Asylum Appls. Comm’n, Switz., Nov. 3, 1999). However, in order for a state to make a determination that a person does not qualify as a refugee under this provision, the state is obliged to first determine whether the individual qualifies as a refugee under the Convention. See R v. Spec. Adj’tor ex parte Hoxha, [2005] 1 W.L.R. 1063; [2005] 4 All E.R. 580, [2005] UKHL 19 (UKHL, Mar. 10, 2005, aff’g Hoxha & B., [2002] E.W.C.A. Civ. 1403, [2002] All E.R. (D) 182 (Oct)) ("[Art] 1C(5), a cessation clause, simply has no application ... at any stage unless and until it is invoked by the State against the refugee in order to deprive him of the refugee status previously accorded to him").


17. See e.g., Denmark: Case unreported and unnamed (Appls. Bd., 2005) reported at ECRE Country Report 2004, supra note 5, rev’g the application of the cessation clause in the case of Afghan refugees since the situation in Afghanistan was not yet sufficiently stable; Germany: Oberverwaltungsgericht [OVG] Saarland [High Admin. Ct.] Sept. 21, 2004, Case 1 R 15/04 (Ger.) (holding that there was no longer a risk of persecution for ethnic Albanians in Kosovo, justifying cessation of refugee status); OVG Schleswig-Holstein, June 16, 2004, Case 2 LB 54/03 (Ger.) (holding that Afghanistan was sufficiently stable after the fall of the Taliban); Verwaltungsgerichtshof [VGH] Hessen, Feb. 10, 2005, Case 8 UE 216/02.A (Ger.) (holding that conditions of stability must be considered and that Afghanistan after the Loya Jirga in June 2002 was sufficiently stable to provide for cessation of refugee status); Verwaltungsgericht [VG] Ansbach, June 29, 2004, Case 19 K 03.31666 (Ger.) (holding that the particularized living conditions were so poor that refugee status could not be withdrawn); Case 6 A 524/04 (VG Braunschweig, Feb. 17, 2005) (same); Case 7 K 2389/01.A (VG Düsseldorf, Feb. 25, 2004) (same); Case 7 K 1517/00.A (VG Frankfurt/Oder, Mar. 2, 2004) (same); Case 9 E 7411/03.A(2) (VG Frankfurt am Main, Jan. 24, 2005) (same); Case 1 E 495/04.A(V) (VG Frankfurt am Main, May 27, 2004) (same); Case 5 E 4425/03.A (VG Frankfurt am Main, Feb. 27, 2004) (same); Case 9 K 4856/03.A (VG Minden, Apr. 26, 2004) (same); Case 7 E 2245/03.A(V) (VG Wiesbaden, Nov. 4, 2004) (same); Case W 7 K 04.30411 (VG Würzburg, Aug. 20, 2004) (same); Romania: Ord. No. 102/2000, Ordinance on the Status and Regime of Refugees in Romania (Nov. 1, 2000), as amended by Ord. 43/2004, Amending and Completing Government Ordinance No. 102/2000 on the Status and Regime of Refugees in Romania (Jan. 29, 2004) (Rom.) (modifying the rules of cessation, withdrawal and cancelling refugee status to reflect the Refugee Convention).

Switzerland: In re O.D. und Kinder, Eritrea, Case 2004, No. 26 Asylrekurskommission [ARK] [Asylum Appls. Comm’n] May 26, 2004 (Switz.) (Eritrea-Ethiopian borderland of the Senafe/Debub region is considered unreasonable due to the humanitarian situation); In re A.G.M. et famille, Angola, Case 2004, No. 32 [Comm’n des Recours des Réfugiés] Sept. 17, 2004 (provinces of Cabinda, Uige, Malanje, Lunda Norte, Lunda Sul, Bié, Mexico and Cuando-Cubango of Angola not to be reasonable, but removal to Luanda and easily accessible cities in the provinces of Cunene, Huila, Namibe, Benguela, Huambo, Cuanza Sul, Cuanza Norte, Bengo and Zaire is considered to be reasonable under certain circumstances (i.e., single men or couples, without young children, no serious medical
general, however, the application of the Refugee Convention appears to be customarily interpreted dynamically with very heavy reliance on its apparent objective and purpose of protecting individuals in need.

Nevertheless, in several cases, general instability was found insufficient to prevent cessation of status, while in others general instability was not even a factor considered unless it was due to lingering effects of the persecution. Yet in other cases, general instability was considered only in the context of a viable internal flight alternative, a topic addressed later. These cases do not appear to arise widely in different jurisdictions, so their evidence of a customary interpretation of the Refugee Convention is limited.

Further, in another set of cases, even where cessation of refugee status occurred preventing protection under the Refugee Convention, some cases held that general conditions of instability still could be grounds for granting subsidiary protection. Such a decision by a state’s court may in fact evidence that the state holds the belief that where the Refugee Convention strictly fails to provide for protection, the state is nonetheless still obliged to grant protection. This consideration will be addressed in more detail in a subsequent section on subsidiary protection, infra sec. III.E.2.

The Refugee Convention does contemplate that although the conditions of persecution are no longer continuing, an individual might still qualify as a refugee if there are compelling reasons arising out of previous persecution for refusing to avail oneself of the protection of the country of nationality; this is sometimes referred to as “exemption from cessation.” The Refugee Convention limits the application of this exception only to individuals who qualify as refu-

18. See e.g., Bundesverfassungsgericht [BVerfG] [Fed. Const. Ct.] July 23, 2004, Case 2 BvR 1056/04 (Ger.) (holding that refugee status may cease and recognition may be withdrawn, as long as the individual received notice of the decision); OVG Schleswig-Holstein, June 16, 2004, Case 2 LB 54/03 (Ger.) (reaching the opposite conclusion in the case of Afghanistan, finding it was sufficiently stable); VG Berlin, Feb. 2, 2004, Case VG 33 X 302.96 (Ger.) (same); VG Dresden, Mar. 16, 2004, Case Case A 7 K 31035/03 (Ger.) (same); VG Gelsenkirchen, Nov. 11, 2004, Case 5a K 8121/95.A (Ger.) (same); VG Neustadt a.d.W, Apr. 26, 2004, Case 5 K 1900/03.NW (Ger.) (same). See also VG Stuttgart, Jan. 7, 2003, Case 5 K 11226/00 (Ger.) (same).

19. See e.g., VG Braunschweig, Nov. 12, 2004, Case 6 A 58/04 (Ger.) (holding that cessation did not apply in particularized cases of ethnic Albanians from Kosovo where, e.g., there were lingering effects); VG Göttingen, Apr. 27, 2004, Case 3 A 519/03 (Ger.) (same); VG Saarland, Nov. 24, 2004, Case 10 K 442/02.A (Ger.) (same).

20. See e.g., VG Düsseldorf, July 15, 2004, Case 6 K 4833/03.A (Ger.) (holding that Kabul was sufficiently stable to present a viable internal flight alternative).

21. See e.g., VG Gelsenkirchen, Nov. 11, 2004, Case 5a K 8121/95.A (Ger.) (holding that poor living conditions justified granting subsidiary protection); VG Wiesbaden, Mar. 30, 2004, Case 7 E 572/04. A(V) (Ger.) (same).

gees under Article 1A(1), not 1A(2), i.e., only to “statutory” refugees whose status was based on conventions prior to the Refugee Convention, not “convention” refugees whose status is based on the particular definition of “refugee” in the Convention. Notwithstanding the explicit terms of the Refugee Convention, there is practice of states extending this “exemption from cessation” protection to convention refugees, even though the UNHCR clearly phrases this interpretation as not legally required by the Refugee Convention. However, given the extent of state practice there may now be a customary norm requiring its application. Again, there is a strongly purposive application of the terms of the Refugee Convention in state practice.

Also not qualifying as refugees under the Refugee Convention are persons

23. See e.g., Canada: Immigration Act § 2(3) (Can.); Jiminez v. Can., F.C. No. IMM-1718-98 (Can.); Finland: Aliens Act, § 36, Feb. 22, 1991 (Fin.) (however, it is unclear whether the standard is the same: only in cases where the person is “evidently no longer stands in need of protection”); Ireland: Refugee Act 1996, No. 17/1996 § 21(2) (Ir.) (“compelling reasons”); Netherlands: Aliens Act 2000 § 27(1) (Neth.) (“pressing reasons of a humanitarian nature”); New Zealand: Re R.S. (135/92) (N.Z.) (“it can no longer be confidently said that the ‘compelling reasons’ exception is confined only to refugees under Article 1A(1)’); Switzerland: Case No. 16, “F.M.,” Rwanda at 139 [ARK] Mar. 23, 1998 (Switz.); the United Kingdom: In re Qafaliaj, 00/HX/01051, [Immigr. & Naturalization Trib.] (U.K.); and the United States: 8 U.S.C. § 208.13(b)(1)(ii) (exempts from denial of status those who are able to demonstrate “compelling reasons for being unwilling to return to his or her country of nationality or last habitual residence arising out of the severity of the past persecution”); Skalak v. Immigration & Naturalization Serv., 944 F.2d 364 (7th Cir. 1991); Lal v. Immigration & Naturalization Serv., 255 F.3d 998 (9th Cir. 2001); Lopez-Galarza v. Immigration & Naturalization Serv., 99 F.3d 954 (9th Cir. 1996); Matter of Chen, 20 I. & N. Dec. 16 (Bd. Immigr. Appls., 1989). But see France: Loi No. 52-893, Loi relative au droit d’asile (July 25, 1952) (providing for the opposite); Luxembourg: Loi du 20 mai 1953 portant approbation de la Convention relative au statut des réfugiés, signée Genève, le 298 juillet 1951 [Refugee Convention], Mém. A-37, Jun 16, 1953 § 703 (same); Portugal: Law No. 15/98 § 36(h), Mar. 26, 1998 (Port.) (same); the United Kingdom: R v. Spec. Adj’t ex parte Hoxha, [2005] 1 W.L.R. 1063; [2005] 4 All E.R. 580, [2005] U.K.H.L. 19 (H. Lords, Mar. 10, 2005) (holding that there was no evidence of a clear and widespread state practice); R v. Sec’y St. ex parte Adan, [2001] 2 A.C. 477 (H. Lords, Apr. 2, 1998) (Slynn, L.) (“I am satisfied, however, that the Geneva [Refugee] Convention in Article 1A(2), does not confer that status. The first matter to be established under the Article is that the claimant is outside the country of his nationality owing to a well-founded fear of persecution. That well-founded fear must, as I read it, exist at the time his claim is established under the Article is that the claimant...”)

24. Milner, supra note 22, at 95-96.

25. See id. at 96 (“according to the Conclusions of the Lisbon Expert Roundtable, “Application of the ‘compelling reasons’ exemption to general cessation contained in Article 1C(5)-(6) is interpreted to extend beyond the actual words of the provision and is recognized to apply to all Article A1(2) refugees. This reflects a general humanitarian principle that is now well-grounded in State practice.”) (citing Joan Fitzpatrick, Current Issues in Cessation of Protection under Article 1C of the 1951 Refugee Convention and Article 1.4 of the 1969 OAU Convention, Excerpt Paper for the Global Consultations on International Protection, U.N.H.C.R., at ¶ 69, (citing in turn the practice of Germany, Ireland, Slovakia, Ghana, Liberia, Malawi, Zimbabwe, Azerbaijan, Lithuania, Canada and the United States) (Apr. 2001)). Also note that Germany, the United States and Canada may be considered specially interested states, see infra note 65 and accompanying text.
who are receiving protection or assistance from UN offices other than the UNHCR such as United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA), or the United Nations Korean Reconstruction Agency, now defunct.\textsuperscript{26} However, if the individual leaves the protection of that agency, and UNRWA is only operating in some portions of the Middle East, then the individual appears to qualify as a refugee under the Refugee Convention (the Helsinki, Finland Administrative Court went so far as to determine that the individual must be granted asylum because he now qualifies under the Refugee Convention).\textsuperscript{27}

In addition, other persons may qualify under (1) or (2) but do not benefit from the protection of non-refoulement. These provisions have been adopted in municipal legislation on refugee status.\textsuperscript{28} There are two classes of persons contemplated. The first class are those falling in the “exclusion” clauses,\textsuperscript{29} including individuals with respect to whom there are “serious reasons for considering” that they have committed a crime against peace, a war crime, or a crime against humanity,\textsuperscript{30} committed a serious non-political crime outside the country of ref-

\begin{enumerate}
\item \textsuperscript{26} Refugee Convention, art. 1D; UNHCR, Background Note on Exclusion Clauses, supra note 12, at ¶ 8. See also UNHCR, Note on the Applicability of Article 1D of the 1951 Convention relating to the Status of Refugees to Palestinian Refugees (Oct. 2002).
\item \textsuperscript{27} See Case No. 15.9.2004/04/1240/7 (Helsinki Admin. Ct., Fin. 2004).
\item \textsuperscript{28} See e.g., Act No. 480/2002 § 13 (Slovak) (providing specific categories of people who are excluded from the protection of the Refugee Convention, essentially adopting the Refugee Convention text).
\item \textsuperscript{29} The UNHCR has requested that states interpret the exclusion clauses restrictively. See UNHCR HANDBOOK, ¶ 116; UNHCR, Background Note on Exclusion Clauses, supra note 12, at ¶ 4; UNHCR, Note on Expulsion of Refugees, U.N.H.C.R. Doc. EC/SCP/3, ¶ 4 (1977). See also X & Y v. Refugee Status Appeals, CIV 2006-404-4213, Auth., High Court, Auckland (N.Z.), 23 Apr. 2007, at ¶ 64 (agreeing that the clauses must be interpreted restrictively).
uge prior to admission as a refugee, or have “been guilty of acts contrary to the purposes and principles of the United Nations.” However, there may be an exception to these provisions for cases involving child soldiers, even though the Refugee Convention does not specifically address this exception. Persons falling within the exclusion clauses do not appear to acquire the status of refugee under the Refugee Convention, as opposed to individuals qualifying as refugees who might later lose that status.

A second class of persons who are refused protection includes individuals lawfully present who: (1) pose a compelling threat to national security or public order, (2) present a danger to the security of the country of refuge, or (3) have been convicted by a final judgment of a particularly serious crime and constitute a danger to the community of the country of refuge. This second class applies to an individual who at some time qualified as a refugee, but who subse-


33. See Case No. AWB 03/26654, Judgement (Dist. Ct. Arnhem, Neth., Oct. 18, 2004) (regarding a former child soldier of UNITA); UNHCR, Background Note on Exclusion Clauses, supra note 12, at ¶¶ 91-93. See also, id. at n.92 (“If the age of criminal responsibility is higher in the country of origin, this should also be taken into account (in the child’s favour).”)

34. UNHCR, Background Note on Exclusion Clauses, supra note 12 at ¶ 10.

35. Refugee Convention, art. 32(1); Dec. No. Nr. III-27-20/2005, (Vilniaus apygardos administracinio teismo sprendimas [Vilnius Dist. Admin. Ct.], Lith., July 15, 2005) (balancing threat to national security and public order against family relations in Lithuania and holding that the need to expel was not necessary in a democratic society).

36. Refugee Convention, art. 33(2). See e.g., Nationality, Immigration and Asylum Act 2002 (Specification of Particularly Serious Crimes) Order 2004, (Aug. 2004) (U.K.) (listing offences defined as serious within the context of Article 33(2) of the Refugee Convention, including offences such as “criminal damage” possibly encompassing graffiti or shoplifting).

37. Refugee Convention, art. 33(2).
Worster: The Evolving Definition of the Refugee in Contemporary Internatio

104 BERKELEY JOURNAL OF INTERNATIONAL LAW [Vol. 30:1

quently loses that status.\textsuperscript{38} The exclusion clause does not appear to be widely used, even if it is potentially applicable.\textsuperscript{39}

The infrequent application of the exclusion clauses, coupled with the child soldiers exception, exemplify the Convention’s dynamic and human-oriented application. At least one court has held that Article 33(2) of the Refugee Convention has been amended by the more absolute prohibitions on \textit{refoulement} provided in the Convention Against Torture.\textsuperscript{40} In \textit{A.G. v Zaoui}, N.Z.S.C. 38, the New Zealand Supreme Court initially noted that the Convention Against Torture was a successive treaty relating to the same subject matter as the Refugee Convention, and not an amendment to the Convention.\textsuperscript{41} However, the court also held that “[t]he prohibition on refoulement to torture has the status of a peremptory norm or \textit{jus cogens} with the consequence that article 33.2 [of the Refugee Convention] would now be void to the extent that it allows for [refoulement in such circumstances].”\textsuperscript{42} Essentially, the court held that the prohibition of torture was \textit{jus cogens},\textsuperscript{43} so \textit{refoulement} to a situation of torture was prohibited and the Refugee Convention may not contain a provision that would permit \textit{refoulement} to a situation of torture.\textsuperscript{44} Although not technically amended, the terms of the Convention have thus been stricken. If the Refugee Convention has not been amended by the practices cited above, then it is worth considering whether the prohibition of \textit{refoulement} in a situation of torture is establishing a customary international legal norm, perhaps even a \textit{jus cogens} norm, alongside the conventional one. In any event, it would appear that the means of interpreting the obligations of the Refugee Convention have been established as overwhelmingly purposive.

Not only do the various examples above imply flexible application of the Convention, the Council of Europe also has expressly urged flexible application. The Council issued Recommendation 773 in an attempt to address the plight of certain individuals denied recognition as refugees. In this recommendation it is not entirely clear whether the problem is that the individuals do not qualify un-

\textsuperscript{38} UNHCR, \textit{Background Note on Exclusion Clauses}, supra note 12 at ¶ 10.


\textsuperscript{41} See \textit{A.G. v Zaoui}, N.Z.S.C. 38 at ¶ 50.

\textsuperscript{42} See \textit{id. at ¶ 51.}

\textsuperscript{43} See also, e.g., \textit{Pros. v. Furundzija, Case No. IT-95-17/1-T, Judgment, ¶¶ 144, 147, 153-4 (Int’l Crim. Trib. Former Yugoslavia Dec. 10, 1998).}

under the Refugee Convention at all, or whether the states of the Council of Europe are not properly fulfilling their international legal obligation to recognize these individuals as de jure refugees under municipal law. The Council recommended that states “apply liberally the definition of ‘refugee’ in the Convention . . . as amended by the Protocol . . ..” A similar policy of dynamic interpretation of the Convention was evidenced above for particular provisions of the Convention, but here it is argued that the entire Convention definition of refugee should be liberally applied.

In sum, the conventional definition of refugee is a complex one. It requires that the person, his background, and his situation satisfy both inclusive and exclusive requirements. However, the conventional definition has shown a remarkable flexibility of liberal interpretation, either based on an evolving interpretation of the Convention or perhaps a supplementary understanding from customary international law. In fact, it appears that it is required for the terms of the Refugee Convention to be interpreted primarily with an eye to the Convention’s object and purpose.

III.
EVOLUTION OF A DEFINITION UNDER CUSTOMARY INTERNATIONAL LAW

As mentioned above, the Refugee Convention has not been amended either explicitly or through practice to provide for a revised definition of refugee; however, customarily it is interpreted in an expansive fashion, relying heavily on its object and purpose. In fact, in some instances cited above, the qualification as a refugee may have been supplemented beyond the express terms of the Convention.

It has been argued that the definition of refugee does not exist under customary international law but only under treaty law. Most scholars of international refugee law have concluded as much. In particular, as far as the European Union is concerned, Kay Hailbronner has concluded, “[T]he assumption of an international legal obligation to grant protection to victims of war, civil war and general violence must still be considered as ‘wishful legal thinking.’” Similarly, the American Society of International Law has concluded that there is no

47. See supra notes 17, 23, and 33.
customary international law obliging states to provide protection to individuals who fall outside the strict terms of the Refugee Convention. Even as active an advocate as Guy Goodwin-Gill has stated that:

[P]ractice reveals a significant level of general agreement not to return to danger those fleeing severe internal upheavals or armed conflict in their own countries . . . . nearly four decades of practice contain ample recognition of a humanitarian response to refugees falling outside the 1951 Convention. Whether practice has been sufficiently consistent over time and accompanied by the *opinio juris* essential to the emergence of a customary rule of refuge, is possibly less certain, even at the regional level.\(^51\)

This Article will question the validity of these conclusions.

The conference that adopted the Refugee Convention immediately adopted a recommendation and attached it to the Final Act, urging states to extend refugee benefits to individuals not qualifying under the narrow terms of the Refugee Convention:

The Conference expresses the hope that the Convention relating to the Status of Refugees will have value as an example exceeding its contractual scope and that all nations will be guided by it in granting so far as possible to persons in their territory as refugees and who would not be covered by the terms of the Convention, the treatment for which it provides.\(^52\)

This statement could be interpreted to acknowledge, or possibly even express *opinio juris*, that a complementary definition would develop under customary international law.

Many authors have attempted to argue that just such a definition under customary international law has arisen. Some have argued that the prevailing restrictive reading of the term “refugee” in the Convention is incorrect, disregards usage of the term prior to the Convention and is not supported by the *travaux préparatoires*.\(^53\) Some have even argued that the Refugee Convention is merely

---


51. GUY S. GOODWIN-GILL, THE REFUGEE IN INTERNATIONAL LAW 171 (1996). However, Goodwin-Gill has subsequently relaxed this perspective to be much more open to the existence of a definition of refugee under customary international law.


one in a collection of human rights instruments that must be read as a whole so that the protections described by the Refugee Convention apply to any person who enjoys some form of non-refoulement from any human rights instrument. Thus, non-refoulement is a general principle and the Refugee Convention is merely one kind of situation in which non-refoulement arises. For example, María-Teresa Gil-Bazo argues that “in addition to refugees within the meaning of the Geneva Convention, there are other categories of individuals that have a right to protection under international law and accordingly, they are ‘refugees’ in a broader sense.” However, many of these authors have not cited extensive practice and opinio juris to support their argument. This Article will attempt to identify practice and opinio juris on point.

A. State Practice Expanding the Definition

As the international relations of states evolves, so too does the law, at least customary international law. It has been observed that increasingly “refugee” flows have been more likely due to “civil wars, ethnic and communal conflicts and generalized violence, or natural disasters or famine—usually in combinations—than individually targeted persecution by an oppressive regime.” As states have shifted their behavior to respond to these crises, we must consider whether they have shifted their understanding of the definition of refugee under customary international law.

It is accepted in the international legal system that binding international law can arise through custom. Discussion of sources commonly cites the Statute of extend to a person ‘unable to obtain from [his or her] country [of origin] permission to return’) (citing Ad Hoc Committee on Statelessness and Related Problems, France: Proposal for a Draft Convention Preamble, U.N. Doc. E/AC.32/L.3 (17 Jan. 1950)). Also note that France may be considered a specially interested state. See infra note 65 and accompanying text.


55. See María-Teresa Gil-Bazo, La protección internacional del derecho del refugiado a recibir asilo en el Derecho internacional de los derechos humanos, in F.M. MARÍNO MENÉNDEZ, ED., DERECHO DE EXTRANJERIA, ASILLO Y REFUGIO 691-2 (2d ed., 2003). See also Spijkerboer, supra note 53; Gil-Bazo, Refugee status, supra note 53, at 10 (citing Goodwin-Gill, Asylum: The Law and Politics of Change, 7 Int’l J. Ref. 7 (1995)) (“The refugee in this broader sense includes not only those who have a well founded fear of persecution, but also those who have a substantial risk to be subjected to torture or to a serious harm if they are returned to their country of origin, for reasons that include war, violence, conflict and massive violations of human rights.”). However, among those who argue that the term “refugee” should be read liberally, some concede that there must still be some element of persecution in order to qualify, although collective persecution would suffice.


57. See generally Statute of the International Court of Justice, art. 38, June 16, 1945; Statute of the Permanent Court of Justice, art. 38, Dec. 16, 1920; A. D’AMATO, THE CONCEPT OF CUSTOM IN INTERNATIONAL LAW (1971); HUGH THRILWAY, INTERNATIONAL CUSTOMARY LAW AND CODIFICATION (1972); Int’l L. Assoc., Statement of Principles Applicable to the Formation of Gen-
Worster: The Evolving Definition of the Refugee in Contemporary Internatio
108 BERKELEY JOURNAL OF INTERNATIONAL LAW [Vol. 30:1

the International Court of Justice for proof that “evidence of a general practice
accepted as law” is law.58 Building on this definition of customary international
law, courts have determined that this source of law has two elements: state prac-
tice and opinio juris sive necessitatis.59 State practice is usually defined as a
widespread and consistent practice followed by states.60 Opinio juris is usually
defined as a subjective belief on the part of the state engaging in the practice that
the practice is required, not merely optional.61

As for the first element of state practice, there is no set number of states
that must engage in the practice before it becomes law. It is accepted that it does
not need to be all states, just many of them.62 However, in the North Sea Conti-
nental Shelf cases, the International Court of Justice stated that the practice of
“specially affected States” is the most significant practice.63 Which states will
be specially affected will vary on a case-by-case basis, depending on the nature

58 See Statute of the International Court of Justice, supra note 57, art. 38(1)(b).
59 See e.g., Military & Paramilitary Activities In & Against Nicaragua, Merits (Nicar. v.
43 (Feb. 20); Case Concerning Right of Passage Over Indian Territory, Merits (Port. v. Ind.) 1960
I.C.J. Reps. 6 (Apr. 12); Fisheries Case (U.K. v. Norw.), 1951 I.C.J. 116 (Dec. 18); Asylum Case
(Colom./Peru), 1950 I.C.J. 266 (Nov. 20); Case of the S.S. “Lotus” (Fr. v. Turk.), 1927 P.C.I.J. (ser.
A) No. 9 (Sept. 7).
60 See Asylum Case, 1950 I.C.J. 276-7 (“in accordance with a constant and uniform usage
practised by the States in question”) (holding that state practices was lacking in the consistency and
certainty required to constitute “constant and uniform usage”); IAN BROWNLIE, PRINCIPLES OF
10, 28 (“only if such abstention were based on their [the states] being conscious of a duty to abstain
would it be possible to speak of an international custom”); MALCOLM N. SHAW, INTERNATIONAL
LAW 75 (6th ed. 2008).
62 See Fisheries Case, 1951 I.C.J. 131, 138; CHARLES DE VISSCHER, THEORY AND REALITY
IN PUBLIC INTERNATIONAL LAW 149 (P.E. Corbett trans., Princeton Univ. rev. ed. 1968); HERSCH
LAUTERPACHT, THE DEVELOPMENT OF INTERNATIONAL LAW BY THE INTERNATIONAL COURT 368
(1958). See also, generally JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, CUSTOMARY
INTERNATIONAL HUMANITARIAN LAW (2005) (surveying a selection of representative states for each
point of law, which has been widely accepted as correctly stating the law on the matter) [hereinafter
HENCKAERTS & DOSWALD-BECK, CUSTOMARY IHL].
63 N. Sea Cont. Shelf Cases, 1969 I.C.J. 43, at ¶ 74; SHAW, supra note 61 at 80 (citing situa-
tions where the practice of only one or two states could be potentially determinative, such as the
practice of the United Kingdom regarding the law of the sea; the practice of the United States and
USSR regarding space law). We might consider that if regional custom could be found, the usage in
such a case would also presumably need to be widespread and consistent albeit only within the re-
gion. See e.g., N. Sea Cont. Shelf Cases, 1969 I.C.J. 43; Asylum Case, 1950 I.C.J. 276-7 (requiring
“constant and uniform usage”); HENCKAERTS & DOSWALD-BECK, CUSTOMARY IHL, supra note 62,
at xlv; M. MENDELSON, ET. AL., FINAL REPORT OF THE COMMITTEE: STATEMENT OF PRINCIPLES
APPLICABLE TO THE FORMATION OF GEN. CUSTOMARY INT’L L. Principle 14, Commentary (c)
(2000).
of the practice being examined. On this basis, it is worth noting which states are specially affected by refugee flows and note their practice and *opinio juris* in particular. This Article cannot hope to survey all states in the world to any satisfying degree of depth and certainty, but it can assess many of them, particularly those whose practice is representative of global practice. One of the first steps, then, is to identify those states that might be specially interested in refugee law.

**B. The Role of Specially Interested States**

There are more than nine million individuals that the United Nations High Commissioner for Refugees has identified as “refugees” deserving protection. Of that number, the states in which the largest numbers of individuals have sought refuge are, beginning with the largest:

Table 1

<table>
<thead>
<tr>
<th>State</th>
<th>Refugees or Persons in Refugee-like Situations</th>
<th>Percentage of Total Global Population of Refugees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syria</td>
<td>1,503,769</td>
<td>15</td>
</tr>
<tr>
<td>Iran</td>
<td>963,546</td>
<td>10</td>
</tr>
<tr>
<td>Pakistan</td>
<td>887,273</td>
<td>9</td>
</tr>
<tr>
<td>Germany</td>
<td>578,879</td>
<td>6</td>
</tr>
<tr>
<td>Jordan</td>
<td>500,281</td>
<td>6</td>
</tr>
<tr>
<td>Tanzania</td>
<td>435,630</td>
<td>5</td>
</tr>
<tr>
<td>China</td>
<td>301,078</td>
<td>3</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>299,718</td>
<td>3</td>
</tr>
<tr>
<td>Chad</td>
<td>294,017</td>
<td>3</td>
</tr>
<tr>
<td>United States</td>
<td>281,219</td>
<td>3</td>
</tr>
<tr>
<td>Kenya</td>
<td>265,729</td>
<td>3</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>240,742</td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>228,959</td>
<td></td>
</tr>
<tr>
<td>Sudan</td>
<td>222,722</td>
<td></td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>177,390</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>175,741</td>
<td></td>
</tr>
</tbody>
</table>

64. The International Committee of the Red Cross (“ICRC”) study gives examples of “specially affected” states in certain situations. See HENCKAERTS & DOSWALD-BECK, CUSTOMARY IHL, supra note 62, at xliv.

65. These numbers are provided on the UNHCR website for statistics dated at the end of 2007. 2007 UNHCR Statistical Yearbook, http://www.unhcr.org/statistics/49a2c7f2.html. The particular chart containing this data is http://www.unhcr.org/static/statistical_yearbook/2007/annextables.zip. The next ten states, in the order of the number of refugees seeking refuge on their territories are: Yemen (117,363 persons); Zambia (112,931); Serbia (97,995); Egypt (97,556); Algeria (94,137); Netherlands (86,587); Ethiopia (85,183); Sweden (75,078); Cameroon (60,137); and Rwanda (53,577). See id.
Perhaps states that experience significant numbers of refugees are “special-
ly affected” in the North Sea Continental Shelf sense and thus more crucial to
and representative in establishing the widespread practice and opinio juris ne-
necessary. Those states actually deal with more cases and experience the effects of
their policies more directly (or, conversely, consider that evidence of state prac-
tice and opinio juris is not necessarily widespread unless it is undertaken by
states that are the major recipients of refugees). Stated differently, perhaps the
way that the majority of individuals in the world seeking refuge are treated, al-
beit by less than a majority of states, is more relevant in establishing custom
than by the way in which they define “refugee.”

It is not entirely clear under international law whether representative states
should also be geographically and culturally diverse in order to establish the ex-
istence of generally customary international law. Insofar as diversity may be
necessary, the above list of states is already fairly diverse with perhaps South
America being the one region that is not represented. Therefore, South Ameri-
can regional practice will also be addressed in the sections below to accommo-
date for its omission from the list of specially interested states above and ensure
geographical diversity in the analysis.

C. International Agreements Defining Refugee Status

1. The Organization of African Unity Convention on Refugees

The 1969, the Organization of African Unity (OAU) Convention on the
Specific Aspects of Refugee Problems in Africa expanded the definition of refu-
gee in the Refugee Convention to include those fleeing “external aggression, oc-
cupation, foreign domination or events seriously disturbing public order.”66 Due
to a lack of documentation on the drafting history of the OAU Convention, there
has been considerable debate about the intention of the drafters, and speculation
has been unhelpful.67 However, note that this instrument was signed by some of
the largest recipients of refugees in the world, specifically: Tanzania, Chad,
Kenya, Uganda, Sudan, Democratic Republic of Congo, Zambia, Egypt, Alge-

ria, Ethiopia, Cameroon, and Rwanda.68

In addition to these major recipients of refugees, the OAU Convention was also signed by Burundi, the Central African Republic, Congo, Gabon, Gambia, Ghana, Guinea, Côte d’Ivoire, Mali, Nigeria, Rwanda, and Sierra Leone, which are all in the upper half of states receiving the most numbers of refugees. States such as South Africa,69 Tanzania,70 and Uganda71 for example, have additionally adopted the OAU definition into municipal law. The second of the states incorporating the Convention criteria, Tanzania, is the single highest recipient of refugees in Africa and the sixth highest recipient in the world.

In addition to these refugee numbers, which magnify the impact of the state practice of those states on the formation of customary international law, also note that Mexico has adopted into state law the definition established by the OAU convention.72 Clearly, Mexico is under no obligation to adopt this definition from another region, but this act demonstrates a growing acceptance outside the region of the norms established by the region. In sum, the OAU Convention, although only binding states in the region under treaty law, has significantly

68. See 2007 UNHCR Statistical Yearbook, supra note 65.
69. See Ruma Mandal, Protection Mechanisms Outside of the 1951 Convention (“Complementary Protection”), ¶ 238, U.N. Doc. PPLA/2005/02 (June 2005) (“In defining refugee status, section 3 of the 1998 Refugees Act incorporates the refugee definitions in Article 1A(2) of the 1951 Convention and Article I(2) of the 1969 OAU Convention (in sub-sections 3(a) and 3(b) respectively).”), http://www.unhcr.org/protect.
70. See id. at ¶ 242 (“Refugee status is defined in section 4 of the 1998 Refugees Act. Section 4(1)(a) incorporates the language of Article 1A(2) of the 1951 [Refugee] Convention while section 4(1)(b) adopts the text of Article I(2) of the 1969 OAU Convention. Section 4(4) of the Act incorporates the exclusion clause in Article 1F of the 1951 Convention, though some of these grounds are also included in the section 4(3) provision on cessation.”). Also note that Tanzania is the sixth highest recipient of refugees and persons in refugee-like situations. Thus, it may be considered specially interested. See supra note 69.
contributed to the formation of a general rule of customary international law due to its highly representative nature of establishing the norms of treatment for a high percentage of the individuals seeking refuge in the world.


The Asian-African Legal Consultative Organization agreed in June 2001 on a set of principles concerning the treatment of refugees, known as the “Bangkok Principles.” Although they are not binding themselves, they could nonetheless serve as a source of opinio juris. The Bangkok Principles define refugee essentially the same way as the Refugee Convention, but they also cover:

(Every person, who, owing to external aggression, occupation, foreign domination or events seriously disturbing public order in either part or the whole of his country of origin or nationality, is compelled to leave his place of habitual residence in order to seek refuge in another place outside his country of origin or nationality).

The Bangkok Principles contain other provisions similar to those in the Refugee Convention, such as dual nationality, cessation, and exclusion. It specifically accepts the concept of refugee sur place. It also provides that a “refugee” may not be expelled “to a State or Country where his life or liberty would be threatened for reasons of race, colour, nationality, ethnic origin, religion, political opinion, or membership of a particular social group.”

A large number of specially interested states has adhered to the Bangkok Principles. This list includes the following states with large influxes of refugees: Syria, Iran, Pakistan, Jordan, Tanzania (also a party to the OAU Convention), PR China, Kenya, Saudi Arabia, Uganda, Sudan, India, Nepal, Thailand, Yemen, Zambia, Egypt, and Cameroon.

75. Bangkok Principles, art. I(1).
76. Id. at art. I(2).
77. Id. at art. I(5).
78. Id. at arts. I(6) – (7).
79. Id. at art. I(3).
80. Id. at art. V(3).
81. See supra note 65 and accompanying text. Also note: Tanzania, Kenya, Uganda, Sudan, Egypt, and Cameroon are all members to the OAU Convention. It is important to note that India submitted a communication when adopting the Bangkok Principles where it opposed the expanded definition. See id. at art. I, note 5. But see Sec. III.E. infra, for a discussion of other aspects of India’s actual practice that may be inconsistent with this apparent statement of opinio juris. It may be that India was opposed to expanding the Convention definition in the instrument purely because of the potential for weakening refugee integration into their host states, but is not opposed to engaging in practices supportive of an expanded definition.
The above states are among the top twenty-five state-recipients of refugees. Additional signatories to the Principles that fall within the top half of states receiving refugees are Bangladesh, Gambia, Ghana, Lebanon, Nigeria, Senegal, Sierra Leone, and Turkey. Of these states, Gambia, Ghana, and Nigeria are also parties to the OAU Convention. The Bangkok Principles have thus provided a very strong statement of *opinio juris* by the major recipients of world refugee flows.

3. The Cartagena Declaration

The 1984 Cartagena Declaration, regarding forced migrants in Central and South America, expresses the same principles as the foregoing documents, even going so far as to explicitly refer to Article I(2) of the OAU Convention as inspiration for its definition of “refugee,” although the two texts do differ. Of the Central and South American states, Costa Rica participated in drafting the Cartagena Declaration but has only received approximately 12,000 refugees, placing it almost in the top fifty states receiving refugees. Ecuador is in the top fifty recipient states, but did not participate in the conference. Brazil and Argentina did not participate, but have received approximately 3,000 refugees each.

Although Brazil and Ecuador neither attended the drafting conference nor signed the Cartagena Declaration, the Declaration’s principles have been adopted into municipal law in both countries. The Declaration, although not legally binding in itself, has been endorsed by the Organization of American States, the UNHCR Executive Committee and by states party to the universal refugee treaties. Further, the Declaration has been cited in turn by the Brasilia Declaration on the Protection of Refugees and Stateless Persons in the Americas, which was signed by Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Mexico, Nicaragua, Pana-

---

84. See Refugee Law No. 9474/97, § I(iii) (Braz.) (persons seeking refuge because of “serious and generalised violations of human rights”) (discussed in Mandal, *supra* note 69 (observing that “[i]n practice, it seems that subsection (iii) is also considered to apply to situations of armed conflict and generalised violence.”)); Pres. Decree 3301/92, arts. 1-2 (Ecuad.) (discussed in Mandal, *supra* note 69 (observing that “[s]ince the beginning of 2003, only 36 out of 5772 individuals recognised as refugees were recognised on the basis of Article 2 of the Decree.”))
ma, Paraguay, Peru, Uruguay, and Venezuela. The Brasilia Declaration “highlighted” the expansive regional definition of refugee, suggesting that the participants at the Brasilia Declaration Conference understood that an expansive legal definition already existed in the region, due to the formally non-binding Cartagena Declaration and supporting instruments. Based on these participants and other factors, and notwithstanding the declaratory nature of the document, the Cartagena Declaration does further crystallize customary international law.

4. The Mercosur Rio de Janeiro Declaration

The Mercado Común del Sur (Mercosur), along with Bolivia and Chile, has also adopted the Rio de Janeiro Declaration on the Institution of Refuge. This declaration provides that “international protection should be given to individuals persecuted for reasons of race, nationality, religion, membership of a particular social group, political opinion or victims of serious and generalized violation of human rights.” Specifically, the states parties proclaimed that they “will study the possibility of including in the refugee definition the protection of victims of serious and generalized human rights violations,” and that they:

[W]ill not apply refoulement measures to a refugee who has been recognised in another Contracting or associate State, to a country where his life, freedom or physical integrity are threatened by reasons of race, nationality, membership of a particular social group, political opinion or serious and generalised violation of human rights, according to the international norms governing this issue.

The practices of Latin American countries is unlikely to heavily influence the formation of customary international law because these countries do not receive the high numbers of refugees that countries in other regions do. However, the Declaration remains important as it expresses an opinio juris. Further, the consideration of geographically diverse practice may be necessary to support a finding of the existence of customary international law. Thus, Latin American practice should be considered regardless of whether states in the region are spe-

86. See UNHCR, Brasilia Declaration on the Protection of Refugees and Stateless Persons in the Americas, Brasilia, Nov. 11, 2010, http://www.unhcr.org/4cdd3fac6.html (Canada and the United States also attended the conference as observers) [hereinafter UNHCR, Brasilia Declaration].
87. Highlighting the contribution of the Americas to strengthen the protection of victims of forced displacement and stateless persons through the adoption of multilateral treaties on asylum, statelessness and human rights. Id. at 1. However, the fact that the states would resolve to promote accession to the instruments does not necessarily imply that they believe that the definition of refugee in those other instruments was not already binding under customary international law. Given the other statements suggesting that an expansive legal definition already applied, at least in the region, one cannot conclude otherwise.
89. Id.
90. Id. at Proclam. 3.
91. Id. at Proclam. 4.
cially interested based on refugee flows.

5. Conclusion on the Role of International Agreements by Specially Interested States

All of the foregoing conventions are expressly regional law or soft law, not universal, conventional international law. However, each expresses the practice and/or opinio juris of some of the most important refugee-receiving states in regards to dealing with refugee flows. The Declaration of States Parties to the 1951 Convention, and/or its 1967 Protocol relating to the Status of Refugees in Geneva in 2001, affirmed “the importance of other human rights and regional refugee protection instruments.”92 It might be that obligations undertaken by states that are in a position to suffer the most burden will signal a stronger existence of opinio juris than states that will not suffer as much from assuming such obligations.

This Section has only examined international agreements, but aside from participation in these agreements, the practice of specially interested states is important as well. Now that specially interested states have been identified, this Article will highlight those states where they appear in subsequent sections of this analysis. Based on the practice and opinio juris evidenced in these documents further defining “refugee” in their regions, and the fact that the relevant states are also specially interested ones in regards to refugee flows, the instruments above potentially show that customary international law has supplemented and broadened the definition of “refugee.” Specifically, it appears that the general customary international law definition of refugee includes persons fleeing serious disturbances of public order. Bearing in mind this tentative conclusion, the next issue is the influence of subsidiary protection.

D. The Influence of “Subsidiary Protection” in International Agreements

Whereas the preceding section examined the customary international legal definition of “refugee”, this Section considers whether subsidiary protection, or the protection available for persons under alternative international legal obligations who clearly fall outside the conventional definition of “refugee” governing in the relevant state, has expanded the customary international legal definition of refugee.

It is admitted that even when the Refugee Convention is applied dynamically, many individuals do not qualify under its definition of “refugee” because of the absence of the element of persecution (or no qualifying form of persecution). However, these individuals may have other justifiable reasons for refusing to avail themselves of the protection of their state of nationality. This group of

persons is often referred to “de facto” refugees: their refuge needs are seen as legitimate but they do not qualify under the Convention. Their needs are often addressed through other agreements or legislation granting them a “subsidiary” status. The question for this Section is whether the extension of subsidiary protection status to these individuals has created a wider notion under customary international law of “refugees” that includes those who are deserving of the non-refoulement and other protections substantively the same as those afforded by the Refugee Convention.

The Conference that negotiated the Refugee Convention specifically encouraged states to grant subsidiary protection to individuals not qualifying under the Refugee Convention.93 On this basis, the intent of the Convention’s drafters can be interpreted to express a desire to cover this expanded group of persons.94 When the European Union was considering the minimum standards directive, the EU Presidency stated that the exclusion provision from subsidiary protection should be modeled on Article 1(F) of the Refugee Convention, but there was no legal obligation to follow the Convention’s terms for subsidiary protection.95 In addition, state courts have held that the definition of refugee remains that given in the Refugee Convention and has not been supplemented by humanitarian assistance to persons in need.96 However, consider whether the practice of subsidiary protection, as a separate institution itself, has expanded the definition of refugee under customary international law.

Certain treaties call for subsidiary protection. Most notably, the International Covenant on Civil and Political Rights (ICCPR)97 and the 1984 Conven-

---

93. See Sohn & Buergenthal, eds., supra note 48, at 102 (“The 1951 Geneva Conference which adopted that [Refugee] Convention expressed at the same time the hope that “all nations will be guided by it in granting so far as possible to persons in their territory as refugees and who would not be covered by the terms of the [Refugee] Convention, the treatment for which it provides.””).
95. See Presidency Note to SCIFA, EU Doc. 13623/02 Asile 59 (30 Oct. 2002). But see Gil-Bazo, Refugee status, supra note 53 (observing that Sweden disagreed with the interpretation and sent its own competing proposal to SCIFA). Also note that Sweden might be a specially interested state, although it is less affected than some other states under consideration.
Evolution of the Refugee

Worster, 2012

Evolution of the concept of the refugee includes the evolution of international law and human rights protections. The Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT), which prohibits refoulement of a person “where there are substantial grounds for believing that he would be in danger of being subject to torture (Art. 3).” There are no exceptions for national security to the CAT or ICCPR obligations of non-refoulement to situations of torture. The European Court of Human Rights (ECHR) establishes a similar protection requirement where there is a “real risk” that the person in question will be subject to inhuman or degrading treatment and punishment. The American and African Charters make similar provisions for torture.


99. See Suresh v. Can. (MCI), [2002] S.C.C. 1, ¶ 78 (Can. Sup. Ct., Jan. 11, 2002) (reaching the opposite conclusion: “We do not exclude the possibility that in exceptional circumstances, deportation to face torture might be justified”); but see U.N. Commission against Torture, Conclusions and Recommendations of the Committee Against Torture: Canada, 34th sess., U.N. Doc. CAT/C/CAN/CO/34/CAN, ¶ 4(a) (May 2004) (criticizing the Suresh decision); U.N. Commission on Human Rights, 85th sess., U.N. Doc. CCPR/C/CAN/CO/5 at 15 (Oct.-Nov. 2005) (“The State party should recognize the absolute nature of the prohibition of torture, cruel, inhuman or degrading treatment, which in no circumstances can be derogated from.”). Note that Syria and Pakistan, specially interested states, see supra note 65, are parties to the CAT.


EU law also makes provision for an expanded class of persons in need. The EU Minimum Standards Directive orders EU member states to receive an asylum application if filed.\textsuperscript{102} The EU Council Directive of April 29, 2004, also orders subsidiary protection for any person who cannot return to the country of origin because of serious harm, which consists of: (1) death penalty or execution; (2) torture or inhuman or degrading treatment or punishment of an applicant in the country of origin; or (3) serious and individual threat to a civilian’s life or person by reason of indiscriminate violence in situations of international or internal armed conflict.\textsuperscript{103} Of course, these directives override EU member state law if not already in conformity with it.\textsuperscript{104}

Also important is the Charter of Fundamental Rights of the European Union, which has recently become legally binding.\textsuperscript{105} The Charter provides for certain rights to subsidiary protection insofar as those rights arise from the constitu-
How this will be specifically interpreted remains to be seen; however, an expression of \textit{opinio juris} that those Charter protections should be binding is already evident.

Also noteworthy is the European Social Charter, which provides that migrant workers lawfully residing within the territories of the state parties shall not be expelled unless they endanger national security or offend public interest or morality. Refugees might conceivably fall in this category if they initially came to the state party as a migrant worker and subsequently become a refugee \textit{sur place}.

In sum, additional sources of international law such as the ICCPR, CAT, ECHR and EU law have mandated specific treatment of certain individuals that is broader than the narrow terms of the Refugee Convention. These sources have not expressly sought to supplement the definition of “refugee,” in contrast to the OAU and Cartagena instruments, but have effectively required refugee-like treatment. The fact that they expressly create subsidiary bases for protection rather than amend the Refugee Convention could, however, suggest that their drafters intended to preserve the Convention definition of refugee as a separate institution, thus holding an \textit{opinio juris} that the definition was not to be expanded. At a minimum, these instruments contribute to a supplementary protection regime under customary international law that would provide \textit{non-refoulement} for individuals who would suffer torture, inhuman or degrading treatment, and punishment. Possible additional grounds might be the imposition of the death penalty and indiscriminate violence, although those grounds are less widespread.

\textbf{E. The Influence of Municipal Law}

Moving from international legal obligations to provide subsidiary protection to municipal legal provisions for other forms of subsidiary protection, states tend to, at a minimum, adopt provisions for refuge in their municipal law that track the 1951 Convention definition. However, states have also adopted regional definitions, such as the OAU or Cartagena definitions, into state law. This may evidence an \textit{opinio juris} of the binding nature of the regional definition. For example, the Cartagena Declaration principles have been adopted into municipal law in Brazil and Ecuador, and the OAU definition has been

\begin{itemize}
\item \textit{See id. at pmbl.}
\item \textit{See European Social Charter, Oct. 18, 1961, Council of Eur., art. 19, C.E.T.S. 035.}
\item \textit{See e.g., 8 U.S.C. § 1101(a)(42)(A) (1993); 8 C.F.R. § 1208.13(a); Tesfamichael v. Gonzalez, No. 04-61180 (5th Cir. Oct. 24, 2006) (stating that to qualify for asylum, an alien must be a “refugee”); Migration Act, 1958 (Cth) (Australia).}
\item \textit{See Sohn & Buegenthal, eds., supra note 48, at 103-04.}
\item \textit{See Refugee Law No. 9474/97 § 1(iii) (adopting language from the Cartagena Declaration: “serious and generalised violations of human rights”); Mandal, supra note 69 (“In practice, it seems that subsection (iii) is also considered to apply to situations of armed conflict and generalised vio-}
\end{itemize}
adopted into municipal law in South Africa, Tanzania, Uganda and, strangely, Mexico.

1. Municipal Law and Practice Concerning Refugee Status

This Article examines the municipal law specifically of those states that have been identified as specially interested. Syria, as discussed above, is perhaps the most specially interested. It is signatory to the Bangkok Principles and its constitution provides for protection for political refugees. However, Syria informally permits the UNHCR to perform refugee status determinations on its behalf, resulting in prima facie recognition of refugee status of Iraqi applicants hailing from the central or southern regions of Iraq and the issuance of asylum-seeker documentation to applicants from the Kurdish-controlled northern region. Syria appears to still be in the midst of reforming its refugee policies, but is doing so under the guidance of the Swiss Government and the UNHCR. These actions could suggest state practice and opinio juris that UNHCR definitions of refugee are obligatory.

Iran is also a specially interested state. It is a party to the Refugee Convention, and has adopted the Convention definition into its municipal law. Iran has not, however, clearly complied with the requirements of the Convention based on its lack of transparency in refugee status determination and expulsion

111. See Pres. Decree 3301/92, arts. 1-2.
112. See Mandal, supra note 69, at ¶ 238 (“In defining refugee status, section 3 of the 1998 Refugees Act incorporates the refugee definitions in Article 1A(2) of the 1951 [Refugee] Convention and Article I(2) of the 1969 OAU Convention (in sub-sections 3(a) and 3(b) respectively.”).
113. See id. at ¶ 242 (“Refugee status is defined in section 4 of the 1998 Refugees Act. Section 4(1)(a) incorporates the language of Article 1A(2) of the 1951 [Refugee] Convention while section 4(1)(b) adopts the text of Article I(2) of the 1969 OAU Convention. Section 4(4) of the Act incorporates the exclusion clause in Article 1F of the 1951 [Refugee] Convention, though some of these grounds are also included in the section 4(3) provision on cessation.”) (internal citations omitted).
115. See supra note 71.
116. See supra note 65.
117. See supra note 75, at sec. III.3.C.ii.
120. See id.
121. See supra note 71.
122. See IRAN CONST. (as amended July 28, 1989); Regulations relating to Refugees (1963).
of refugees for violating technical requirements of area of registration. However, it is a signatory of the Bangkok Principles and it permits the UNHCR to conduct refugee status determinations on its territory.

Pakistan, although not a party to the Refugee Convention and having no municipal legislation on point, has delegated the refugee status determination procedure to the UNHCR, at least for Afghans seeking refuge—although individuals with Afghani nationality constitute the single largest group of refugees in Pakistan. Pakistan thus may be considered to have expressed a positive opinio juris regarding the expansive practice of the UNHCR. Additionally, Pakistan is a party to the ICCPR and CAT, as noted above, and signatory to the Bangkok Principles.

Germany, the United Kingdom, and France are all specially interested states. Together with the European Union’s other member states, they are bound to comply with European legislation on refugees, so the policies of the European Union are clearly supported by the practice of specially interested states. Further, European states widely comply with UNHCR recommendations, so in their cases UNHCR recommendations supported by state compliance may be strong evidence of opinio juris as well as state practice.

Jordan is not a party to the Refugee Convention. However, it does have...
Worster: The Evolving Definition of the Refugee in Contemporary Internatio
122 BERKELEY JOURNAL OF INTERNATIONAL LAW [Vol. 30:1

da definition in its Constitution covering political refugees and has signed the Bangkok Principles. In addition, Jordan has signed a Memorandum of Understanding (MOU) providing that refugees recognized as such by the UNHCR will be permitted to enjoy that status in Jordan for six months, during which time the UNHCR will locate countries for resettlement. Even if those persons are not resettled in that time frame, Jordan appears to continue to respect the UNHCR determination and permits the refugees to remain on its territory. This practice suggests recognition and acceptance—possibly constituting state practice and/or opinio juris—that the UNHCR’s definition is correct, and is all the more significant as evidence of customary international law in that Jordan is not a party to the Refugee Convention and yet is a specially interested state.

Tanzania is a party to the Refugee Convention and the OAU Convention, in addition to being a signatory of the Bangkok Principles. Although, its commitment to living up to those standards has been questioned, there does not appear to be any effort by Tanzania to articulate any failure as contributing to the formation of a new norm of customary international law. The practice of Tanzania is to permit the UNHCR to observe its screening procedures for applicants for refugee status, and to intervene in the procedures with legal arguments. When Tanzania proposed repatriating massive numbers of Rwandan refugees, it did so with UNHCR approval. As with the other states noted above, this practice suggests that Tanzania may hold an opinio juris that involvement and standards applied by the UNHCR are obligatory.

China is a party to the Refugee Convention and its municipal law provides asylum for political reasons, although it is unclear whether this practice is actually carried out. China is also a signatory of the Bangkok Principles. Although China does permit UNHCR to conduct status determinations on its be-

133. See supra note 75, at sec. III.3.C.ii.
136. See supra note 65 for analysis of status as a specially interested state; see also supra note 75, at sec. III.3.C.
140. See supra note 65 for analysis of status as a specially interested state; P.R.C. CONST. art. 32 (amended Dec. 4, 1982); See USCRI, World Refugee Survey 2009 – China, (June 17, 2009), http://www.unhcr.org/refworld/docid/4a40d2a3c.html [hereinafter USCRI, China].
141. See supra note 75, at sec. III.3.C.ii.
half, it prohibits UNHCR from doing so near the border with North Korea.\footnote{142}

The practice and \textit{opinio juris} drawn from this policy is more ambiguous, although it could be seen as supportive of the UNHCR definition of refugee generally, similar to the analysis of other states above.

Chad\footnote{143} is a party to the Refugee Convention and the OAU Convention, and its municipal law provides for asylum and protection of political refugees.\footnote{144} Chad has also agreed to an MOU with the UNHCR specifically providing for \textit{non-refoulement}.\footnote{145} In 2007, the UNHCR and Chadian Government jointly proposed a draft law on asylum for Chad, but the draft was not approved, so the country still lacks legislation on point.\footnote{146} Chad is, however, recognizing individuals fleeing from the violence in Darfur and the Central African Republic as \textit{prima facie} refugees, with the condition that they remain in the refugee camps.\footnote{147} Those leaving camps may have individualized refugee status determinations, apparently applying the definition in the Refugee Convention and OAU Convention.\footnote{148} This practice suggests that Chad considers individuals fleeing generalized violence and instability as refugees.

The United States is also a specially interested state.\footnote{149} It is not a party to the Refugee Convention, although it is somewhat incongruously a party to the Refugee Protocol and the CAT. Nonetheless, the United States applies the definition provided in the Refugee Convention in its municipal law, though merging the recognition of status and benefit of \textit{non-refoulement} with the application for asylum.\footnote{150} It exempts terrorists and those providing “material support” to terrorists from eligibility for refuge or asylum, although some exceptions have been introduced. The United States also exempts individuals that commit “aggravated felonies” from \textit{non-refoulement} or a grant of asylum.\footnote{151} Based on these policies, the United States appears to have the policy and \textit{opinio juris} that the Convention definition of refugee applies possibly alternatively through customary international law.

Kenya\footnote{152} is a party to the Refugee Convention and the OAU Convention,

\begin{itemize}
\item \footnote{142}{See USCRI, China, supra note 140.}
\item \footnote{143}{See supra note 65 for analysis of status as a specially interested state.}
\item \footnote{144}{See supra note 75, at sec. III.3.C. See also CHAD CONST. art 46. (amended Mar. 31, 1996).}
\item \footnote{145}{See Memorandum of Understanding, Chad – UNHCR, reported in USCRI, \textit{World Refugee Survey} 2009 – Chad, (June 17, 2009), http://www.unhcr.org/refworld/docid/4a40d2a271.html [hereinafter USCRI, Chad].}
\item \footnote{146}{See id.}
\item \footnote{147}{See id.}
\item \footnote{148}{See id.}
\item \footnote{149}{See supra note 65 for analysis of status as a specially interested state.}
\item \footnote{150}{See Immigration and Nationality Act, 8 U.S.C. § 1158, 208(b) (1986).}
\item \footnote{151}{See id.}
\item \footnote{152}{See supra note 65 for analysis of status as a specially interested state.}
\end{itemize}
as well as having signed the Bangkok Principles. It has also adopted a refugee statute, but it is unclear whether that law has or needs to have implementing regulations in order to be internally binding. However, the UNHCR has received the delegated authority to administer refugee status determinations and operate refugee camps. Based on these practices, Kenya appears to adhere to the practice and opinio juris that the broader notion of refugee in the OAU Convention is the legal definition of the term.

Saudi Arabia is not a party to the Refugee Convention and reserves the right to grant political asylum only where its public interest is served. There does not appear to be any implementing legislation codifying this policy and, in the interim, the policy is very restrictively applied. However, Saudi Arabia has signed the Bangkok Principles and has further agreed to an MOU with the UNHCR, providing UNHCR with the authority to conduct refugee status determinations on its behalf. Although the practice and opinio juris of Saudi Arabia are mixed, there is some acknowledgement that the practice and opinio juris of the UNHCR embody the appropriate standard.

Uganda is a party to the Refugee Convention and the OAU Convention, as well as being a signatory of the Bangkok Principles. It adopted the more liberal definition of refugee provided therein within its municipal legislation.

Sudan is a party to the Refugee Convention and the OAU Convention,

153. See supra note 75, at sec. III.3.C.
155. See USCRI, World Refugee Survey 2009 – Kenya, (June 17, 2009), http://www.unhcr.org/refworld/docid/4a40d2aa76.html [hereinafter USCRI, Kenya] (reporting that there was no Minister for Immigration and Registration of Persons and no regulations in force, although there was a Commissioner for Refugee Affairs who was legally vested with the authority to make refugee status determinations).
156. See id.
157. See supra note 65 for analysis of status as a specially interested state.
159. USCRI, World Refugee Survey 2009 - Saudi Arabia, (June 17, 2009), http://www.unhcr.org/refworld/docid/4a40d2b071.html [hereinafter USCRI, Saudi Arabia] (reporting that recognition of refugee status and/or grants of asylum are only accepted from individuals legally admitted and in possession of a durable residence permit).
162. See supra note 65 for analysis of status as a specially interested state.
163. See supra note 75, at sec. III.3.C.
165. See supra note 65 for analysis of status as a specially interested state.
and signatory to the Bangkok Principles. It adopted the broader definition in the OAU Convention and Bangkok Principles into its municipal law. The express terms of the municipal legislation do not provide for non-refoulement, although it appears to be granted in practice. Although, Sudan permits the UNHCR to monitor its refugee status determinations, it does not allow for intervention. Therefore, the practice and opinio juris are not clearly in favor of one or another legal standard. However, there appears to be a general practice and opinio juris in favor of the OAU definition, which is not as broad as the UNHCR mandate of protection.

In the Democratic Republic of the Congo (DRC), practices around refoulement suggest an opinio juris that the OAU Convention definition of refugee is the applicable standard. The DRC is a party to both the Refugee Convention and the OAU Convention. It has adopted these definitions into municipal law. Reports by NGOs from the country confirm that individuals qualifying under the OAU Convention are not generally being expelled. However, Amnesty International has reported on a massive, forced return of Rwandan refugees. Amnesty’s characterization of the return as a violation of international law does not appear to have been contested, perhaps buttressing the OAU Convention as the applicable standard. However, Amnesty’s argument is focused on the forcible means of the return, not the general legal right to expel.

Canada is a party to the Refugee Convention and CAT, and has implemented those obligations into municipal law. Canada also suspended all deportations, not involving individuals who are a security or criminal threat, to Afghanistan, Burundi, DRC, Haiti, Iraq, Liberia, Rwanda, and Zimbabwe. This suspension suggests an opinio juris that return to situations of instability is impermissible. Further, Canada accepts refugees for resettlement based to some degree on UNHCR classification. Thus, Canada appears to support the practice and opinio juris of UNHCR.

166. See supra note 75, at sec. III.3.C.
169. See id.
170. See supra note 65 for analysis of status as a specially interested state.
172. See Amnesty Int’l, Great Lakes Region, supra note 137, at 3.
173. See supra note 65 for analysis of status as a specially interested state.
176. See id.
India is supportive of UNHCR practices although it is not a party to the Refugee Convention and does not appear to have a law on refugees. However, indirectly, India provides for refugee status based on its constitutional principles. India has served on the UNHCR’s Executive Committee and has signed the Bangkok Principles. Notwithstanding the formal denial of refugee status, India practices a policy of non-refoulement (especially for Tibetans and Sri Lankans, and to some degree also for Bhutanese and Nepalese). India formally denies the UNHCR a binding legal role in refugee status determinations. Nevertheless, it does permit the UNHCR to operate within the country—indicating a level of support for UNHCR practices. Therefore, India could be considered to hold an opinio juris in favor of an expanded definition of refugee.

Nepal similarly is not a party to the Refugee Convention and has no laws on refugees, although it does grant de facto refugee status and is a signatory of the Bangkok Principles. Nepal practices non-refoulement of individuals recognized to be refugees, especially Bhutanese and Tibetans. Nepal also has permitted the UNHCR (albeit somewhat inconsistently) to conduct refugee status determinations, recognized such determinations, and cooperated with the UNHCR operations assisting refugees in resettlement.

Finally, Thailand is also not a party to the Refugee Convention and has no refugee laws. However, similarly to India, Thailand’s practice shows support for the Refugee Convention principles. Thailand has signed the Bangkok Principles and, perhaps more significantly, has permitted the UNHCR to conduct status determinations in the past, although that practice was recently suspended. The UNHCR operations were suspended when Thailand adopted its own de facto status determination procedure, in which it screens individuals in refugee

177. See supra note 65 for analysis of status as a specially interested state.


179. See supra note 75, at sec. III.3.C.ii. But see the discussion of the possible impact of India’s reservation, supra note 81.


181. See supra note 65 for analysis of status as a specially interested state.


183. See supra note 75, at sec. III.3.C.ii.


185. See supra note 75, at sec. III.3.C.ii.
camps and admits a certain qualifying quota. This practice essentially amounts to a national adoption of UNHCR practice, where Thailand is under no conventional legal obligation to do so.

In conclusion, the practice and *opinio juris* of the specially interested states above demonstrates the application of an expanded definition of refugee. This expanded definition is based both on the direct practices of the states, and on the practice of the UNHCR in setting and applying refugee status determination standards on behalf of those states. Moreover, state practices refusing to apply an expanded definition or attempting to curtail the work of the UNHCR have been considered violations of the rules on refugees. Where states act in ways that are successfully characterized as violations of the law, an alternative customary norm does not develop—instead the rule being violated is reaffirmed. The legal characterization of restrictive refugee definition practices as violations of international law indicate that the expanded definition of refugee is well accepted in the international community.

2. Municipal Law and Practice Concerning Subsidiary Protection

Many states provide for some form of subsidiary protection under their national law and this provision could be evidence of practice and *opinio juris* of an expanded definition of refugee. This protection is only sometimes mandated by the states’ international legal obligations. For example, the specially interested states of France, Germany, the United Kingdom, and the United States provide for subsidiary protection either in their international legal obligations or by virtue of *opinio juris*.

186. See USCRI, *World Refugee Survey 2009 – Thailand*, (June 17, 2009), http://www.unhcr.org/refworld/docid/4a40d2b4c.html. Thailand has also expressed the intention to create a formal procedure that would replace the informal one, although it is unclear whether formal means *de jure*. See id.


188. See *FR. 5 TH REP. CONST. art. 53-1* (adopted Oct. 4, 1958) (“the authorities of the Republic shall remain empowered to grant asylum to any foreigner who is persecuted for his action in pursuit of freedom or who seeks the protection of France on other grounds”); *FR. 4 TH REP. CONST.* (adopted Oct. 27, 1946); ECRE, *Complementary/Subsidiary Forms of Protection in the EU Member States: An Overview: France*, http://www.ecre.org/files/survcompro.pdf (noting that the refugee authority and appeals board has discretion to grant Constitutional asylum to persons “fighting for freedom” who do not qualify as Convention refugees); *Loi no. 52/893 relative au droit d’asile* (the “Asylum Law”), art. 2 (II) (July 25, 1952) (risk of the death penalty; torture or inhuman or degrading punishment or treatment; serious, direct and individualised threat to his life or person because of generalised violence resulting from internal or international armed conflict); *Loi no. 99/586* (July 12, 1999) (“Loi Chevènement”), § 13 (rejected asylum seeker who, if returned, would face a threat to his life or freedom or would be at risk of treatment contrary to Article 3 of the ECHR). Also note that France may be considered a specially interested state based on the receipt of high numbers of refugees and persons in refugee-like situations. *See supra* note 65.

States, all provide in their domestic legislation for subsidiary protection for a class of persons wider than the conventional definition of refugee. Additionally, many states other than those in the top twenty-five in terms of refugee recipients have also adopted similar subsidiary protection regimes, including Australia, Austria, Belgium, Canada, Denmark, Ecuador, Finland,
Greece, Ireland, Israel, Italy, Luxembourg, the Netherlands.
Worster: The Evolving Definition of the Refugee in Contemporary Internation
130 BERKELEY JOURNAL OF INTERNATIONAL LAW [Vol. 30:1

Portugal,205 Russia,206 Spain,207 Sweden,208 and Switzerland.209 There is, of

(Oct. 12, 2000), art. 3 (granting suspension of judicial deportation when deportation is not possible, especially when alien’s life is in danger.).

200. See Immigrant Act of 1999, § 3(6) (1999) (allowing leave to remain for humanitarian or other reasons, such as illness, family connections and personal considerations); Refugee Act of 1996 § 17(6) (1996) (allowing for a discretionary right to remain under protective status if asylum application was withdrawn, denied or revoked; this protection status has never been used); Refugee Act of 1996 § 5(2), Criminal Justice (United Nations Convention Against Torture) Act, § 4 (2000) (prohibiting refoulement if life or freedom (including sexual assault) would be threatened on account of his race, religion, nationality, membership in a particular social group or political opinion, danger of being subject to torture).

201. See Belcha v. Min. of Int. Aff’rs & Trib. for Jud. Rev., AdmAp 2028/05 (Tel Aviv Dist. Ct., Isr. Feb. 8, 2006), reprinted at ILDC 290 (IL 2006) (expressing willingness to expand the non-refoulement protection beyond the narrow class of Convention refugees with sufficient proof of a well-founded fear of persecution to those not formally qualifying as “refugees” but deserving protection from a risk of persecution, but only as a temporary measure).

202. See ITALY CONST. art. 10, ¶ 3 (granting political asylum to those not allowed to exercise in his own home country the democratic liberties guaranteed by the Italian Constitution). Note that this was traditionally considered a hortatory statement but was later interpreted to have binding legal effect. See Dec. No. 4674 (Ct. Cass. May 26, 1997); In re Abdullah Ocalan (Trib. Rome Oct. 1, 1999) discussed in ECRE, Residence Permits on Humanitarian Grounds; Complementary/Subsidiary Forms of Protection in the EU Member States: An Overview: Italy, http://www.ecre.org/files/survcompro.pdf; Law No. 189/02 (providing for asylum explicitly as a consequence of ECHR, art. 3); Legis. Decree No. 286/98 “Testico Unico,” art. 5(6) (applying to those that do not qualify as Convention refugees but cannot be returned due to serious humanitarian reasons or constitutional or international obligations of the Italian State).

203. See Act of 18 Mar. 2000, art. 13 (applying to persons denied refugee status but who cannot be returned for practical reasons such as refusal to re-admit or health problems).

204. See Aliens Act 2000 §§ 26, 27, 29 (granting “residence status for humanitarian reasons” to those at risk of being subject to torture or inhuman or degrading treatment or punishment, pressing humanitarian reasons, for whom return would constitute an exceptional hardship due to violence in the country of origin, human rights violations, or if a spouse or parent has the same nationality and was granted a residence permit within three months of the applicant’s application for admission); Aliens Act 2000 § 43 (if decision on the asylum application is extended beyond the typical six-month period because of insecurity in the country of origin, situation justifying the grant of a residence permit is expected to be of a short duration, number of applications lodged by persons from a particular country or region is so large the Minister cannot reasonably decide them within the six-month time limit); Aliens Act 2000 § 45(4) (asylum applications were rejected but return is temporarily impossible); Aliens Act 2000, art. 14, Aliens Decree, art. 3.6 (granting residence permits to those unable to leave the Netherlands through no fault of his or her own, such as stateless persons and unaccompanied minors). Also note that the Netherlands might be considered a specially interested state, although it is not so highly affected by refugee flows as some other states examined. See supra note 65.

205. See 1998 Asylum Act, art. 8 (granting residence permit to those not qualifying for Convention refugee status for humanitarian reasons, e.g., serious lack of security resulting from armed conflicts or systematic violation of human rights); Aliens Act of 2003 (Decree Law No. 34/2003), art. 87, revoking Aliens Act of 2001 (Decree Law No. 4/2001), art. 55 (granting residence permit in case of irregular stay to those: gravely ill, with relative who has a humanitarian residence permit, are married to or live with Portuguese nationals or aliens with residence permits, or have minor children living in Portugal; artistic, scientific, economic or social activity of high interest for the country; diplomats who have worked in Portugal for more than three years), art. 88 (granting exceptional residence permit for reasons of national interest).

206. See 1997 Law on Refugees, arts. 1(1), 2, 12(2) (granting temporary refuge to those not
course, no bright line dividing the top twenty-five recipient states from others in terms of which are considered specially interested or not. Some of the states that are not within the top twenty-five might also be considered “specially interested,” or at least highly influential, due to their rates of receipt of individuals in need of protection, especially, for example, in relation to their population levels, resources, etc. Common bases for subsidiary protection by these states include: (1) the risk of torture or degrading punishment,\textsuperscript{210} (2) the imposition of capital punishment,\textsuperscript{211} (3) impossibility or futility of return,\textsuperscript{212} (4) existence of a state of armed conflict,\textsuperscript{213} (5) environmental disaster or deprivation of resources,\textsuperscript{214} (6) strong ties to the state or family ties,\textsuperscript{215} (7) grave illnesses,\textsuperscript{216} (8) gender or qualifying as permanent refugees for humanitarian reasons, armed conflicts, serious health problems, instability in the country of origin, family unity).

\textsuperscript{207} See Law 5/1984, Reglamento de Aplicación de la Ley de Asilo [Regulating Refugee Status and the Right of Asylum], arts. 17(2), (3), approved by Royal Decree 203/1995 (10 Feb. 1995), modified by Royal Decree 864/2001 (20 Jul. 2001) and Royal Decree 1325/2003 (25 Oct. 2003), art. 31.3 (granting residence permits to persons not qualifying for refugee status but who are displaced persons under a serious risk of exposure to systematic or generalised human rights violations), art. 31.4 (granting leave to remain to those refused asylum for humanitarian reasons or reasons of the public interest, such as people fleeing conflict and other serious disturbances; prohibiting refoulement for humanitarian reasons); Regulation on Temporary Protection, Royal Decree 1325/2003 (25 Oct. 2003), art. 2.

\textsuperscript{208} See Aliens Act, 2005, ch. 12 §§ 1, 2 (prohibiting refoulement to in a situation where there is a risk of a resulting situation of danger), ch. 3, § 3 (granting protection to those who do not qualify as Convention refugees but have a well-founded fear of being sentenced to death or corporal punishment or torture or other inhuman or degrading treatment or punishment, external or internal armed conflict, environmental disaster preventing return, or a well-founded fear of persecution because of his/her sex or sexual orientation), ch. 8, §§ 1-4 (prohibiting refoulement of those not granted refugee status or subsidiary protection, if reasonable grounds exist for believing that they will face capital or corporal punishment or torture or other inhuman or degrading punishment or treatment, or persecution), ch. 2, § 5b (providing for reconsideration of a decision of expulsion for those otherwise in need of international protection or if decision is contrary to humanity), ch. 2, §§ 4(5), 5 (humanitarian reasons, such as certain physical or mental handicaps or diseases or home country is in a state of war). Also note that Sweden might be a specially interested state, although it is less affected than some other states, see supra note 65.


\textsuperscript{210} Austria, Belgium, Denmark, Finland, Germany, Greece, Ireland, Netherlands, Sweden, and the United Kingdom. Note that Germany might be considered a specially interested state. See supra note 65. Also note that the Netherlands might be considered a specially interested state, although it is not so highly affected as some other states examined. See supra note 65.

\textsuperscript{211} Austria, Denmark, Finland, Germany, Sweden, and the United Kingdom. Also note that Germany and the United Kingdom might be specially interested states, see supra note 65; and that Sweden might also be so considered, although it is less affected than some other states, see supra note 65.

\textsuperscript{212} Austria, Belgium, Denmark, Germany, Greece, Luxembourg, and the Netherlands.

\textsuperscript{213} Belgium, Denmark, Finland, Germany, Greece, the Netherlands, Spain, Sweden, and the United Kingdom.

\textsuperscript{214} Denmark, Finland, Greece, Portugal, and the United Kingdom.

\textsuperscript{215} Belgium, Denmark, Ireland, Netherlands, and the United Kingdom.

\textsuperscript{216} Belgium, Greece, Ireland, Luxembourg, Portugal, and Sweden.
sexual orientation persecution, and (9) other violations of human rights. In addition, many of these aforementioned states conflate terms of “refuge” and “asylum.” For example, they grant “refuge” to individuals qualifying on grounds not mandated by the Refugee Convention, suggesting that the customary international legal definition of “refugee” may encompass more than Convention “refugees.” As an example of the convergence between refuge and asylum, consider that from 2000 to 2002, European states granted protection to approximately 70,000 applicants each year, of which approximately 57,000 subsequently received asylum.

In addition to protections for risk of torture or other subsidiary grounds, some states also apply a general proportionality test to the expulsion of any person for any reason, including individuals whose refugee status was either not recognized or terminated. This test usually weighs the need for expulsion against the dangers facing the person on return. Just as subsidiary rules have effectively expanded the definition of “refugee” beyond the narrow meaning in the Refugee Convention, general proportionality concerns may have done so as well. In applying Article 33(2) of the Refugee Convention, though, the New Zealand Supreme Court found that the Refugee Convention did not require this proportionality test. Accordingly, with only two reported states applying such a test and one rejecting it, there is insufficient practice on point on which to comfortably base a rule of customary international law.

Many states also apply the same or similar standards in determining subsidiary protection as they do in determining refugee status, although they maintain a formal distinction between the two statuses and can apply slightly differing standards on certain aspects. This practice might suggest an opinio juris to

217. Austria, Denmark, Finland, and Sweden.
218. The Netherlands and the United Kingdom.
221. See e.g., Israel: Belcha v. Min. Int. Aff’rs, AdmAp 2028/05, reprinted at ILDC 290 (although petitioners did not qualify for refugee status, the life-threatening situation they would face upon return was disproportionate to the problems with their continued stay, justifying a temporary stay of deportation pending designation of a safe third country); Switzerland: In re M.C.C., Somalie, Case 2006, No. 2 (ARK Dec. 13, 2005) (finding that the chaotic situation and the permanent state of violence in Somalia rendered removal unreasonable because it would be disproportionate).
222. See AG v. Zaoui, (2005) N.Z.S.C. 38, ¶ 42 (Sup. Ct., N. Zealand June 21 2005), reprinted at ILDC 81 (NZ 2005). Article 33(2) [of the Refugee Convention] did not, however, invite a ‘proportionality’ or ‘sliding scale’ approach. It stated a single standard and was to be applied in its own terms by reference to the danger to the security of New Zealand without any individualized balancing or weighing of the particular risk of deportation to the individual. Id.
223. See ECRE Country Report 2004, supra note 5 (“Changes were brought about after the introduction of subsidiary protection into French legislation. Grounds for the use of exclusion clauses are the same for subsidiary protection as for Convention Status, except for the concept of ‘non-political crime’ which is wider for subsidiary protection than for Geneva Convention Status because
maintain the separate categories of refugee and subsidiary protection, and deny that there is any international legal basis for an expanded definition beyond the Refugee Convention. Consider, for example, the language of the United States’ Ninth Circuit Court of Appeals:

“[The] customary international law of safe haven and nonreturn is not a separate basis for jurisdiction before immigration court [a defense against removal by a municipal court].”

However, given the position of specially interested states regarding the UNHCR definition and the widespread and consistent practice of providing and granting subsidiary protection with comparable content, there may be customary international law requiring states to provide, at least, subsidiary protection in some of these situations, though not necessarily a customary expansion of the definition of refugee. In any event, that may be a difference without significance if it results in an expanded class of persons deriving a right of non-refoulement.

F. Practice and Opinio Juris of Entities Other Than States

It is highly contentious whether the acts of entities other than states can contribute to the formation of customary international law. Where the acts of international organizations have been so considered, they are often explained as relevant because they are the collective expression of the practice of states. However, this analysis suffers from the weakness of characterizing some acts as those of the states within the international organization and those of the international organization proper, i.e., exercising a “will of its own.”

It is unconvincing that international law accepts the practice of international organizations themselves, as opposed to the practice of states within and through international organizations, as contributing to customary international law. However, given that this argument is on-going and that the trend appears increasingly in favor of accepting the practice of international organizations as contributory, this Article next examines the practice of organizations. In particular, it will be noted where the practice of organizations is more indicative of the practice and opinio juris of states, such as instances where states have made their opinio juris known through the organization or where they have potentially

it includes host countries.”

224. See Galo-Garcia v. Immigr. & Naturalization Serv., 86 F.3d 916 (9th Cir. 1996).
225. See e.g., Prosecutor v. Tadić, Case No. IT-94-1-l, Decision on Defense Motion for Interlocutory Appeal on Jurisdiction, ¶¶ 113-15 (Oct. 2, 1995) (not clearly distinguishing between the acts of the states within the organization or the acts of the organization proper when relying on the practice of international organizations for establishing customary international law); RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES, § 102, Rep’s n.2 (1987) (“The practice of states that builds customary law takes many forms and includes what states do in or through international organizations”). See also HENRY G. SCHERMERS & NIELS M. BLOKKER, INTERNATIONAL INSTITUTIONAL LAW § 33, 44-44A (4th rev. ed., 2003) (discussing the notion of a “will of its own”). Of course, that the characterization of some acts as either being attributable to the organization or the member states is difficult to make.
adopted the practice and *opinio juris* of the organization.

1. United Nations High Commissioner for Refugees Practice and *Opinio Juris*

Kay Hailbronner has argued that the practice of the UNHCR is not state practice and therefore does not contribute to the formation of customary international law.226 It is most likely that it is the practice of states that contributes to the formation of customary international law, not the practice of institutions, and that Hailbronner is therefore correct insofar as her precise argument stands. However, there is a distinction that acts of international organizations are capable of embodying practice and *opinio juris* of states where the act in question is not an act of the organization *per se*, but the act of states within and through the organization.227 A good example of this type of act is the voting records of states in the UN General Assembly. Acknowledging that a resolution of the General Assembly is an act of the organization with its prescribed legal effects, the voting behavior of a member state remains the practice of the state with its own legal effects in the form of the potential contribution to the formation of customary international law.228 In addition, the acts of the organization might also be said to embody the practice of the state where the state delegates its state functions to the international organization,229 or where states cite the mandate of

226. See e.g., Kay Hailbronner, *Non-Refoulement and Humanitarian Refugees*, 26(4) VA. J. INT’L L. 857, 869 (1986) (“Although the UNHCR fulfills its functions with the agreement of states, it remains a special body entrusted with humanitarian tasks … the fact that the UNHCR continues to care for the interests of de facto refugees cannot be considered evidence of an *opinio juris* by states.”).

227. See supra note 225.

228. See *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 I.C.J. 226, ¶ 70 (July 8). The Court notes that General Assembly resolutions, even if they are not binding, may sometimes have normative value. They can, in certain circumstances, provide evidence important for establishing the existence of a rule or the emergence of an *opinio juris*. To establish whether this is true of a General Assembly resolution, it is necessary to look at its content and the conditions of its adoption; it is also necessary to see whether an *opinio juris* exists as to its normative character. Or a series of resolutions may show the gradual evolution of the *opinio juris* required for the establishment of a new rule.

the UNHCR as generally supervisory. It is not a stretch to understand that where a state so delegates its functions, it is, at least, expressing an opinio juris that the legal standards applied by the organization are the correct ones, and, at most, that the opinio juris of the organization could be attributed to that state. An example appropriate here is the delegation by states to the UNHCR of the refugee status determination function, studied extensively above under the practice of specially interested states. Therefore, the practice of the UNHCR cannot easily be dismissed, and, in fact, it might be representative of the opinio juris and practice of states.

The Codification Division of the UN Secretariat suggested that the practice of the UNHCR and states, and opinio juris, might even be seen through a particular lens: that they may wish to refrain from designating certain individuals as refugees even when they qualify as such, because the recognition of status would have political effects on the state now recognized to be a persecuting state. If this is correct, then it is worth being more aggressive in identifying instances where states and/or the UNHCR might consider the relevant persons to be refugees because it is known that those actors will avoid communicating that fact, even if they regard themselves under an obligation to protect.

The UNHCR was established to address the needs of refugees, but its competence and mission have been extensively expanded by the UN General Assembly, without necessarily expanding the core definition of its mission. This practice arguably expands the international community’s understanding of refugee. It is interesting to note that the UNHCR’s mandate has been expanded to cover those who have a “serious threat to their life, liberty or security of person in their country of origin as a result of armed conflict or serious public disor-

---

230. See UNHCR – Brasilia Declaration, supra note 86. It is interesting to note that the Brasilia Declaration discusses the work of the UNHCR as covering refugees, stateless persons, and internally displaced persons. Since the UNHCR mandate also covers persons in “refugee-like” situations of flight out of the state from situations of external aggression, occupation, foreign domination or events seriously disturbing public order, and such individuals cannot be classified as stateless or internally displaced, the logical conclusion is that the parties to the Brasilia Declaration understand those individuals covered by the expanded UNHCR mandate to be refugees per se. See id.

231. See U.N. Secretariat Memo, supra note 48, at ¶ 158.


233. In particular, it expands the understanding of specially interested states, such as China, Jordan, Kenya, Pakistan, Saudi Arabia, Syria and Tanzania, that appear to have adopted the practice and opinio juris of the UNHCR based on the fact that they have, albeit sometimes formally and sometimes informally, delegated the refugee status determination procedure to the organization. See USCRI, China, supra note 140; USCRI, Jordan, supra note 135; USCRI, Kenya, supra note 155; USCRI, Pakistan, supra note 127; USCRI, Saudi Arabia, supra note 159; USCRI, Syria, supra note 119; USCRI, Tanzania, supra note 138. Germany, the United Kingdom, and France might be considered to be highly influenced by the UNHCR definition of individuals falling under its mandate, though not having delegated status determination to the organization. See USCRI, Europe, supra note 130. The practice and opinio juris of Nepal as per the UNHCR is more ambiguous. See USCRI, Nepal, supra note 184.
Worster: The Evolving Definition of the Refugee in Contemporary Internatio

der.234 This mandate should include:

[A]ll persons who, owing to external aggression, occupation, foreign domination or events seriously disturbing public order in either part of the whole of their country of origin or nationality are compelling to leave their place of habitual residence in order to seek refuge in another place outside the country of origin or nationality.235

Interestingly, this expansion of the mandate of the UNHCR closely tracks the expanded definition of refugee that has developed through regional agreements discussed above. It is important to note that the Refugee Convention itself was not amended to expand the refugee definition. The only expansion was in the UNHCR organizational mandate. However, the expansion of the mandate of the agency charged with ensuring the protections of the Refugee Convention might suggest an opinio juris of states that some supplementary “refugee” definition exists.236

It could be that war and aggression are seen as falling within the broad notion of persecution. The European Council on Refugees and Exiles has noted that “it is hard to conceive of a recent war or civil war situation which has not resulted in or been motivated by persecution for one of the grounds in Article 1A(2) of the Refugee Convention.”237 This statement suggests that war or other comparable disruptions might alone qualify as persecution. However, the ECRE does not reach this conclusion specifically, subsequently noting that the UNHCR Executive Committee at its 49th Session observed only that “the increasing use of war and violence as a means to carry out persecutory policies against groups targeted on account of their race, religion, nationality, membership of a particular social group, or political opinion.”238 While war could be a means of persecution, it was not persecution per se and the individuals who fled did not qualify as refugees. However, this interpretation makes it difficult to understand why the UNHCR’s mandate would be expanded if the individuals it was now charged with protecting were not being persecuted and falling under the Refugee Convention on that basis. It could suggest that the expansion of the


238. See id. at ¶ 30.
mandate was not an effort to expand the meaning of persecution, thus keeping
the Convention largely intact, but rather the development of a supplementary
definition of refugee.

Arguing against the UNHCR’s mandate as supplementing the Refugee
Convention definition of refugee, the UNHCR’s mandate is still limited to
“those in refugee-like situations,” and the terms of its statute under which it acts
appear to permit it to offer assistance to individuals who are not formally refu-
gees.239 Thus, the mandate of the UNHCR appears to consider these individuals
under the expanded mandate as not literally qualifying as “refugees.” Therefore,
the mandate alone does not appear to expand the conventional definition of “ref-
ugee”, even if the practice of that office could potentially do so. Additionally,
states often refuse to accept individuals for settlement by the UNHCR if the pro-
tected individual does not qualify as a refugee under the Refugee Convention.240
The foregoing forces the conclusion that the expansion of the UNHCR mandate
most likely does not, in itself, expand the conventional definition of refugee.
However, even if it is not expanding the conventional definition, it may contrib-
ute to the formation of customary international law on the provision of refuge to
a broader scope of protected individuals.

2. Council of Europe Practice and Opinio Juris

The Council of Europe’s practice is also relevant. Recommendation (2001)
18 of the Committee of Ministers of the Council of Europe proposes that Europe
take a common approach to refugee qualifications, suggesting that the risk of
torture or inhuman or degrading treatment or punishment; a threat to life, securi-
ty or liberty because of indiscriminate violence arising out of situations such as
armed conflict; or other reasons recognized by legislation or practice in a mem-
ber state, all establish a person as a “refugee.”241 We can consider this the
 opinio juris of the states of the Council of Europe, especially since it was adopt-
ed by the Committee of Ministers, and more likely to be expressions of opinio
juris by states through the organization, rather than an expression by the organi-
zation itself.

239. UNHCR Statute, supra note 234, at art. 9 (providing that the High Commissioner may
“engage in such activities … as the General Assembly may determine, within the limits of the re-
sources placed at his disposal”); U.N. Office for the Coordination of Humanitarian Affairs, Guiding
240. See Mandal, supra note 69, at ¶ 265 (discussing US policy).
241. See Council of Europe Parliamentary Assembly Recommendation 773 (1976) on the Situa-
tion of De Facto Refugees; Council of Europe Parliamentary Assembly Recommendation 817 (1977)
on Certain Aspects of the Right to Asylum; Council of Europe Parliamentary Assembly Recommen-
dation 1327 (1997) on the Protection and Reinforcement of the Human Rights of Refugees and Asy-
lum-Seekers in Europe; Council of Europe Parliamentary Assembly Recommendation 1525 (2001)
on the United Nations High Commissioner for Refugees and the Fiftieth Anniversary of the Geneva
Convention.
3. EU Practice and Opinio Juris

Turning to EU law, the Qualification Directive clearly applies to refugees as defined in the Refugee Convention, but it establishes a separate category for other non-refugee persons who deserve protection. At least three of the member states of the European Union are specially interested states, and all member states, including the specially interested ones, are bound to comply with EU law. Accordingly, EU practice and opinio juris must carry significant weight, even if we refuse to accord it formal significance as contributing to the formation of customary international law. It reflects the practice of a considerable number of states in the world, including at least three specially interested ones.

In the Qualification Directive, refugees are defined according to the Refugee Convention, with the difference that EU nationals are exempt from the definition. Insofar as its treatment of armed conflict, the Directive, and especially Article 15(c), draws on the language in the OAU Convention, the Cartagena Declaration and the UNHCR’s widened mandate to establish categories of persons deserving protection. However, it implicitly considers these to be situations outside the definition of “refugee” since it classifies them as falling under the definition of “subsidiary protection.” In Elgafaji v. Staatssecretaris van Justitie, Case C-465/07, 2009 E.C.R., the European Court of Justice clarified the qualification for subsidiary protection as not needing to be “specifically targeted” for harsh treatment, but rather only to have suffered a “serious and individual threat” due to indiscriminate violence—meaning that there was an inverse relationship between the level of violence and the specificity of the threat. However, the Court did not attempt to broaden the definition of refugee, keeping clearly within the terms of subsidiary protection. This language is evidence that the European Union does not have the opinio juris that the definition of refugee is broader than the Refugee Convention, although it does express an opinio juris that there may be a norm of supplementary subsidiary protection.

4. Other Instances of International Organization Practice and Opinio Juris

Arguably, there is opinio juris for an expanded definition of refugee based on UN Declarations dealing with enforced disappearances and extra-legal executions, extradition treaties, and international humanitarian law. However,
by their own terms these declarations do not appear to strive for an expanded
definition of refugee but only expand or narrow prohibitions on expulsion.248

Some have suggested that the 1992 Rio Declaration, the UN Framework
Convention on Climate Change, the Convention on Biological Diversity, and
Agenda 21 contain obligations to adopt a program of common responsibility for
sustainable development and prevention of environmental refugees.249 Could
this obligation include an obligation to receive “environmental refugees,” thus
expanding the definition of refugee in customary international law? Lacking ex-
plicit treaty law or clear evidence of customary international law, this appears to
be a weak argument.250 The UN Codification Division concluded that “interna-
tional law has yet to confer refugee status on victims of environmental condi-
tions,”251 which appears to be the correct assessment.

5. Conclusion on International Organization Practice and Opinio Juris

Based on the foregoing, the practice of international organizations is rather
inconsistent. In general, it suggests a practice and opinio juris to expand the
scope of protection to deserving individuals, but it also shows a reluctance to use
the term “refugee” explicitly to expand the conventional definition in customary
international law. Instead, the approach seems to be to provide for supplementary
categories of subsidiary protection. It may very well be that customary in-
ternational law now provides for a norm of subsidiary protection.

However, this limited contribution by international organizations, com-
bined with the contribution of specially interested states, the primary makers of
customary international law, might result in expansion of the definition under
customary international law to cover a wider scope of deserving individuals. In
particular, it appears that there is a definition of refugee under customary inter-

247. See generally Sibylle Kapferer, External Consultant, UNHCR, Dep’t of Int’l Protection, Protection Policy & Legal Advice Sec., The Interface between Extradition and Asylum, LEGAL & PROT. POL’Y RESEARCH SER. (Nov. 2003) (discussing the relationship between extradition and asy-

tions relating to POWs “do not create individual remedy for relief from deportation”).

249. See Arthur C. Helton, The Legal Dimensions of Preventing Forced Migration, 90 PROC. AM. SOC’Y INT’L L. 546 (1996); but see L.F. (Ukr.) v. Min. Int., Case No. 5 Azs 38/2003-58 (Sup. Admin. Ct., Cz. Rep. Feb. 25, 2004) (holding that environmental conditions, such as the damage to the region surrounding the Chernobyl nuclear reactor, could not serve as the basis for a claim to ref-


251. See U.N. Secretariat Memo, supra note 48 (citing David Adam, 50m environmental refu-

national law that covers, in addition to conventional refugees: (1) individuals who would suffer torture, inhuman or degrading treatment and/or punishment upon return; and (2) individuals who are fleeing a threat to life, security or liberty due to external aggression, armed conflict, occupation, foreign domination or events seriously disturbing public order, including widespread indiscriminate violence. Individuals that might suffer imposition of the death penalty may in the future be added specifically to that list, but they do not fall under the current international law definition. The only exceptions would be cases where the imposition of the death penalty in that state would result in situations of torture or inhumane punishment.

IV.
CUSTOMARY INTERNATIONAL LAW NARROWING THE DEFINITION

The above discussion has focused on the expanding definition of refugee and concluded that the definition has most likely been expanded to cover more individuals than just those facing a risk of persecution. This finding, in turn, means that more individuals can qualify for refugee protection. However, refugee qualification has also been contracting in some important ways that may be relevant. If it is accepted that customary international law may have supplemented the Convention in a “positive” sense (i.e., expanding the definition to offer protection to a wider range of persons), then it must be considered whether customary international law may be operating in a “negative” fashion (i.e., restricting the persons and situations covered). This narrowing of the definition might go further than the terms of the Refugee Convention. The sources of international law, principally treaty and custom, are generally perceived to be equal, so that a custom can, in theory, operate to restrict obligations incurred under a treaty, although the evidentiary demands for such a change may be rather high.252

A. Territorial Application

One method that states have been adopting to narrow the application of refugee law is to interpret the territorial application of the Refugee Convention restrictively. Although this is not formally an aspect of the definition of refugee under customary international law, it does impact the determination of when an individual is outside his country of nationality. The US Supreme Court has found that the Refugee Convention does not apply outside of the territory of the United States,253 and the Immigration and Nationality Act (which includes the refugee and asylum provisions) does not apply on Midway Island because it is

252. See e.g., Delimitations of the Continental Shelf (U.K./Fr.), Award, ¶ 47, 18 R.I.A.A. 3 (June 30, 1977).

not formally a part of the United States. Russia has adopted a similar interpretation in following domestic legal definitions of its territory. These interpretations appear to leave a vacuum between the fact that the individual in question is clearly outside his country of nationality and the territorial application of treaty obligations. It does not appear to attempt to modify the definition of refugee. In addition, aside from certain unusual territorial situations such as Midway Island, states do not appear to claim non-application in their metropolitan territory, so this exception is rare and not widespread.

B. Internal Flight or Relocation Alternative

Internal flight or relocation within the state of nationality has become an alternative to refugee status developed in case law. Courts have found that relocation within the state of nationality can allow for the individual not to face the danger needed for refugee qualification. The language of the Refugee Convention does not explicitly provide for internal flight as a discrete basis defeating refugee status, and the notion has been criticized. However, the courts of many states have found, on a widespread and fairly consistent basis, that the exception is inherent in the definition.

Although not entirely clear on what provisions of the Refugee Convention this interpretation rests, it appears to be a combination of the inclusion clauses, the cessation clause (making a refugee capable to re-avail himself of the protection of the state of nationality) and the non-refoulement obligation against returning the individual to the state where he would face persecution. Additionally, the internal flight rule has been argued to be inherent in the concept of the need for international protection. Because of this unclear basis, it is also hard to determine whether the existence of an internal flight alternative means that the individual does not qualify (even de facto) as a refugee under the Refugee Convention or whether the individual may qualify (because he was persecuted or faces a real risk of persecution), but falls into an exception of the rule of non-

254. See In re Li, 71 F. Supp. 2d 1052 (D. Hawaii 1999). But see Rasul v. Bush, 542 U.S. 466 (2004) (finding that for habeas corpus, the application of US law required that the US exercise effective, as opposed to merely formal, sovereignty, such as over Guantanamo Bay).


refoulement. This Article will address the internal flight alternative as an exception to the rules on qualification.

Although the internal flight requirement is applied by a number of states, there are several different ways that this exception might be applied. The UK House of Lords considered two approaches to the interpretation of internal flight. The first is that the internal flight alternative should be restrictively applied to situations where: (1) there is genuine access to the area of domestic protection (existence of financial, logistical, or other barriers); (2) the protection is meaningful (meets basic norms of civil, political, and socio-economic human


rights); and (3) the protection is not illusory or unpredictable. The second approach is to compare the situation for persons with the relevant characteristics in the individual’s area of residence and the proposed area of protection. This approach arose out of UK case law that tends to merely consider whether it is “unduly harsh” to return an individual to the alternate region of the state. The House of Lords held in favor of the second approach because the first approach was not mandated by the EU Council Directive 2004/83/EC, the Refugee Convention did not appear to provide a basis for it, there was no sufficient practice to support it, and the Lords thought it would be strange to permit an individual to escape the general conditions of life that all his fellow countrymen were subject to just because he would suffer persecution in one area of the country.

The cases in other countries appear to fall on either end of this spectrum or somewhere in between. Countries such as France and Switzerland appear to apply a test closer to that of the first approach, whereas countries such as Germany appear to take the second approach. The European Union has established a slightly different test of whether the individual could lead a “normal life” in another part of the country, taking into account social and economic criteria, as well as the possibilities of finding a job. Also note that France may be considered a specially interested state based on the receipt of high numbers of refugees and persons in refugee-like situations. See supra note 65 and accompanying text.

261. See id.
262. See id.
263. See Sec’y St. Home Dep’t v. A.H., [2007] U.K.H.L. 49; Januzi v. Sec’y St., [2006] U.K.H.L. 5, ¶ 47 (“The words “unduly harsh” set the standard that must be met for this to be regarded as unreasonable.”); Karanakaran v. Sec’y St. Home Dep’t, [2000] E.W.C.A. Civ. 11 (holding that when assessing the reasonableness of internal flight alternative, the decision-maker should simply ask: would it be “unduly harsh” to expect the applicant to settle there?).
265. See ECRE Country Report 2004, supra note 5 (“The Commission examines the possibility for applicants to have a ‘normal life’ in another part of the country, taking into account social and economic criteria, as well as the possibilities of finding a job.”). Also note that France may be considered a specially interested state based on the receipt of high numbers of refugees and persons in refugee-like situations. See supra note 65 and accompanying text.
266. See e.g., In re M.C.C., Dec. 2006/2 – 015 (holding that in certain conditions return to Somalia might be reasonable, e.g., special strong ties to the safe region, able to establish a stable existence, ability to rely on a functioning family/clan structure, etc.; however, the court determined that it was not reasonable in the situation at hand).
267. See e.g., Case No. 1 LA 79/04 (OVG Schleswig-Holstein Oct. 7, 2004) (only examining the risk of persecution in the safe area); Case No. 1 K 3266/01, VG Arnsberg (17 Mar. 2004) (same); Case No. 25 K 3188/03.A, VG Düsseldorf (16 Dec. 2004) (same); Case No. 6 K 4833/03.A, VG Düsseldorf (15 July 2004) (same); Case No. A 11 K 10417/02, VG Karlsruhe (10 Mar. 2004) (same); Case No. 2 E 1598/02.A, VG Kassel (2 June 2004) (same); Case No. 2 A 94/01, VG Lüneburg (26 Feb. 2004) (same); Case No. 1 A 2944/01, VG Oldenburg (17 May 2004). But see e.g., Case No. 2 LB 54/03, OVG Schleswig-Holstein (16 June 2004) (examining the degree of effective government); Case No. 8 UE 216/02.A, VGH Hessen (10 Feb. 2005) (same); Case No. VG 33 X 302.96, VG Berlin (2 Feb. 2004) (same); Case No. A 7 K 31035/03, VG Dresden (16 Mar. 2004) (same); Case No. 7 K 1517/00.A, VG Frankfurt/Oder (2 Mar. 2004) (same); Case No. 5a K 8121/95.A, VG Gelsenkirchen (11 Nov. 2004) (same); Case No. 9 K 4856/03.A, VG Minden (26 Apr. 2004) (same); Case No. 5 K 1900/03.NW, VG Neustadt a.d.W (26 Apr. 2004) (same); Case No. 7 E 2245/03.A(V), VG Wiesbaden (4 Nov. 2004). Also note that Germany may be considered a specially interested state, see supra note 65 and accompanying text.
life” in the alternate region.268 The ECHR has held only that establishing an internal flight alternative test is acceptable as long as reprehellement to persecution does not result.269

Because the internal flight/relocation alternative exception is widespread, and also because the test for an internal flight alternative could be argued to be inherent in the refugee qualification regime, it appears to be part of refugee law under customary international law. Which of the variations of the exception controls is not entirely clear; however, a broader test for a relocation alternative is more likely to be the approach favored by the House of Lords. The reason to favor this interpretation is that, since the exception partly arises from the terms of the Refugee Convention itself, which merely requires non-refoulement to persecution (as well as torture and situations of serious internal disturbances under customary international law), it follows that an individual is not relieved of the general conditions of life of his countrymen that do not rise to the level of persecution.

C. Safe Third Country or Country of Origin Policies

States also apply safe third country and/or safe country of origin tests to refuse claims without further review.270 Similar to the discussion above, there are

---

268. See Qualif. Dir., supra note 103, at art. 8(1) (providing that states may reject claims for refugee status if there is an internal flight alternative); ECRE Country Report 2004, supra note 5 (“The Commission examines the possibility for applicants to have a ‘normal life’ in another part of the country, taking into account social and economic criteria, as well as the possibilities of finding a job”).

269. See Sheekh v. Neth., Appl. No. 1948/04, Eur. Ct. H.R. (2007) (holding that the indirect removal of an alien to an intermediary country did not affect the responsibility of the expelling contracting state to ensure they were not, consequently, exposed to treatment contrary to Article 3 of the ECHR). There was no reason to hold differently where expulsion was to a different area of the country.

two approaches to applying these tests. The first approach is considering a safe third country as a factor in the assessment of whether the person qualifies under the Convention. The second approach is using the safe third country test as a means to refuse claims outright without analysis if the individual is coming from a country that has been deemed safe. This second possibility is often paired with expedited processing procedures. It is this second practice that will be addressed here.

It goes without saying that no state is entirely safe. The UNHCR reports on the states of origin of refugees or other persons of concern. The list is surprisingly universal, including most states in the world. Although there are clearly states that produce a significant amount of refugees,\(^{271}\) even Gibraltar, Palau, the Turks and Caicos Islands, Brunei, Luxembourg, St. Kitts and Nevis, Samoa, Tuvalu, San Marino, Nauru, Norway, Finland, Tonga, Timor-Leste, Iceland, Lesotho, Malta, Andorra, Cyprus, Macao, and Ireland produce a handful of refugees each year.\(^{272}\)

Some states have provided under their municipal laws and international agreements for the discretion to refuse claims where the individual is in transit from a safe country.\(^{273}\) The arguments for these policies are: (1) that in order to manage the countless number of applications, it is efficient and in line with the

---

271. See Table 2, 2007 UNHCR Statistical Yearbook, http://www.unhcr.org/4981b1942.html (for statistics dated at the end of 2007). The states from which the most refugees come are as follows: Iraq (2,279,245), Afghanistan (1,909,911), Sudan (523,032), Somalia (455,356), Burundi (375,715), D.R. Congo (370,386), Palestinian Territories (335,219) and Vietnam (327,776).

272. See Table 2, supra note 271; see also, id. for additional refugee statistics on small numbers of refugees from a variety of states and territories, e.g., Hong Kong (11), New Zealand (13), Bahamas (14), Denmark (14), Botswana (16), Sweden (16), Belize (17), and the Maldives (17).

Refugee Convention to assess certain states as safe and deny all applicants from that state, and (2) individuals should not engage in forum-shopping for the adjudication of their refugee status claim but should instead apply in the first safe country they reach. Based on these arguments, governments designate certain states as “safe” and prohibit application from their nationals and, if their state of origin is not safe, return individuals to the first safe state that the individual reached upon fleeing. There is no provision in the Refugee Convention for these policies, since the definition of refugee only considers whether the person is outside his state of nationality and whether he qualifies. Nonetheless, the United States has adopted safe third country legislation,274 as has Finland,275 France,276 Germany,277 Ireland,278 Switzerland,279 among others, as have the non-EU European states of Belarus280 and Norway.281 Many of the European states were quick to designate other EU and EFTA states as “safe.” In addition, the Qualification Directive defines refugees as third party nationals or stateless persons, meaning that EU nationals are excluded from the definition.282 What is curious is that of the other EU states not already mentioned above as producing refugees (Luxembourg, Finland, Malta, Cyprus, and Ireland, and the EU-linked entities of

274. See INA § 208(a)(2)(A).
277. See Asylum Procedure Act, June 26, 1992, § 26a, BGBl. I S.1430, Ann. I (July 1, 1992) (explicitly designating safe third countries and including all EU states and EFTA states); GUY GOODWIN-GILL, THE REFUGEE IN INTERNATIONAL LAW 348, nn. 98-9 (2nd ed. 1998) (reporting on the Swiss and German laws on safe third countries, specifically that both Switzerland and Germany designated Bulgaria, the Czech Republic, Gambia, Ghana, Hungary, Poland, Romania, Senegal, and the Slovak Republic as safe); Sam Blay & Andreas Zimmermann, Current Development, Recent Changes in German Refugee Law: A Critical Assessment, 88 Am. J. Int’l L. 361 (Apr. 1994). Also note Germany’s importance as a specially interested state, see supra note 65.
278. See Refugee Act 1996, sec. 4(a), as amended; ECRE Country Report 2005, supra note 102, at 158 (“Designations of safe country of origin status were made, and continued in force in 2005 regarding Romania, Bulgaria, Croatia and South Africa; however Nigeria was not included in this safe country of origin list.”); European Council on Refugees and Exiles, ECRE Country Report 2003, supra note 39 (Ireland has only placed Nigerians in the expedited process under this ground).
279. See Law on Asylum Procedure of 1982 (border police were allowed to reject asylum applicants if the applicants had been “safe from persecution in another country”); GOODWIN-GILL, REFUGEE, supra note 277, at 348 nn.98-9.
Gibraltar, Iceland, San Marino, Norway, and Andorra), the UNHCR reports the following states as countries of origin of refugees: Austria (23 refugees from Austria in other states), Belgium (60), Bulgaria (3,311), Czech Republic (1,384), Estonia (262), France (101), Germany (129), Greece (92), Hungary (3,386), Italy (90), Latvia (662), Lithuania (466), the Netherlands (43), Poland (2,915), Portugal (32), Romania (5,306), Slovakia (342), Slovenia (52), Spain (41), and the United Kingdom (200). To this list of EU countries, we can also add the EFTA state of Switzerland (31). The Qualification Directive therefore appears to be an attempt to redefine the meaning of refugee and exclude individuals qualifying under the Refugee Convention. In addition to exclusion of other European nationals, Germany and Switzerland have both also designated Gambia, Ghana, and Senegal as “safe,” and excluded their nationals from qualifying as refugees accordingly. This act is strange in that the UNHCR names those states as the country of origin of 1,267, 5,060, and 15,896 refugees, respectively, so their designation as “safe” does not appear to be entirely accurate. The United Kingdom has designated all of the countries of the former USSR as safe counties. In sum, certain states have asserted the right to redefine the qualifications for being a refugee and limit the application of the Refugee Convention.

In addition to this policy, some states have also created an expedited processing mechanism for certain applicants for recognition of refugee status. The United Kingdom has adopted twenty-four-hour expedited processing for individuals from nationals of the following countries: Ghana, India, Nigeria, Pakistan, Poland, Romania, and Uganda. This expedited processing appears to have a normative effect in that applications reviewed through expedited processing resulted in almost a 100% refusal rate (5,735 refusals, 3 grants, and 996 pending). The result is that individual cases that are selected for expedited processing have been prejudged as failing. The number of refugees from those states of origin mentioned as reported by UNHCR has already been noted immediately above, but recall the numbers for India (20,463), Nigeria (13,902), Pakistan (31,858), and Uganda (228,959). Again, this policy appears inconsistent with the significant refugee flows originating from those states.

Courts have accepted the right of states to refuse to accept claims based on safe third country criteria, but have demanded that the claims be assessed on their merits rather than procedurally refused based on the stated country of transit (or on other criteria deemed to apply to the country, such as merely having signed the Refugee Convention or the ECHR). Additionally, the UNHCR

283. See Table 2, supra note 271.
284. Id.
287. See e.g., Bulgaria: In re Abdusalam, Case NB-46/2004 (Sofia City Ct., 3-G Dep’t 9 Feb.
Executive Committee has argued that states should not deny asylum simply because it may be sought elsewhere.\footnote{288} Clearly, the adoption of safe third country policies appears to be at odds with the UNHCR determination of state of origin of refugees, and the policies accordingly appear to violate the Refugee Convention when they do not permit substantive assessment of qualifications under it. In addition, some have argued that safe third country policies violate the ECHR.\footnote{289}

In sum, there appears to be a growing trend in favor of safe third country or country of origin policies. The states that impose such policies are not the most

\footnote{288. See UNHCR, Executive Committee Note No. 15 (XXX), ¶ (b)(iv). \textit{See also UNHCR, \textit{Note on International Protection}, supra note 236 (advising that often claimants are sent to a safe third country but that country fails to accept responsibility for the individual who is then returned to their country of origin); UNHCR, \textit{Conclusions on the International Protection of Refugees adopted by the Executive Committee of the iWHCR Programme} 33 (1980); Sztucki, J., \textit{The Conclusions on the International Protection of Refugees Adopted by the Executive Committee of the UNHCR Programme}, 1 INT’L J. REF. L. 285 (1989).}

\footnote{289. See Kathleen Marie Whitney, \textit{Does the European Convention on Human Rights Protect Refugees from “Safe” Countries?}, 26 GA. J. INT’L & COMP. L. 375 (Spr. 1997) (arguing that the ECHR prohibits states from expelling refugees to “safe” countries).}
representative ones, although a few specially interested states are included. Even if a rule of customary international law has been crystallized, it is important to note that those policies may not operate as procedural denial of claims. Instead, it appears that the safe country policies must operate at a level of holistic substantive review. Therefore, the definition of refugee as a person outside his country of nationality has evolved to cover only a person outside his country of nationality who has not arrived in any other safe country yet.

D. Prohibitions on Applying for Recognition of Refugee Status

States have sometimes imposed certain regulations that entirely prohibit the filing of an application for refugee status (or demand an automatic denial of an application) under certain circumstances. These provisions do not narrow the definition of refugee, rather they assess the broader scope of refugee status. They will nonetheless be considered here briefly. Note that in some of these situations, the burden of proof is placed on the applicant to establish that he has no disqualifying acts or conditions. 290

Mandatory denial of asylum or withholding of removal is required in some states if the applicant is a “terrorist” 291 or is a former Nazi, or “genocidaire.” 292 Although it may be acceptable under the Refugee Convention to refuse refugee status to those who have persecuted others, 293 committed a particularly serious crime, 294 or a serious non-political crime 295 or are otherwise a danger to state security, 296 these cases appear to refuse refugee status based on another status—that of simply being a “terrorist,” “Nazi,” or “genocidaire”—rather than being based on the specific culpable acts the person undertook. The recent judgment of the European Court of Justice in Bundesrepublik Deutschland v. B, Case C-57/09, 1990 E.C.R. and Bundesrepublik Deutschland v. D., Case C-101/09, 2010 E.C.R. (a joined case) seems to have disposed of any traction this rule may have had as a seed of customary international law when it held that “terrorist” status

290. See e.g., 8 C.F.R. §§ 208.13(c)(2)(ii), 1208.13(c)(2)(ii); Chay-Velasquez v. Ashcroft, 367 F.3d 751 (8th Cir. 2004); Ahmetovic v. Immigration & Naturalization Serv., 62 F.3d 48 (2d Cir. 1995) (holding that international legal obligations did not compel finding that the burden of proof was unlawful). Also note that the United States might be considered a specially interested state, see supra note 65.


292. See e.g., INA § 237(a)(4)(D).


alone cannot be a bar to an individualized refugee status determination when applying the usual criteria for such status under international law. 297

In addition, some states claim that they may impose a time limit within which the individual must apply for recognition of refugee status, although some states do not. 298 Austria permits applications up to three months after entry; 299 Belgium had a time limit policy but abolished it; 300 Germany and the Czech Republic require an application for recognition to be filed within forty-eight hours of admission to their respective territories; 301 Spain also had a time limit policy until the Spanish Supreme Court ruled that it was a violation of the obligation to consider claims under the Refugee Convention; 302 the law in Ukraine stipulated strict time limits for submission of applications for refugee status (five days for asylum seekers who crossed the Ukrainian borders legally and three days for those who crossed the borders illegally 303), but this law was revised to require an application simply “without delay”; 304 and the United States requires applicants to file within one year of arrival (unless there are changed circumstances). 305

Moreover, several states have introduced an obligation that an alien must not have committed any acts of fraud, misrepresentation, or other falsehood during their migration. Failure to comply can result in mandatory denial and expul-

297. See Case C-57/09, Bundesrepublik Deutschland v. B, 1990 E.C.R.; Case C-101/09, Bundesrepublik Deutschland v. D., 2010 E.C.R. Note that the ECJ initially makes reference to the Refugee Convention as the applicable standard under international law, although it also refers to EU Council Directive 2004/83/EC, discussed infra, at 54, for the substance of the refugee definition. However, the question in the case principally revolved around the need for an individualized assessment of the exclusionary clause, not the precise definition of refugee existing under international law. See id. at ¶¶ 67, 94.

298. See ECRE Country Report 2006, supra note 280 (e.g., there are no time limits for the submission of applications for refugee status on the territory of Belarus).

299. See Asylum Act 2005, Bundesgesetzblatt (“BGBl”) [Legal Gazette] Pt. I, No. 100/2005 (rendering claims “manifestly unfounded” if filed more than three months after entry, after a residence ban has been executable, or if the asylum-seeker is believed to have deceived the authorities with regard to his/her identity, nationality or submits false documents).

300. See Royal Decree of 3 Feb. 2005 (abolishing time limit of 8 days following arrival, however application after 8 days must be justified or it can still be declared inadmissible).


302. See Decision of 23 June 2004 (Sup. Ct., Sp.) (holding that the time limit imposed by Spanish Regulation, R.D. 203/1995, art. 5.6-Asylum, could not be applied to presume the application was manifestly unfounded and must assess the merits).


304. See id. (“Despite the fact that the time limits for submitting an asylum application were removed from Ukrainian refugee legislation in 2005, and replaced with the term ‘without delay’ – this is in practice interpreted literally and can still be given as a reason for a refusal to accept an asylum application on formal grounds.”).

305. See e.g., INA § 208(a)(2)(B), (D); 8 U.S.C. § 1158(a)(2)(B), (D); Joaquin-Porras v. Gonzalez, 435 F.3d 172 (2d Cir. 2006).
Among these states are Austria, Bulgaria, and Finland, although there are examples of state practice with the opposite results. For instance, falsehood alone could not be a reason for refusal to examine the application in Ireland. There does not appear to be any basis in the Refugee Convention for this treatment. If the person qualifies as a refugee under the Convention, then he may not be returned regardless of falsehood.

Last, a variety of miscellaneous practices exist. Some states impose a mandatory denial of asylum if a previous asylum application was denied and there is no proof of changed circumstances, in others there is the isolated practice of refusing to apply Article 31(1) of the Refugee Convention that requires states to overlook illegal entry or presence for asylum seekers, so applications may be refused on that basis; and in even others the issuance of a visitor visa—required for individuals of nationality of the major refugee producing states—can be conditional upon a guarantee that the visiting person will not apply for a permanent stay, including asylum, again making it possible to refuse refugee status on that basis.

These are interesting prohibitions on lodging refugee applications, but they are different from one another and rather diffuse. None appear to attract the widespread and consistent practice sufficient to form a rule of customary international law. The only rules among these practices that could arguably form the basis for a rule of customary international law are the prohibition on applications by terrorists and other similarly designated persons, the rules on time limits for applications, and the rules against fraud. However, the prohibitions on applications by terrorists and similar individuals are not widespread, and in any event they are most likely not permissible as a procedural bar. The rules on time limits are so varied in their specific length and the countries with fraud rules are so un-

306. See Asylum Act 2005, (rendering claims “manifestly unfounded” if filed more than three months after entry, after a residence ban has been executable, or if the asylum-seeker is believed to have deceived the authorities with regard to his/her identity, nationality or submits false documents).

307. See Law on Asylum and Refugees, art. 17(2) (2002), as amended Apr. 2005 (allowing revocation of refugee or humanitarian status in cases where the refugee has used a false identity or has concealed material information related to his/her case).

308. See Aliens Act, § 108 (2004) (“if the applicant has, when applying for asylum deliberately or knowingly given false information which has affected the outcome of the decision, or concealed a fact that would have affected the outcome of the decision.”); ECRE Country Report 2005, supra note 102 (“There have for example been cases of resettlement in which officials have discovered that the person had given misleading information prior to selection as a quota refugee to Finland.”).

309. See Ref. No. 22, 2006 (Ref. Appl. Trib. 2006) (applicant from Zimbabwe changed story at the appeal stage and was found to have failed to tell the truth at Questionnaire and Interview stage).

310. See e.g., INA § 208(a)(2)(C), (D); 8 U.S.C. § 1158(a)(2)(C), (D); § 1129a(c)(7)(C)(ii); Zheng v. Mukasey, 509 F.3d. 869 (8th Cir. 2007).


representative that they are unlikely to be considered as rules of customary international law. In addition, in some of those cases the states have already moved to suspend or modify their time limit requirements, sometimes recognizing their non-compliance with international law in the process.

E. Recognition of Refusals by Other States

Some states also assert a right to expel an individual claiming refugee status without assessing the person’s qualification, based on a notion similar to *res judicata*, i.e., by recognizing the prior decision of another state as to a person’s refugee status.\(^{313}\) Additionally, if no prior status determination took place and the individual was found to likely be a refugee, then EU states comply with the Council Directive ordering the person to be returned to the state where the status determination should have taken place, as discussed above.\(^{314}\) If he was not found to be a refugee in the status determination, then he may be returned to his state of origin without further inquiry.\(^{315}\) These rules also do not appear to have become rules of customary international law because the practice is isolated to EU and EEA states. This practice must be viewed in the context of the EU system of handling refugee applications and safe third country of origin policies. Given that this practice is relatively limited, it seems a hard argument to consider that it amounts to customary international law.

F. Manifestly Unfounded Applications

Another trend in refugee claims is for governments to either refuse to consider applications or to consider them in an expedited fashion by characterizing the claim as “manifestly unfounded” based on specific conditions present in the application. States that have adopted this policy include Australia,\(^{316}\) Austria,\(^{317}\) Belgium: Amendment of Aliens Act (Sept. 1, 2004), art. 8 (implementing EU Dir. 2001/40/EC on mutual recognition of expulsion decisions (May 28, 2001)) (Minister of Home Affairs can recognise an expulsion decision taken by an administrative authority of another EU Member State bound by the Directive and expel the person based on a danger to public security or non-compliance with the legislation on entry and residence); Switzerland: (creating a ground for inadmissibility to the state due to a negative asylum decision from an EU or EEA country, except for subsequent persecution), reported in ECRE Country Report 2004, supra note 5; the United Kingdom: Yogathas v. Sec’y St. Home Dep’t, [2002] U.K.H.L. 36 (H. Lords, Oct. 17, 2002).

313. See Granting Refugee Status Dir., supra note 282. See also, e.g., Belgium: Amendment of Aliens Act (Sept. 1, 2004), art. 8 (implementing EU Dir. 2001/40/EC on mutual recognition of expulsion decisions (May 28, 2001)) (Minister of Home Affairs can recognise an expulsion decision taken by an administrative authority of another EU Member State bound by the Directive and expel the person based on a danger to public security or non-compliance with the legislation on entry and residence); Switzerland: (creating a ground for inadmissibility to the state due to a negative asylum decision from an EU or EEA country, except for subsequent persecution), reported in ECRE Country Report 2004, supra note 5; the United Kingdom: Yogathas v. Sec’y St. Home Dep’t, [2002] U.K.H.L. 36 (H. Lords, Oct. 17, 2002).

314. See Council Regulation (EC) No. 343/2003 (Feb. 18, 2003) (establishing the criteria and mechanisms for determining the Member State responsible for examining an asylum application lodged in one of the Member States by a third-country national; this regulation, known as the “Dublin Convention”, integrates the principles of the 1990 Dublin Convention into the community context).

315. See Granting Refugee Status Dir., supra note 282.

316. See Somaghi v. Min. Immigr. Local Gov’ts, & Ethnic Aff’rs, (1991) 31 F.C.R. 100 (Fed. Ct., Austl.) (the court was not persuaded to grant refugee status to “… a person whose sole ground for refugee status consists of his own actions in his country of residence designed solely to establish the circumstances that may give rise to his persecution if he should return to the country of origin.”).
Bulgaria, Hungary, the Netherlands, Norway and the United States. The Qualification Directive also permits states to refuse claims that are “manifestly unfounded.” Under the Qualification Directive, states may inquire whether actions by the applicant that give rise to a well-founded fear of persecution were taken merely as a pretext for claiming refugee status.

However, the manifestly unfounded analysis is arguably in violation of the Refugee Convention. The terms of the Convention make no mention of good faith or bad faith actions on the part of the refugee that lead to the acquisition of refugee status. The Refugee Convention merely seeks to prevent removal to situations of persecution (aside from clearly articulated cases where the claimant is considered undeserving of refugee status). Without the discovery of conclusive indications that any practice and opinio juris is attempting to specifically reverse the treaty, we cannot agree that a new, contradictory rule has emerged under customary international law. Recalling the language from Delimitations of the Continental Shelf, 18 R.I.A.A. 3, Award, ¶ 47, (June 30, 1977):

“[T]he Court recognises both the importance of the evolution of the law of the sea which is now in progress and the possibility that a development in customary law may, under certain conditions, evidence the assent of the States concerned to the modification, or even termination, of previously existing treaty rights and obligations. . . . [O]nly the most conclusive indications of the intention of the parties to the [treaty] to regard it as terminated could warrant this Court in treating it as obsolete and inapplicable . . . .”

Therefore, it will be important to find conclusive indications to support the ex-

317. See Asylum Act 2005 (providing for subsidiary protection rendering claims manifestly unfounded or “disallowed if filed more than three months after entry, after a residence ban has been executable, or if the asylum-seeker is believed to have deceived the authorities with regard to his/her identity, nationality or submits false documents).

318. See Case No. N 3229/2004-Budali (Sofia City Ct., 3-B Dep’t 28 Mar. 2005) (holding that claim was not contradicted by information regarding Algeria’s human rights record, so it was not manifestly unfounded).


322. The policies sometimes include actual refoulement. See Sale v. Haitian Ctrs. Council, 509 U.S. 155 (1993) (holding that the US government could return Haitians directly to Haiti, without access to a refugee determination, if the Haitians were interdicted on the high seas).

323. See Granting Refugee Status Dir., supra note 282.

324. See Qualif. Dir., supra note 103, at art. 4(3)(d).

325. Delimitations of the Continental Shelf (U.K./Fr.), Award, ¶ 47, 18 R.I.A.A. 3 (June 30, 1977).
There appear to be two different applications of this policy. The first consists of cases where the individual would not qualify under the law even if the facts could be proven (compare to “summary judgment”). The second consists of cases where there is merely a lack of evidence. Courts have found that the manifestly unfounded policy cannot be applied in situations where the claim is manifestly unfounded only due to a lack of evidence.326

As for the other area of application—the “summary judgment” processing—this policy has been criticized, especially when the designation of “manifestly unfounded” appears to be very liberally used.327 That being said, the UNHCR appears to have accepted the manifestly unfounded assessment practice provided that it is not applied to deny legitimate refugees the status they deserve.328 Providing an expedited “summary judgment” processing may very well assist legitimate applicants by directing further resources to careful evaluation and processing of legitimate applications. The acceptance of this practice by the UNHCR specifically would appear to be fairly conclusive so that, to the degree that the policy contradicts the Refugee Convention, it has reversed it through customary international law.

Thus it would appear that the presumption of refugee status, and the rights accrued to those with presumptive refugee status, do not apply in cases where the person clearly does not qualify. In cases where the person clearly does not qualify, the state need not consider that the person might be a de facto refugee in its treatment of the person. The manifestly unfounded policies appear to be regarded as unlawful where they operate to refuse claims based purely on evidentiary concerns, but not where they operate to efficiently dismiss claims that could never succeed on the law. In the latter, during the pendency of the mani-

326. See Case NB-604/2004 - Dai Dzyu Huang (Sofia City Ct., Dep’t 3-G June 1, 2004) (holding that had the claimant given coherent and plausible statements and made a genuine effort to substantiate their story, the application should be referred for further consideration as the manifestly unfounded criteria should not be legally applied purely on the basis of lack of evidence (the benefit of the doubt principle)).

327. See V AJN & NJCM v. Neth., Ct. Appl., The Hague, ILDC 143 (NL 2002) (holding that mandatory detention policy at the Application Centre (such as at Schiphol Airport) during the accelerated 48-hour procedure for “manifestly unfounded” claims violated Art. 5 of the ECHR because it restricted freedom of movement by demanding that individuals who leave automatically withdraw their asylum application; also observing that the UNHCR had criticized the increasingly liberal interpretation of “manifestly unfounded”); Chiara Martini, Is the EU abandoning non-refoulement?, 25 FORCED MIGR. REV. 62 (May 2006) (reporting that under the “accelerated procedures” provision, a wide range of asylum claims – more than 80% according to Amnesty International – are arbitrarily judged to be “manifestly unfounded”).

328. Follow-up on Earlier Conclusions of the Sub-Committee on the Determination of Refugee Status, inter alia, with Reference to the Role of UNHCR in National Refugee Status Determination Procedure Date: 3 Sep. 1982 International Protection (SCIP), U.N. Doc. EC/SCP/22/Rev.1 (expressing acceptance of measures for manifestly unfounded or abusive applications for refugee status but expressing concern that genuine applications not be overlooked).
festy unfounded claim, the state may not even need to consider the applicant as a presumptive refugee.

G. Diplomatic Assurances

The topic of diplomatic assurances remains controversial and the United Nations Special Rapporteur on Torture\(^\text{329}\) and others\(^\text{330}\) have argued that diplomatic assurances cannot relieve a state of its non-refoulement obligation. Nonetheless, the UK Government has concluded MOU with Jordan, Libya, and Lebanon in order to provide blanket assurances.\(^\text{331}\) The Council of Europe Commissioner for Human Rights argued that “[t]he weakness inherent in the practice of diplomatic assurances lies in the fact that where there is a need for such assurances, there is clearly an acknowledged risk of torture or ill-treatment.”\(^\text{332}\) This is echoed by the UN Independent Expert on the Protection of Human Rights and Fundamental Freedoms while Countering Terrorism: “[T]he mere fact that such assurances are sought is arguably a tacit admission by the sending State that the transferred person is indeed at risk of being tortured or ill-treated.”\(^\text{333}\)

The Human Rights Committee established the standard that assurances may be accepted provided the state “institute[s] credible mechanisms for ensuring compliance by the receiving State with these assurances from the moment of ex-

\(^{329}\) See Theo van Boven, U.N. Special Rapporteur on Torture, Report to the General Assembly, U.N. Doc. A/60/316, ¶ 51 (2005) (“It is the view of the Special Rapporteur that diplomatic assurances are unreliable and ineffective in the protection against torture and ill-treatment.”); Committee Against Torture, U.N. Doc. CAT/C/CR/33/3, ¶ 4 (Nov. 2004) (expressing concern at the United Kingdom’s reliance on diplomatic assurances to refoul); UNHCR, Executive Committee Conclusion No. 30 (XXXIV) – 1983, The Problem of Manifestly Unfounded or Abusive Applications for Refugee Status or Asylums (“recognized the substantive character of a decision that an application for refugee status is manifestly unfounded or abusive, the grave consequences of an erroneous determination for the applicant and the resulting need for such a decision to be accompanied by appropriate procedural guarantees.”).


The Special Rapporteur on Torture reached a similar conclusion when he stated that assurances would be acceptable where “the receiving State has provided an unequivocal guarantee to the extraditing authorities that the persons concerned will not be subjected to torture or any other form of ill-treatment, and that a system for monitoring the treatment of such persons has been put into place to ensure that they are treated with full respect for their human dignity.”\textsuperscript{335} The successor to the Special Rapporteur, however, concluded that “[i]n the situation that there’s a country where there’s a systematic practice of torture, no such assurances would be possible.”\textsuperscript{336} This is similar to the holding in \textit{Agiza v. Swed.}, Comm. No. 233/2003, U.N. Doc. CAT/C/34/D/233/2003, \textsuperscript{¶} 13.4 (Comm. Ag. Torture, May 24, 2005), where the Committee against Torture found that the assurances were insufficient to permit expulsion because:

At the outset [. . .] it was known, or should have been known, to the State party’s authorities at the time of the complainant’s removal that Egypt resorted to consistent and widespread use of torture against detainees, and that the risk of such treatment was particularly high in the case of detainees held for political and security reasons. [The CAT also noted that the complainant was implicated in terrorist activities with national security implications.] In the Committee’s view, the natural conclusion from these combined elements, that is, that the complainant was at a real risk of torture in Egypt in the event of expulsion [. . .] The procurement of diplomatic assurances, which, moreover, provided no mechanism for their enforcement, did not suffice to protect against this manifest risk.\textsuperscript{337}


A distinction may be drawn between assurances given by a state that it will not apply the death penalty (through a legal process) and assurances by a state that it will not resort to torture (an illegal process). We would signal the difficulty in relying too heavily on assurances by a state that it will refrain from torture in the future when it has engaged in illegal torture or allowed others to do so on its territory in the past . . . The former [death penalty] are easier to monitor and generally more reliable than the latter [torture].\textsuperscript{338}


It appears that assurances may be acceptable but only where they are genuine. Courts have blocked expulsions in such cases where the assurances were not credible.339 Perhaps there is a presumption against their being genuine inherent in the fact that they are being requested, i.e., without credible concerns for the risk to the person being return, assurances would not be requested. In any event, a state proposing to expel a person cannot rely on assurances from a state with a record of violations. Additionally, the state expelling a person to a state from which assurances were requested must institute an effective monitoring capacity.340

Another situation in which a state could not issue assurances sufficient to relieve the expelling state of its non-refoulement obligations is where the individuals responsible for the potential persecution, torture, or other inhumane treatment are affiliated with non-state agents or rogue state agents that the state is unable or unwilling to control.341 This, however, falls within the usual test of whether there are substantial grounds for believing that there is a real risk of the unacceptable treatment.342

As for the definition of refugee under customary international law, it appears that a state can defeat a claim of persecution (or other grounds of refugee qualification) by issuing assurances that such persecution (or other qualifying acts) will not occur. The quality of the assurances must then be assessed for their reliability, but might result in acceptance and refoulement of the individual to the state. Therefore, the definition of refugee has been modified from the exist-
ing flight from a qualifying situation to flight from a qualifying situation where the state does not or cannot offer genuine assurances of the situation not occurring.

V. CONCLUSION

This Article has examined the various ways in which customary international law is changing the definition of refugee in international law. It has attempted to balance the competing demands of international law, the state-centered and human-centered interests, in order to reach what is hopefully a convincing conclusion about the state of contemporary customary international refugee law.

First, this Article examined the Refugee Convention itself and its current interpretation. The initial and overriding conclusion to be drawn from the interpretation of the Convention is that it is usually interpreted with an emphasis on the teleological method, possibly in line with the intention of the drafters. This interpretive technique was applied in cases of the inclusion/exclusion and cessation provisions of the Convention. In some cases the interpretive technique has been used so as to verge on amending the treaty’s explicit terms.

Second, this Article examined customary international law as applied in cases of refugees or individuals in refugee-like situations. Many scholars have concluded that there is no definition of refugee under customary international law. This Article refutes this conclusion by drawing on the extensive practice of states. As a preliminary matter, this Article identified the specially interested or specially affected states in matters of refugee law. This aspect of the formation of customary international law appears to be largely omitted in most analyses of customary international law on point. It was submitted that specially interested states in this case are those states that experience inward refugee flows, as measured by statistical studies of such flows by the UNHCR. With specially interested states identified, the analysis of the formation of customary international law on the definition of refugee can be appreciated in an entirely new light, especially the influence of certain regional instruments. This finding suggests that under customary international law, the definition of refugee may include: individuals persecuted on the basis of gender or sexual orientation, individuals fleeing from external aggression, occupation or other serious disturbances of public order, possibly including massive violations of human rights and/or torture, or even the imposition of the death penalty.

Continuing with the analysis of customary international law, this Article considered the influence of the widespread practice of “subsidiary protection,” both under international law and municipal law. In many instances, however, this examination suggested that the subsidiary protection was enacted partly with the purpose of insulating the definition of refugee from further development under customary international law. Nevertheless, although the formal cate-
gories were in most cases retained, the consideration and actual treatment of individuals shows, at the minimum, that there may be a growing customary international legal obligation to provide for subsidiary protection and, at most, that subsidiary protection is now defining the new outer parameters of the definition of the refugee.

This Article has also considered, albeit cautiously, the contribution of the practice of international organizations to the definition of refugee. It still does not appear that contemporary international law has specifically accepted the practice of international organizations as contributing to the formation of customary international law, that is, as opposed to the practice of states within and through international organizations, where there is much more acceptance. The practice of international organizations suggests, again, a growing norm of subsidiary protection, if a conservative approach is taken to the appreciation of the formation of customary international law. Taking a more aggressive approach, there might even be something more significant happening. However, the overall practice appears a little too inconsistent for finding a new norm.

Last, this Article has had to look at the other side of the formation of customary international law: not the expansion of the definition, but the contraction. There are many instances of state practice attempting to narrow the definition, although the great majority of them are inconsistent, singular, or clearly perceived to be violations of refugee law rather than evolving customary law. One of the more significant developments in the narrowing of the definition is the growth of the concept of internal flight or relocation alternatives to regions within the state that are considered safe. From a comparative study of the practice of states on relocation alternatives, the law appears to permit, at a minimum, that individuals may be returned to a different region of a state where there is genuine access to meaningful protection, not illusory or unpredictable protection (e.g., the House of Lords “first approach,” discussed in Januzi v. Sec’y St. Home Dep’t, [2006] U.K.H.L. 5 (H. Lords Feb. 15, 2006)). However, some states adopt the more restrictive “second approach” (that conditions in the proposed region of the state not be so different from the conditions of non-persecuted persons in the original region), and that approach may eventually crystallize into customary international law, though it does not appear to have done so yet. Safe third country and country of origin designation policies are also becoming increasingly popular; however, these policies do not appear to have moved into customary international law. First, they directly violate the Refugee Convention and therefore any customary international legal analysis will need to see evidence of clear intent to reverse a well-established treaty norm. Second, those blanket rules have failed in some jurisdictions for precisely that reason of their failure to consider cases on their merits, thus the states themselves have determined that their proposed policies were in violation of the norms on refugees, preventing the contrary norm from emerging in customary international law. A

number of other potential narrowing measures were each in turn found to have not crystallized into customary international law, including practices of automatic refusals of manifestly unfounded applications. The final practice that was considered was the reliance on diplomatic assurances that the acts against the person would not occur. This practice is not provided in the Refugee Convention as an exception to non-refoulement, but it appears to have crystallized in customary international law. That being said, the practice comes with the express condition that the assurances provide a credible basis for ensuring that the prohibited acts not occur. Where states have a systematic practice of engaging in the prescribed acts, assurances cannot be used as an exception to non-refoulement.

Therefore, the evolving definition of a refugee under conventional and customary international law is:

I. A person who, owing to a well-founded fear of being subjected to a situation of

   (1) persecution for reasons of race, religion, nationality, membership of a particular social group or political opinion;
   (2) torture, inhuman or degrading treatment and/or punishment; and/or
   (3) a threat to life, security or liberty due to events seriously disturbing public order; that

       (a) is so widespread that it exists in all parts of the state of origin where the person could flee and also exists in every state the person reached upon leaving his state of origin; and
       (b) is unable to be cured by credible, reliable and genuine assurances offered by the state of origin, and any other state that the individual previously reached, of the situation not occurring to that individual;

   and

II. Such person is outside the country of his or her nationality of former habitual residence and is unable or, owing to such fear, unwilling to avail himself of the protection of that country or return to it.

Based on the above, there is customary international law in the field of refugee law. Finding that such law exists does not, however, necessarily mean that refuge is only available to an expanded group of persons. International law is not always so kind. Instead, it means that protection is available for more persons, but also that that protection is also limited by additional rules. In the field of refugee law, there is usual balance between state freedom of action and state limitation on action, with states demanding increasingly liberal moral standards from themselves, but showing increasing reluctance to live by those standards. People in need of protection fall somewhere between these two extremes.
2012

The "Right to Be Forgotten": Reconciling EU and US Perspectives

Steven C. Bennett

Recommended Citation
Available at: http://scholarship.law.berkeley.edu/bjil/vol30/iss1/4
The “Right to Be Forgotten”:
Reconciling EU and US Perspectives

By
Steven C. Bennett*

Recent developments in the European Union (EU) have highlighted the potential for the development of a “right to be forgotten.” For United States (US) companies, especially those operating on the Internet, the development and enforcement of such a right could prove to be quite problematic. This Article outlines the practical implications of such a right, pointing the way toward possibilities for reconciliation of US and EU views on the application of a right to be forgotten.

I. RECENT EU DEVELOPMENTS

Viviane Reding was recently accepted to a position as European Commissioner for Justice, Fundamental Rights and Citizenship. She previously served as Commissioner for Information Society and Media. In those roles, she has served as an important spokesperson and advocate for the development of EU privacy protection.¹ At an American Chamber of Commerce gathering in June 2010, Reding suggested that her “paramount goal” in her new position is to “ensure that people have a high level of protection and control over their personal infor-

---

¹ The Author is a partner at Jones Day in New York and teaches Conflicts of Law at Hofstra Law School. The views expressed are solely those of the author and should not be attributed to the author’s firm or its clients.

Reding emphasized that “[i]nternet users must have effective control of what they put online and be able to correct, withdraw or delete it at will,” labeling this right of control “a right to be forgotten.” Reding, moreover, noted that “[a]ll companies that operate in the European Union must abide by our high standards of data protection and privacy.”

Later in 2010, an EU press release announced that the European Union planned to propose “a new general legal framework for the protection of personal data in the European Union covering data processing operations in all sectors and policies of the European Union.” This “comprehensive” new legal framework would be subject to negotiation between the European Parliament and the European Council of Ministers. The EU announcement specifically mentioned the “right to be forgotten,” described as the right of individuals to “have their data fully removed when it is no longer needed for the purposes for which it was collected.” A more comprehensive EU white paper, released simultaneously, referenced the same concept and outlined the EU plan for modernization of EU privacy law to address “globalization and new technologies. . . .” Similar EU

3. Id. Reding noted, however, that the area of data protection and privacy “needs clarity, not red tape.” She suggested that “industry self-regulation” could “work well” in this area. Id.; see generally Viviane Reding, The Upcoming Data Protection Reform for the European Union, 1 INT’L DATA PRIV. L. 3 (2011) (“Rapid technological developments and globalization have profoundly changed the world around us, and brought new challenges to the protection of personal data. . . .

Globalization has seen an increasing role of third countries relating to data protection, and has also led to a steady increase in the processing of the personal data of Europeans by companies and public authorities outside the European Union.”), available at http://idpl.oxfordjournals.org/content/1/1/3.full.
6. See Data Protection Reform – Frequently Asked Questions, supra note 4 (stating that the announcement, moreover, pointed specifically at social networking site information: “People who want to delete profiles on social networking sites should be able to rely on the service provider to remove personal data, such as photos, completely”).
7. See Comprehensive Approach on Personal Data Protection, supra note 5, at 5. The EU white paper suggested a need to examine ways of “clarifying the so-called ‘right to be forgotten,’ i.e.
explanations of the “right to be forgotten” have followed. The European Union, moreover, has emphasized that “privacy standards for European citizens should apply independently of the area of the world in which their data is being processed.”

Recent developments in Spain and Italy have amplified public discussion on the right to be forgotten. In early 2011, Spanish data protection authorities demanded that Google remove links to online news articles on grounds that the articles contained out-of-date information which infringed on the privacy of Spanish citizens. At about the same time, Italy announced that it would regu-
late Internet content sites as if they were television broadcasters and would impose an obligation to publish “corrections” to libelous content. These developments followed a case in Germany in 2009, where two murderers, who had completed their prison sentences, sued to remove references to their crime from Internet postings. In early 2010, moreover, an Italian court found several Google executives guilty of violating Italian privacy law by permitting a video of abuse of a disabled boy to persist on its online video service. These developments suggested the broad uses to which a comprehensive “right to be forgotten” might be implemented.

II. US RESPONSES

Many US commentators, confronted with the suggestion of development of a “right to be forgotten,” accused EU regulators of “foggy thinking” incon-


See Peter Fleischer, Foggy Thinking About The Right To Oblivion, Mar. 9, 2011 (Google global privacy counsel suggests that right to be forgotten may be “used to justify censorship”), available at http://peterfleischer.blogspot.com/2011/03/foggy-thinking-about-right-to-oblivion.html; see also Tessa Mayes, We Have No Right To Be Forgotten Online, Mar. 19, 2011 (UK commentator
sistent with fundamental US values (such as freedom of expression and of the press). A representative of Facebook, for example, suggested that the EU approach was to “shoot the messenger,” in that the “source of the content” rather than the “places where the content is shared,” should be the focus of any efforts to promote privacy controls. Others suggested that the EU approach could create a property right in information (which otherwise does not exist), and suggests “[t]he right to be forgotten is a figment of our imaginations;” “a right to be forgotten is about extreme withdrawal, and in its worst guise can be an antisocial, nihilistic act”), available at http://www.guardian.co.uk/commentisfree/libertycentral/2011/mar/18/forgotten-online-european-union-law-internet.

15. See Timothy Ryan, The Right To Be Forgotten: Questioning The Nature Of Online Privacy, May 2, 2011 (“Sometimes, the right to information ought to outweigh the right to privacy. What incentive will there ever be for a journalist to rake muck if the information can simply be taken down upon request?”), available at http://www.psfk.com/2011/05/the-right-to-be-forgotten-questioning-the-nature-of-online-privacy.html; Spanish Claim “Right To Be Forgotten” On Web, Apr. 20, 2011 (“In the United States, we have a very strong tradition of free speech [and] freedom of expression. We would strongly caution against any interpretation of the right to be forgotten that infringes upon that.”) (quoting Justin Brookman, Center for Democracy and Technology, Privacy Project), available at http://www.cbsnews.com/2100-205_162-20055718.html; Jennifer L. Saunders, Across Jurisdictions And Web Domains, Questions Of Privacy And Online Anonymity Persist, Apr. 15, 2011 (noting “tug-of-war” involving “privacy-related question of accountability when one individual’s postings defame or expose personal information about another”), available at https://www.privacyassociation.org/publications/2011_04_15_across_jurisdictions_and_web_domains_questions_of_privacy_and; L. Gordon Crovitz, Forget Any “Right To Be Forgotten”, Nov. 19, 2010 (“Any regulation to keep personal information confidential quickly runs up against other rights, such as free speech, and many privileges, from free Web search to free email.”), available at http://abletueau.wordpress.com/2010/11/19/forget-any-right-to-be-forgotten/; L. Gordon Crovitz, Get Used To It—The Internet Is Forever, Nov. 10, 2010 (“Regulators have no reason to dictate one right answer to these balancing acts among interests that consumers are fully capable of making for themselves.”).

16. See Kelly Fiveash, Facebook Tells Privacy Advocates Not To “Shoot The Messenger:” You Have No Right To Be Forgotten, Argues Big IT, Mar. 23, 2011, http://www.theregister.co.uk/2011/03/23/facebook_shoot_messenger/; Pichayada Promchertchoo, Facebook Questions EU “Right To Be Forgotten”, Mar. 23, 2011 (right to be forgotten is “opposite” of what most users want, according to Richard Allan of Facebook), http://www.techweekeurope.co.uk/news/facebook-questions-eu-right-to-be-forgotten-24509; see also Jason Walsh, When it Comes To Facebook, EU Defends The “Right To Disappear”, Apr. 6, 2011 (“criticism is coming from American technology companies and some advocates who come down on the side of freedom of expression online, over the right to privacy;” “the typical US response is to encourage more personal responsibility and education of users”), http://www.csmonitor.com/World/Europe/2011/0406/When-it-comes-to-Facebook-EU-defends-the-right-to-disappear; Ron Miller, We May Not Have A “Right To Be Forgotten” Online, Mar. 14, 2011 (“There’s really no way to remove every trace of anyone on the Internet, even if there were a law in place requiring it.”), http://www.intemetrevolution.com/author.asp?section_id=1047&doc_id=204757.

17. See Larry Downes, Europe Reimagines Orwell’s Memory Hole, Nov. 16, 2010, (“Information isn’t property, at least not as understood by our industrial-age legal system or popular metaphors of ownership. Information, from an economic standpoint, is a virtual good. It can be ‘possessed’ and used by everyone at the same time. . . And, whether the law says so or not, it can’t be repossessed, put back in the safety deposit box, buried at sea, or ‘devoured by the flames,[]’”), http://techliberation.com/2010/11/16/europe-reimagines-orwells-memory-hole/; Adam Thierer, An Internet Eraser Button To Protect Privacy? Unwise & Probably Impossible, Apr. 19, 2011, (suggest-
duce a “bureaucratic nightmare,”\textsuperscript{18} which might interfere with “business demands” for data.\textsuperscript{19}

Yet, some US commentators appeared more receptive to at least a limited version of the right to be forgotten.\textsuperscript{20} Some, for example, suggested that children
Bennett: The "Right to Be Forgotten": Reconciling EU and US Perspectives

THE “RIGHT TO BE FORGOTTEN” 167

should have the ability to erase information posted improvidently, out of youthful lack of judgment. 21 Others noted that the concept of “data minimization” (a form of the right to be forgotten) has long been a central element of “fair information practices” under various US laws, and could be expanded. 22 Still others noted that the concept of “forgive and forget” embodies a fundamental human value, 23 and that US law (bankruptcy, credit reporting and criminal law, among others) actually does recognize at least some elements of a “right to be forgot-


22. See Justin Brockman, Europe Revisiting Privacy Law is Opportunity, not Catastrophe, CTR. FOR DEMOCRACY & TECH. (Nov. 12, 2010), available at http://cdt.org/blogs/justin-brockman/europe-revisiting-privacy-laws-opportunity-not-catastrophe (“The concept of data minimization—including deleting data no longer necessary to achieve a consumer purpose—has been a bedrock concept of Fair Information Practices . . . for years.”). A guide recently released by the Department of Homeland Security (“DHS”), for example, aims to help “federal privacy practitioners” understand how to build a “privacy culture.” Department of Health Services Privacy Office, Guide to Implementing Privacy 3 (Jun. 3, 2010), available at http://www.cio.gov/documents/DHS-Privacy-Office-Guide_June-2010.pdf. The DHS guide identifies, as an essential fair information practice, the rule that “DHS should only collect PII [personally identified information] that is directly relevant and necessary to accomplish the specified purpose(s) and only retain PII for as long as is necessary to fulfill the specified purpose(s).” The Fair Credit Reporting Act (“FCRA”) also provides a model for regulation to minimize the use of out-of-date and inaccurate information. Fair Credit Reporting Act, 15 U.S.C. § 1681. See Protecting Privacy in Online Identity: A Review of the Letter and Spirit of the Fair Credit Reporting Act’s Application to Identity Providers, CTR. FOR DEMOCRACY & TECH. (Feb. 26, 2010), available at http://cdt.org/policy/protecting-privacy-online-identity-review-letter-and-spirit-fair-credit-reporting-act%E2%80%99s-application-n (noting that the FCRA “is one source of some of the necessary protections and may already apply to entities providing or using identity-related services”).

23. See Seaton Daly, Le Droit a L’oubli—Can We Achieve “Oblivion” on the Internet? EMERGING BUS. ADVOCATE (Mar. 15, 2011), http://emergingbusinessadvocate.wordpress.com/2011/03/15/le-droit-a-loubli-can-we-achieve-oblivion-on-the-internet/ (“The debate over privacy has more to do with the universal right to control our image than anything else. . . . Who we perceive ourselves to be, and what people really see us for—that’s the debate, and citizens are mandating that we get some of that control back [that] companies have taken from us.”); Jeffrey Rosen, The Web Means the End of Forgetting, N.Y. TIMES (July 21, 2010), http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html?pagewanted=all. (“[S]ome legal scholars have begun imagining new laws that could allow people to correct, or escape from, the reputation scores that may govern our personal and professional interactions in the future.”); VIKTOR MAYER-SCHÖNBERGER, DELETE: THE VIRTUE OF FORGETTING IN THE DIGITAL AGE (2009) (noting without some form of forgetting, forgiveness may be difficult); DAVID SHENK, DATA SMOG: SURVIVING THE INFORMATION GLUT (1997) (noting risks of “information fatigue syndrome” in modern social and technical milieu); Daniel J. Solove, A Taxonomy of Privacy, 154 U. PA. L. REV. 477, 532 (2006) (“People grow and change, and disclosures of information from their past can inhibit their ability to reform their behavior, to have a second chance, or to alter their life’s direction.”).
The breadth of US reactions to developments in the European Union regarding the right to be forgotten suggests a need for serious thinking about the means available to reconcile US and EU views on the subject. The remainder of this Article focuses on both substantive and procedural mechanisms to effect such reconciliation.

III.
RECONCILING SUBSTANTIVE EU AND US VIEWS

The United States and the European Union have traditionally held widely differing views on data privacy.25 To some degree, those differences remain unresolved.26 The European Union generally adheres to a high degree of government involvement in protection of this fundamental right.27 US privacy law has,


27. See James Gordley, When is the Use of Foreign Law Possible? A Hard Case: The Protection of Privacy in Europe and the United States, 67 LA. L. REV. 1073 (2007) (differences between the European Union and the United States may turn in part on degree to which authorities view in-

http://scholarship.law.berkeley.edu/bjil/vol30/iss1/4
by contrast, largely developed in a “patchwork,” with a “reactive” array of state and federal statutes and common law doctrines. The United States, moreover, has traditionally emphasized freedom of expression over privacy, as a fundamental value.

Some suggest that a right to be forgotten simply cannot exist in the United States. Yet, a brief review of developments in US privacy law suggests the extent to which EU and US notions of the right to be forgotten might be reconciled. More than 120 years ago, the seminal Warren and Brandeis article on privacy focused on the degree to which “unseemly gossip,” coupled with “modern enterprise and invention” (including the telephone and photography) had contributed to “mental pain and distress,” caused by invasions of privacy. Early cases thereafter suggested at least the possibility of claims for privacy invasion based on harmful reference to out-of-date information. The Restatement (Second) of Torts, moreover, expressly recognized a potential tort claim for “publicity given to private life.”


29. See IAN BALLON, E-COMMERCE AND INTERNET LAW § 26.01 (2011) (“The US approach to data privacy is very different from that of some of our major trading partners, like the EU” and noting that “the United States has placed greater emphasis on free speech and access to information”).


33. See RESTATMENT (SECOND) OF TORTS, § 652D (“One who gives publicity to a matter concerning the private life of another is subject to liability to the other for invasion of his privacy, if the matter publicized is of a kind that (a) would be highly offensive to a reasonable person, and (b) is not of legitimate concern to the public.”); Patrick J. McNulty, The Public Disclosure of Private Facts: There is Life after Florida Star, 50 DRAKE L. REV. 93 (2001); John A. Jurata, Jr., The Tort that Refuses to Go Away: The Subtle Reemergence of Public Disclosure of Private Facts, 36 SAN DIEGO L. REV. 489 (1999). These elements are largely derived from William L. Prosser, Privacy, 48 CALIF. L. REV. 383 (1960). These concepts, while based in 19th century (and even earlier) doctrine, retain their vitality in the modern technological age. See, e.g., Jordan Segall, Google Street View: Walking the Line of Privacy—Intrusion upon Seclusion and Publicity Given to Private Facts in the Digital Age, 10 U. PITT. J. TECH. L. & POL. 23 (2010); Andrew Lavoie, The Online Zoom Lens: Why
In *Melvin v. Reid*, decided in 1931, for example, a homemaker, who had once worked as a prostitute and who had been wrongly accused of murder, became the subject of a feature film ("The Red Kimono") seven years after her acquittal, based on the facts of her trial. Although not specifically referencing a right to be forgotten, the court, permitting suit against the film-maker, noted: “One of the major objectives of society as it is now constituted, and of the administration of our penal system, is the rehabilitation of the fallen and the reform of the criminal.” The court held that the unnecessary use of the plaintiff’s real name inhibited her right to obtain rehabilitation. Similarly, in *Briscoe v. Reader’s Digest Association, Inc.*, decided in 1971, the court held that a publisher’s reference to the plaintiff’s prior crimes might infringe on his ability to obtain rehabilitation.

These kinds of cases have largely been overruled based on First Amendment concerns. In a series of opinions, the US Supreme Court held that newsworthy, true stories are protected by freedom of the press, although they may conceivably cause embarrassment or other harm to the stories’ subjects. In Internet Street-Level Mapping Technologies Demand Reconsideration of the Modern-Day Tort of ‘Public Privacy’, 43 GA. L. REV. 604 (2009); Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, SOC. SCI. RESEARCH NETWORK (Nov. 2004), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=629283 (question of what is “private” information cuts across many areas of American law, including Fourth Amendment, trade secrets, patents, evidence, the constitutional right of information of information privacy, and the Freedom of Information Act); Daniel J. Solove, *The Virtues of Knowing Less: Justifying Privacy Protections against Disclosure*, 53 DUKE L.J. 967, 976 (2003) (“Even in the current age, when information is king, sometimes less access to information is the soundest policy choice”).

---


34. See, e.g., Gates v. Discovery Comms., Inc., 101 P.3d 552 (Cal. 2004) (corporation not liable to offender for publishing facts obtained from public records); Wilan v. Columbia County, 280 F.3d 1160, 1163 (7th Cir. 2002) (noting that the *Melvin* case, paternalistic in doubting the ability of people to give proper rather than excessive weight to a person’s criminal history, is dead); Ostergren v. McDonnell, 643 F. Supp. 2d 758 (E.D. Va. 2009), (permitting republication of public documents containing sensitive information), rev’d, Ostergren v. Cuccinelli, 615 F.3d 263 (4th Cir. 2010); see generally ALLEN, supra note 20 (“Current interpretations of tort law do not favor granting relief under privacy tort theories to people whose once-public pasts have been resurrected by the media for public comment and discussion. The First Amendment and the common law mandate wide freedom for speaking truth, accurate news reporting and artistic expression.”); Patricia Sanchez Abril, *A (My)Space of One’s Own: On Privacy and Online Social Networks*, 6 NW. J. TECH. & INTELL. PROP. L. 73, ¶ 21 (2007) (in US, “causes of action that primarily protect one’s reputation, dignity, or privacy . . . have traditionally been both anemic and anomalous”). A variety of concerns other than the First Amendment may also affect views on the wisdom of expanding the right to privacy into a right to be forgotten. See generally Kent Walker, *Where Everybody Knows Your Name: A Pragmatic Look at the Costs of Privacy and the Benefits of Information Exchange*, 2000 STAN. TECH. L. REV. 2 (encouraging “rational review” of the benefits and costs of information exchange); Eugene Volokh, *Freedom of Speech, Information Privacy, and the Troubling Implications of a Right to Stop People from Speaking about You*, 52 STAN. L. REV. 1049 (2000) (suggesting that “a right to have the government stop people from speaking about you” may have “unintended consequences”).

35. See, e.g., The Florida Star v. B.J.F., 491 U.S. 524, 532 (1989) (“[I]f a newspaper lawfully obtains truthful information about a matter of public significance then state officials may not consti-
short, “[c]urrent interpretations of tort law do not favor granting relief under privacy tort theories to people whose once-public pasts have been resurrected by the media for public comment and discussion.” Yet, there are still some circumstances where US courts will accept privacy claims, even where the matter is worthy of media attention.

Outside the context of newsworthy stories, US courts have been less inclined to insist on unrestrained access to information. In Nixon v. Warner Communications, Inc., for example, the Supreme Court recognized a general right to inspect and copy public records, but also suggested that courts must exercise their supervisory powers to preclude access to information for “improper purposes,” such as to “gratify private spite or promote public scandal.” Similarly, in US Dep’t of Justice v. Reporters Comm. for Freedom of the Press, the Supreme Court recognized, in the context of a Freedom of Information Act re-
tionally punish publication of the information, absent a need to further a state interest of the highest order.”); Smith v. Daily Mail Publishing Co., 443 U.S. 97 (1979) (justification for prior restraint of publication requires showing that the state’s action furthers a state interest of the “highest order”); Oklahoma Publishing Co. v. District Court, 430 U.S. 308 (1977) (invalidating district court order enjoining newspapers from publishing name and picture of juvenile offender); Cox Broad. v. Cohn, 420 U.S. 469 (1975) (press could not constitutionally be exposed to tort liability for truthfully publishing name of rape and murder victim released to public in official court records); see generally Arminda Bradford Bepko, Public Availability or Practical Obscurity: The Debate over Public Access to Court Records on the Internet, 49 N.Y.L. SCH. L. REV. 967 (2005). Despite these opinions, even court records are not universally available to the public. See Peter A. Winn, Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information, 79 WASH. L. REV. 307, 309 (2004) (noting various criminal procedure rules that protect reputations, such as secrecy of grand jury proceedings, search warrant applications, and pre-sentence reports); id. at 311 (“[C]ourts tend to protect personal information when the purpose of access is not related to facilitating public scrutiny of the judicial process”).

38. See Allen, supra note 20, at 59 (“The First Amendment and the common law mandate wide freedom for speaking truth, accurate news reporting and artistic expression”).

39. See M.G. v. Time Warner, Inc., 107 Cal. Rptr. 2d 504 (2001) (Little League players and coaches had a privacy claim after video showed team’s group photo in a story about the team manager’s molestation of several pictured team members).

40. See Daniel J. Solove, A Tale of Two Bloggers: Free Speech and Privacy in the Blogosphere, 84 WASH. U. L. REV. 1195, 1199 (2006) (“Regarding the marketplace of ideas, truth must be weighed against other values, and the truth about a private person’s personal life is often not of much importance. Therefore, a balance between free speech and privacy might achieve these interests more effectively than merely protecting speech at all costs”). The precise division between “public” and “private” information, however, has sometimes been difficult to discern. See Charles N. Davis, Electronic Access to Information and the Privacy Paradox: Rethinking ‘Practical Obscurity’, ArXiv (2001), http://arXiv.org/ftp/cs/papers/0109/0109083.pdf (cases show that records “rarely fall” into “neat” categories).

41. 435 U.S. 589 (1978); see also Nixon v. Adm’r of Gen. Serv., 433 U.S. 425, 457 (1977) (noting that the President probably had a privacy right in some recordings, which was over-ruled by the Presidential Recordings Act).

42. See 435 U.S. 589, 598 (1978) (“[F]iles could serve as reservoirs of libelous statements for press consumption,” or as “sources of business information that might harm a litigant’s competitive standing”).

quest, that a “privacy interest” may exist in “keeping personal facts away from the public eye.”\textsuperscript{44} Indeed, the Court specifically noted that increased accessibility of information, as a result of “compilation of otherwise hard-to-obtain information” that “would otherwise have surely been forgotten,” threatened to affect the “privacy interest in maintaining the practical obscurity” of information.\textsuperscript{45}

There is, moreover, at least some suggestion in US case law that First Amendment concerns are diminished in the context of international communications. For example, in \textit{Yahoo!, Inc. v. La Ligue Contre Le Racisme Et L’Antisémitisme},\textsuperscript{46} the Ninth Circuit, reversing a lower court decision that refused on First Amendment grounds to enforce a French court decision compelling Yahoo to halt sales of Nazi memorabilia (illegal in France) on its site, noted: “[t]he extent of First Amendment protection of speech accessible solely by those outside the United States is a difficult and, to some degree, unresolved issue. . . .”\textsuperscript{47} More recently, in \textit{Holder v. Humanitarian Law Project},\textsuperscript{48} the US

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{44} See \textit{id.} at 769. Thus, the central purpose of the Freedom of Information Act (to protect the “‘citizens’ right to be informed about ‘what their government is up to’”) would “not [be] fostered by disclosure of information about private citizens that is accumulated in various governmental files but that reveals little or nothing about an agency’s own conduct.” \textit{id.} at 773. Indeed, the Court remarked that “both the common law and the literal understandings of privacy encompass the individual’s control of information concerning his or her person.” \textit{id.} at 763; see also \textit{Nat’l Archives & Records Admin. v. Favish}, 541 U.S. 157, 166 (2004) (privacy interest “at its apex” when records concern private citizens); \textit{Whalen v. Roe}, 428 U.S. 589, 599 (1977) (noting “individual interest in avoiding disclosure of personal matters”); id. at 605 (noting “threat to privacy implicit in the accumulation of vast amounts of personal information in computerized data banks”); \textit{Dep’t of the Air Force v. Rose}, 425 U.S. 352, 381 (1976) (republishing of information that may have been “wholly forgotten” can cause separate harm, which “cannot be rejected as trivial”).
\item \textsuperscript{45} See 489 U.S. at 780; \textit{id.} at 763-64 (“Plainly, there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearing-house of information”).
\item \textsuperscript{46} 433 F.3d 1199 (9th Cir. 2006) (en banc) (holding French decision unenforceable in the US), rev’d 169 F. Supp. 2d 1181 (N.D. Cal. 2001). The \textit{Yahoo!} decisions in the US followed a French court ruling that Yahoo! was required to “take any and all measures of such kind as to dissuade and make impossible any consultations by [Internet] surfers calling from France to sites [that] infringe upon the internal public order of France, especially the site selling Nazi objects.” UEJF \& LICRA v. Yahoo!, Inc., T.G.I. Paris, May 22, 2000, \textit{translated at} http://www.juriscom.net/txt/jurisfr/cti/y auctions20000522.htm. The Ninth Circuit ultimately held that Yahoo! could not bring a claim to invalidate the French decision, but was required to wait and respond to French plaintiffs, if they sought to enforce the decision in the US. See generally Marc H. Greenberg, \textit{A Return To Lilliput: The LICRA v. Yahoo! Case And The Regulation Of Online Content In The World Market}, 18 \textit{Berkeley Tech. L.J.} 1191 (2003) (overview of case and its jurisdiction implications).
\end{itemize}
\end{footnotesize}
Supreme Court upheld portions of the USA PATRIOT Act that criminalize speech when it is “coordinated” with “foreign terrorist” organizations. These, and other decisions, suggest that First Amendment protections are not absolute, at least in the context of foreign-related speech.

Viewed from the other side of the Atlantic, the European Union certainly does not disregard freedom of expression and freedom of the press as essential values. Indeed, recent EU pronouncements and court decisions expressly recognize the need to balance rights of privacy with freedom of expression. Various international declarations similarly support the need for such a balance.

49. See id. at 2747 (referencing USA PATRIOT Act, 18 U.S.C. § 2339B).

50. See, e.g., Meese v. Keen, 481 U.S. 46, 48 (1987) (upholding limits on distribution of foreign political propaganda in the US); Kleindeinst v. Mandel, 408 U.S. 753, 769 (1972) (US citizens may be denied personal access to foreign speakers). In the recent case of G.D. v. Kenny, 205 N.J. 275, 15 A.3d 300 (2011), the New Jersey Supreme Court observed: “In a free society, the right to enjoy one’s reputation free from unjustified smears and aspersions must be weighed against the significant societal benefit in robust and unrestrained debate on matters of public interest.” See id., 15 A.3d at 304. Thus, the court held that, despite the existence of a conviction expungement statute, “an offender has no protected privacy interest in expunged criminal records.” Id. at 308.


52. See WERRO, supra note 30 at 289 (suggesting that “in the context of a conflict between the right to be forgotten and the freedom of the press, the European Court will balance the competing interests and may well consider that in certain cases privacy rights trump the right to publish”); Oreste Pollicino & Marco Bassini, Internet Law In The Era Of Transnational Law, EUI Working Papers RSCAS 2011/24 at 29, available at http://eadmus.eui.eu/bitstream/handle/1814/16835/RSCAS_2011_24rev.pdf?sequence=1 (noting that “[f]reedom of expression” is protected in various European conventions, and in European case law).


54. International Conference of Data Protection and Privacy Commissioners, International
Thus, within Europe there is room for discussion of the need for balancing essential values in establishing the right to be forgotten.55

Regulators and legal theoreticians on both sides of the Atlantic, moreover, recognize that harmonizing international data protection laws may be key to maintaining the health of the world’s Internet-based economy.56 Indeed, the risk


56. See Article 29 Data Protection Working Party & Working Party On Police And Justice, The Future Of Privacy 6, Dec. 1, 2009 (“The establishment and functioning of an internal market requires that personal data should be able to flow freely from one Member State to another, while at the same time a high level of protection of fundamental rights of individuals should be safeguarded.”); Miguel P. Maduro, So Close And Yet So Far: The Paradoxes Of Mutual Recognition, J. EUR. PUB. POLICY 814, 817 (2007) (mutual recognition of national values is key to effective Internet regulation); see also EuroISPA, Personal Data Protection In The EU, supra note 55, at 5 (“The current rigid EU rules applying to the transfer of data to third countries do not seem adequate for the cross-border data flows in a globalised economy.”); see generally Steven C. Bennett (ed.), A Privacy Primer For Corporate Counsel, Ch. 9 (2009) (discussing international business challenges in privacy arena); Yaman Akdeniz, Case Analysis Of League Against Racism And Antisemitism (LICRA), French Union Of Jewish Students v. Yahoo! Inc., Yahoo France, 1 ELEC. BUS. L. REP. 110, 6 (2001) (noting “[t]he value of the Internet as a social, cultural, commercial, educational and entertainment global communications system the legitimate purpose of which is to benefit and empower online users, lowering the barriers to the creation and the distribution of expressions throughout the world”)
that EU data restrictions might prevent US companies from doing business in the European zone led the US Department of Commerce to develop a “Safe Harbor” construct\(^\text{57}\) with the input and approval\(^\text{58}\) of the EU.\(^\text{59}\) This approach has generally been successful\(^\text{60}\) and could be expanded.\(^\text{61}\) An array of other means to promote harmonization of US and EU views on the balance between privacy and free expression exist, and governments have pursued these options in recent years.\(^\text{62}\) Additionally, the existence of long-standing concepts of “fair information practices” provides a solid base for common discussion among regulators.\(^\text{63}\)

Recent political developments in the United States suggest that US regulators and law-makers may be particularly receptive to discussions on the merits

(quotations omitted).


\(^{62}\) See generally Bernhard Maier, How Has The Law Attempted To Tackle The Borderless Nature Of The Internet?, 18 INT’L J. L. & INFO. TECH. 142 (2010) (summarizing efforts at harmonization of international law regarding the Internet); Rustad & Koenig, supra note 12 (suggesting means to promote harmonization, including ALI/UNIDROIT consultation, treaty negotiations and expansion of the Hague Convention on Jurisdiction and Foreign Judgments, and choice of law/choice of forum provisions in user agreements).

\(^{63}\) See David Banisar, The Right To Information And Privacy: Balancing Rights And Managing Conflicts, at 7, 2011, http://wb.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/Right%20to%20Information%20and%20Privacy.pdf (“Since the 1960s, principles governing the collection and handling of [private] information (known as ‘fair information practices’) have been developed and adopted by national governments and international bodies”).
of enhanced privacy protection. In December 2010, the FTC staff issued a “preliminary” report, aimed at providing a “broad privacy framework to guide policymakers, including Congress and industry.” The FTC Report called for a wholesale “re-examination” of the FTC’s approach to privacy protection.

Shortly after the FTC released its 2010 report, the Department of Commerce issued its own report on “Commercial Data Privacy and Innovation in the Internet Economy” (“Commerce Report”). The Commerce Report set out four main goals for US privacy protection policy: (1) to enhance consumer trust online through the recognition of “revitalized” fair information practice principles; (2) to enhance consumer trust online through the recognition of “revitalized” fair information practice principles; (3) to enhance consumer trust online through the recognition of “revitalized” fair information practice principles; (4) to enhance consumer trust online through the recognition of “revitalized” fair information practice principles.


Id. at 19.


Id. at 3-7.

The Commerce Report suggested that fair information principles should “enhance[e] transparency, encouraged [greater detail in purpose specifications and use limitations, and foster[ the development of verifiable evaluation and accountability programs. . . .” Id. at 30. The report noted “lengthy and complex” disclosures that “fail to inform,” and suggested that “privacy rights depend on [consumer] ability to understand and act on” company privacy policies. Id. at 31 (documents written in “legalese” are “typically overwhelming” to the average consumer) (quotations omitted). The Report encouraged “reduced length and greater simplicity and clarity” in privacy disclosures. Id. at 33.
(2) to encourage the development of “voluntary, enforceable” privacy codes of conduct through “collaborative efforts” with government;\(^70\) (3) to encourage “global interoperability;”\(^71\) and (4) to ensure “nationally consistent” privacy rules.\(^72\)

These stated goals for US data protection policy certainly recognize the global nature of information technology issues. Indeed, EU data policy developments have, to some degree, pushed the world toward uniform standards of data protection, and have spurred US regulators to action.\(^73\) Moreover, EU developments have sparked US interest in dialogue with EU authorities.\(^74\) Additional dialogue on the subject of data protection and privacy should develop.\(^75\)

\(^70\). The Commerce Report suggested the need to promote development of “flexible but enforceable codes of conduct,” to address “emerging technologies and issues not covered” by current fair information practices. Id. at 41; see id. at 42 (noting risk that privacy practices may “ossify”).

\(^71\). The Commerce Report noted that “[d]isparate approaches” to data privacy can “create barriers” to trade, “harming both consumers and companies.” Id. at 53. The report reviewed a host of options for “greater harmonization and international interoperability.” Id. at 54 (citing creation of a global privacy standard, adoption of a treaty or convention to govern cross-border data flows, an enhanced US privacy framework that “can be more easily supported abroad,” increased Department Of Commerce international advocacy for US interests, more “focused and coordinated” US government advocacy of the US position internationally, creation of “accountability certifications,” such as binding corporate rules, application for “adequacy” status with the European Union, and development of a US framework that “furthers harmonization” of international privacy laws, including the EU directive). Id. at 54-55. The report suggested the possibility to take harmonization work “to the next level,” by creating “binding trade commitments” to “steer the world toward global privacy protection interoperability.” Id. at 56.

\(^72\). The Commerce Report suggested the need for a “comprehensive” commercial data security breach framework, using federal preemption, to prevent the “maze” of disparate state laws from becoming “costly and burdensome” to business. Id. at 57 (quotations omitted). Any new federal privacy framework should, however, “seek to balance the desire to create uniformity and predictability across State jurisdictions with the desire to permit States the freedom to protect consumers and to regulate new concerns” from emerging technologies that could “create the need for additional protection[.]” Id. at 61.

\(^73\). BALLON, supra note 29 (noting that EU adoption of data privacy Directive “pushed the [FTC] to become increasingly active in the area of internet privacy;” today, the FTC “plays a prominent role in shaping debates over privacy protection and in encouraging compliance”); Moving Forward, supra note 64, at 81-82 (noting that the FTC has joined with 12 EU regulators to launch the Global Privacy Enforcement Network, and that the FTC has been officially admitted to annual conference of data protection and privacy commissioners, and DOC officials have become “increasingly visible” in meetings with EU data protection authorities).


\(^75\). In December 2010, the EU’s Reding asserted that US officials were “unprepared” and “uninterested” in negotiating over data privacy issues. See Cyrus Farivar, EU Official Dissatisfied With American Response To Data Protection Concerns, DEUTSCHE WELLE, Dec. 12, 2010, http://www.dw.de/dw/article/0,,14729661,00.html (“They [the US] have not even appointed a negotiator.”) (quoting Reding). In January 2011, a Department of Commerce representative summarized
Given the breadth of developments in technology and usage of the Internet, and given the increasing globalization of Internet-based commerce, changes in the substantive standards for privacy appear almost inevitable. Thus, EU plans to revisit the data protection directive to improve harmonization within the European Union itself may offer a particularly good opportunity for such dialogue with US authorities.

At a minimum, US engagement of EU authorities can provide clarification of the EU view of the right to be forgotten. The development of a single EU standard for such a right could also provide greater certainty and predictability


77. Hon, Millard & Walden, supra note 27 (noting “important national differences in data protection laws” within EU, including degree of civil liability and penalties for violations).


79. It is possible, of course, that EU regulators may conclude that no single, useful standard for the right to be forgotten can be formulated, and that the issue might be better left for further discussion as law, technology and Internet usage develop. See Stuart Robertson, Hasty Legislation Will Make A Mess Of Europe’s “Right To Be Forgotten;” The Ethics Of Online Deletion, THE REGISTER, Nov. 12, 2010, http://www.theregister.co.uk/2010/11/12/privacy_legislation/ (“[T]he problems [with a right to be forgotten] are technical, ethical and legal. Most of all they are complex, and EU legislators would be fools to write laws covering such sensitive ground in any kind of hurry”).
Bennett: The "Right to Be Forgotten": Reconciling EU and US Perspectives 2012] THE “RIGHT TO BE FORGOTTEN” 179

to foreign businesses operating in the European Union. 80 Substantive reconciliation of EU and US data protection law need not necessarily take on the full debate about freedom of expression versus privacy. 81 Various minimalist solutions might emerge. For example, regulators in the European Union and United States could discuss the scope of the right to be forgotten, and at least agree on certain minimum standards in the area. 82 Further, the discussions might include methods to support self-regulation (or “co-regulation”) to promote aspects of the right to be forgotten. 83 The authorities might also address the specific form of

80. Industry Joint Statement On The Review Of The EU Legal Framework For Data Protection, Mar. 15, 2011 (European communications trade groups suggest that application of the EU Directive has not “resulted in the harmonised framework and level playing field necessary to establishing certainty for data controllers and individuals”; EuroISPA, Personal Data Protection in the EU, supra note 55, at 1 (noting that data protection “Directive has failed in creating a harmonized framework across the EU,” and calling for application of rules “horizontally,” to create a “level playing field”); Eric Pfanner, EU Seeks To Bolster Web Privacy; Data Protection Rules Will Be Updated With New Internet Services In Mind, INT’L HERALD TRIBUNE, Nov. 5, 2010, (“Technology companies have also been calling for an update of EU privacy rules . . . [because] there are too many different interpretations of existing legislation across the 27-country bloc.”); EU Wants To Give People Power To Vanish From Internet, A GENCE FRANCE PRESSE, Nov. 4, 2010, http://www.eubusiness.com/news-eu/consumer-privacy.6sw (“The lack of harmonization of data protection rules creates enormous challenges for entrepreneurs who are trying to use emerging technologies to expand into new markets.”) (quoting Jonathan Zuck, President of Association for Competitive Technology); Hill, supra note 64.

81. FTC Staff Comments, supra note 74, at 8 (noting that question of “international standards” for data protection and privacy presents “a highly complex and technical subject in which there remain significant unresolved political and policy debates”); id. at 9 (given “lack of consensus,” FTC supports “efforts to promote more consistency and inter-operability,” but suggests that “binding general international standards at this stage are premature”); EU Council Conclusions On Personal Data Protection, supra note 53, at 3, 5 (noting that protection of personal data transferred to countries outside the European Union is “one of the most complex issues in the course of the review” of the Directive; and suggesting that “development of universal principles” is “of utmost importance because of the globalised nature of data processing”).


83. See UK Advertising Association, supra note 55, at 4 (current cross-border data transfer regime is “not effective and is neither consistent nor business-friendly”) (suggesting “self-regulatory solutions” to “complement” legal framework); EFAMRO / ESOMAR, supra note 82, at 10 (“Self-regulation provides a level of detail and granularity that is impossible to achieve in national or supranational legislation and encourages sector-specific authoritative guidance and regulation.”); see also CTR. FOR DEMOCRACY & TECH., supra note 24, at 9 (suggesting use of “coregulatory approaches” to privacy governance in EU, as means to provide “international harmonization”); Ira S. Rubinstein, Privacy And Regulatory Innovation: Moving Beyond Voluntary Codes, Mar. 2010, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1510275 (emphasizing value of Safe Harbor systems for promoting privacy protection). A large body of literature exists on the developing use of co-regulation in government. See, e.g., Peter S. Rank, Co-Regulation Of Online Consumer Personal
implementation of any right to be forgotten. Finally, to the extent that certain technical solutions (such as systems for auto-deletion of information) might address concerns underlying the right to be forgotten, regulators could (and should) discuss means to facilitate development and use of such technology.

84. Thus, for example, the European Union and the United States might agree on some form of “notice-and-takedown” of content approach, to shield website purveyors from unanticipated liability. See Center for Democracy & Technology, supra note 24, at 12 (“[A] person demanding takedown of content she did not create should be required to obtain a judicial or administrative determination that the content in question is illegal and should be removed.”); Cynthia Wong, Don’t Blame The Messenger, July 29, 2010, http://www.america.gov/st/democracyhr-english/2010/July/20100727142911enelrahc0.9917871.html (noting use of “notice-and-takedown systems” to deal with copyright and other problems).

85. Systems of auto-deletion of information have been proposed. See Liam J. Bannon, Forgetting as “A Feature, Not a Bug”: The Duality of Memory and Implications for Ubiquitous Computing (2006), http://www3.unin.it/events/alpis06/download/prog/16_Bannon_2.pdf (suggesting that mechanisms of forgetting are of central concern to human psychology and that electronic tagging systems for information to “time-stamp material and contain something like a sell-by date” should be developed); Chris Conley, The Right to Delete 57 (2010), http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1158/1482 (“By building an expiration date into the content that we create, we could indeed address some of the privacy concerns that persistence presents.”); Roxana Geambasu, Tadayoshi Kohno, Amit A. Levy & Henry M. Levy, Vanish: Increasing Data Privacy With Self-Destructing Data, PROCEEDINGS OF THE 18TH USENIX SECURITY SYMPOSIUM 299 (2009), http://vanish.cs.washington.edu/pubs/usenixsec09-geambasu.pdf; see generally Adam Thierer, Two Paradoxes of Privacy Regulation (Aug. 25, 2010), http://techliberation.com/2010/08/25/two-paradoxes-of-privacy-regulation/ (advocating “user-empowerment tools” as best means to protect privacy). The concept of auto-deletion has won supporters, both in the EU and in the US. See INTERNATIONAL CONFERENCE OF DATA PROTECTION AND PRIVACY COMMISSIONERS, Draft Resolution on Privacy Protection in Social Network Services 1, 3 (Oct. 17, 2008), http://www.justice.gov/INR/donlyres/F8A79347-170C-4EEF-A0AD-155554558A5F/24477/2008.pdf (“[I]t can be very hard – and sometimes even impossible – to have information thoroughly removed from the internet once it is published.”) (suggesting that providers of social network services should “allow users to easily terminate their membership, delete their profile and any content or information that they have published on the social network”); CTR. FOR DEMOCRACY & TECH., supra note 24, at 11 (Center “supports empowering individuals to delete data they themselves have created” but would strongly resist measures to “delete comments on other websites” as presenting “a high risk that one user’s right to be forgotten will unduly hamper others’ free expression rights and leave intermediaries with the difficult, potentially impossible, task of ‘disentangling’ individuals’ data”). Auto-deletion systems, however, may present certain technical problems for implementation. See Fleischer, supra note 14 (noting technical problems with auto-delete systems).

86. See EuroISPA, Personal Data Protection in the EU, supra note 55, at 1 (noting “it is important to address the impact that future innovations can produce on privacy through non-legislative measures, such as the use of privacy-enhancing technologies, privacy-by-design and industry self-
Despite the progress to date, and prospects for additional efforts at harmonization, the fact remains that US and EU views of privacy protection (and the right to be forgotten, in particular) are currently in conflict. So long as such conflict exists, a significant procedural question arises: What is the scope of the jurisdiction of EU authorities to regulate and adjudicate the activities of actors operating outside the European Union, where some effects of that activity arguably arise within the European Union? In an inter-connected world, such a scenario inevitably arises. Effective enforcement of any right, including the
right to be forgotten, depends upon a system of jurisdiction that permits effective enforcement of that right. However, the problem of conflicting jurisdictions over Internet activity presents an enormous conundrum, which commentators have long recognized.

Traditionally, the exercise of jurisdiction was based on the concept of “sovereignty” over territory. Modern conflicts of law theory, however, recognizes the potential for expansion of jurisdiction based on the nationality of the defendant, protection of national interests, universal jurisdiction for offenses harmful to humanity, and the “passive personal” theory of adverse effects on a citizen subject to a state’s jurisdiction. In the essentially borderless region of cyberspace, the concept of sovereignty over specific physical territory becomes particularly problematic. Under these conditions, unless nations generally agree on a “law

90. Sarudzai Chitsa, Name Calling on the Internet: The Problems Faced by Victims of Defamatory Content in Cyberspace, Paper No. 48, CORNELL LAW SCHOOL INTER-UNIVERSITY GRADUATE STUDENT CONFERENCE PAPERS (2011), http://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=1077&context=lps_clacp (noting potential “conflict of laws nightmare” in context of Internet content regulation); Rustad & Koenig, supra note 12 (quotation omitted) (“The global Internet’s legal environment makes it inevitable that one country’s laws will conflict with another’s—particularly when a Web surfer in one country accesses content hosted or created in another country.”); Timofeeva, supra note 88 (noting “heated” discussion regarding “multiple overlapping conflicting jurisdictions” claiming authority over Internet activities).

91. From the outset, US commentators recognized that the EU Directive presented challenges to US regulatory authorities. See generally Fred H. Cate, Data Protection Law and the European Union Directive: The Challenge for the United States, 80 IOWA L. REV. 431 (1995); see also Alexander Gigante, Ice Patch on the Information Superhighway: Foreign Liability for Domestically Created Content, 14 CARDOZO ARTS & ENT. L.J. 523 (1996); Kuner, supra note 87 (“While the fundamental, high-level principles of data protection law are similar across regions and legal systems, the details of the law differ substantially. Given these differences, it is not surprising that data protection law has been the subject of an increasing number of jurisdictional disputes, many of them involving [the European Union and the United States]”).


93. See United States v. Yunis, 681 F. Supp. 896 (D.D.C. 1988) (reviewing theories); Timofeeva, supra note 88, at 201 (quotation omitted) (although international law “sets little or no limit on the jurisdiction which a [particular] state may arrogate to itself[,]” several “established principles are more or less recognized in all jurisdictions”).

94. See Am. Libraries Ass’n v. Pataki, 969 F. Supp. 160, 168-69 (S.D.N.Y. 1997) (“The unique nature of the Internet highlights the likelihood that a single actor might be subject to haphazard, uncoordinated, and even outright inconsistent states that the actor never intended to reach and
of the Internet," a consistent, understandable “effects” standard of jurisdiction naturally becomes the most likely substitute for a strict “sovereignty” approach. An unlimited “effects” doctrine, however, could be used to justify virtually unlimited jurisdiction. Online intermediaries such as social networks and other facilitators of Internet content are “at the front lines” of this problem, as their content may be viewed from a computer anywhere on the planet.

The divergence of views on Internet-based jurisdiction appears in US case law and EU official pronouncements. To a large extent, US courts have fol-

possibly was unaware were being accessed. Typically, states’ jurisdictional limits are related to geography; geography, however, is a virtually meaningless construction on the Internet.

95. See generally David R. Johnson & David G. Post, Law and Borders – The Rise of Law in Cyberspace, 48 STAN. L. REV. 1367 (1996) (arguing for development of Internet law independent of individual countries). But see Frank H. Easterbrook, Cyberspace and the Law of the Horse, 1996 U. CHI. LEGAL F. 207 (1996) (suggesting no need exists for creation of separate law of the Internet); Timothy S. Wu, Cyberspace Sovereignty? – The Internet and the International System, 10(3) HARV. J.L. & INFO. TECH. 647, 649 (1997) (arguing that Johnson and Post’s descriptive assumptions – “that the ‘territorial’ powers of the world will, or already do, respect an emergent cyberspace sovereignty” and that “state regulation of the Internet will be impossible or futile” – are incorrect) (“Internet regulation, although difficult, is possible and stands to become increasingly so regardless of its desirability on normative grounds”).

96. Bernhard Maier, How Has the Law Attempted to Tackle the Borderless Nature of the Internet?, 18(2) INT’L J. L. INFO. TECH. 142, 142 (2010) (regulations must address fact that actions may not physically take place in territory, but still have effects there).

97. Jessica R. Friedman, A Lawyer’s Ramble Down the Information Superhighway: Defamation, 64 FORDHAM L. REV. 794, 803 (1995) (defendants potentially liable for suit in every jurisdiction where access to defamatory content may be had from the Internet); Kulesza, supra note 94 (the fact of “enabling contents to be available within a certain territory cannot be [a] basis” for the exercise of the “prerogatives of the ruling sovereign;” rather, the “result of such [a] practice would be the ultimate insecurity of the Net”); Uerpmann-Wittzack, supra note 55, at 1256 (“[Jurisdiction based on an unqualified effects doctrine would not only infringe the sovereignty of other states, but it would also collide with the principle of Internet freedom.”)); Pollicino & Bassini, supra note 52, at 9 (“[I]f the Internet makes websites accessible anywhere, and proper jurisdiction [arises] in any state where a harm occurs due to their contents, two paths are feasible: either the contents must comply with all the relevant jurisdictions where the website can be accessed, or access to such contents may be limited to those countries which has [sic] not outlawed them”).


100. See Toys “R” Us, Inc. v. Step Two, S.A., 318 F.3d 446, 452 (3rd Cir. 2003) (noting that
followed the reasoning in Zippo Mfr. Co. v. Zippo Dot Com, Inc.,101 in which the court referenced a “sliding scale” for determining whether Internet activity could form the basis for personal jurisdiction.102 At one end of the spectrum, according to the Zippo court, are situations where a defendant clearly does business over the Internet by entering into contracts with residents of a foreign jurisdiction.103 At the other end of the spectrum are situations where a defendant has simply posted information on an Internet site.104 In the middle are cases where an interactive website permits the user to exchange information with the host computer.105 In those cases, “the exercise of jurisdiction is determined by examining the level of interactivity and commercial nature of the exchange of information that occurs....”106 Although the Zippo test has not proved infallible in its application,107 its essential notion, that purely “passive” operation of a website should not form the sole basis for exercise of personal jurisdiction, seems relatively well established in US law.108

In contrast, EU interpretations on the reach of European privacy law (and

---

102. Id. (“[T]he likelihood that personal jurisdiction can be constitutionally exercised is directly proportionate to the nature and quality of commercial activity that an entity conducts over the Internet”).
103. Id. (jurisdiction is proper where contacts involve “knowing and repeated transmission of computer files over the Internet”).
104. Id. (“passive” website “does little more than make information available”).
105. Id.
106. Id.
related concepts such as defamation) appear to extend beyond the bounds of the prevailing US test for jurisdiction. The EU Data Protection Directive (the “Directive”) adopted in 1995 does not expressly state that its provisions apply to the activities of non-EU entities but does purport to apply EU substantive law to any organization that uses means within the European Union to collect or process personal data. In 2000, the European Union issued a further directive “on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market,” which again did not expressly touch on the reach of EU jurisdiction. In 2002, the Article 29 Data Protection Working Party (the “Working Party”), an advisory group associated with the European Union, issued its “working document on determining the international application of EU data protection law to personal data processing on the Internet by non-EU based websites.” The Working Party suggested that an online interaction between a website operator with no legal establishment in the European Union and an individual residing in the European Union may suffice to trigger coverage under EU data protection law. The full reach of this Working Party

109. Chen, supra note 107, at 436 (EU approach to Internet jurisdiction “markedly different” from US approach; EU is “highly regulatory”); id. at 445 (EU “country-of-destination” approach may be “overly broad and an unfair burden on [internet] sellers”).


111. See Lokke Moerel, Back To Basics: When Does EU Data Protection Law Apply?, 1 INT’L DATA PRIV. L. 92, 92 (2011) (“It is much debated when the data protection laws of the EU Member States apply in international situations. . . . The lack of guidance in the Directive on key concepts of applicable law and jurisdiction has lead to unacceptable differences in the manner in which the provision is implemented in the Member States.”); see generally Eleni Kosta, Christos Kalloniatis, Lilian Mitrou and Evangelia Kavakli, The “Panopticon” of Search Engines: The Response of the European Data Protection Framework, 15 REQUIREMENTS ENGINEERING 2 (2010) (noting “heated debate” among European privacy professionals on whether EU data protection framework applies to search engine providers that process data from outside the EU).


115. See id. at 15 (noting that the Working Party is convinced that a high level of protection of individuals can only be ensured if web sites established outside the European Union but using equipment in the European Union as explained in this working document respect the guarantees for
opinion has not been tested, but on a technical reading of the EU Directive, anyone who posts personal information about another person on his or her own social networking profile or uses personal information from another person’s profile could be deemed a “data controller” subject to the data protection obligations of the Directive. In that event, recognition of a “right to be forgotten” could have very broad consequences.

Indeed European case law tends to extend well beyond US views on the reach of jurisdiction, based on Internet activity. In substance, so long as actions on the Internet have known “effects” in a European state, EU courts (and, by implication, EU regulators) may exercise jurisdiction. In addition to the Yahoo! case, where a French court held Yahoo! responsible for permitting sale of Nazi-themed materials in France, and a host of other similar cases, the personal data processing, in particular the collection, and the rights of individuals recognized at European level and applicable anyway to all web sites established in the European Union).

116. See Lokke Moerel, The Long Arm Of EU Data Protection Law: Does The Data Protection Directive Apply To Processing Of Personal Data Of EU Citizens By Websites Worldwide?, 1 Int’l Data Privacy L. 28 (2011) (“The conclusion is that the interpretation given by the Working Party is contrary to the legislative history of [the Data Protection Directive]. . . . Although the attempt of the Article 29 Working Party to provide protection to EU nationals is commendable, this result should be achieved by amendment of the applicability rule. . . .”); Wang, supra note 99, at 240 (“[T]here is still no clear indication of the creation of a special regime of jurisdiction rules for e-commerce cases. . . . Even if efforts were made to draft a specific regulation or convention, it would still take time and efforts to come into force.”); see also Kuner, supra note 87; Chen, supra note 107.

117. Daniel B. Garrie, Hon. Maureen Duffy-Lewis, Rebecca Wong & Richard L. Gillespie, Data Protection: The Challenges Facing Social Networking, 6 BYU INT’L L. & MGMT. REV. 127, 131-32 (2010) (to require “every user” to comply with Directive is “unrealistic objective”); id. at 133 (“impractical,” and “not customary” for users to “ask permission before posting another’s personal information, such as a photo or video”). The key under EU law, may turn out to be the extent to which information is accessible beyond a group of “self-selected contacts.” See id. at 143 (citing Sweden v. Lindquist, 2003 E.C.R. 1-12971 (holding that personal use exception to EU Directive does not apply where personal information is accessible by anyone on the Internet, rather than a limited number of self-selected contacts); Denis T. Rice, Jurisdiction Over Privacy Issues On The Internet, (Jul. 15, 2003), http://www.martindale.com/business-law/article_Howard-Rice-Nemirovsky-Canady-Falk_17400.htm (“the notion of accessibility as the basis for jurisdiction is far from dead” in EU law, citing Yahoo! Inc. v. La Ligue Contre Le Racisme, 433 F.3d. 1199 (9th Cir. 2006)).


119. Chris Brummer, Territoriality As A Regulatory Technique: Notes From The Financial Crisis, 79 U. CINN. L. REV. 101, 109 (2010) (“[R]egulators can assert jurisdiction extraterritorially wherever foreign companies engage in conduct that has effects in the country asserting jurisdiction; “this kind of strategy has been used to most spectacular effect” in EU antitrust actions); Marike Vermeer, Unfair Competition Online And The European Electronic Commerce Directive, 7 ANN. SURVEY INT’L & COMP. L. 87, 94-96 (2001) (noting that European law will often point to “lex loci delicti,” or “market” effects rule; these rules are not “effective,” as they permit “too many national laws” to apply).

120. See generally Timofeeva, supra note 88 (noting examples of cases in Germany, France and Italy, and suggesting that “the effects principle as applied in asserting jurisdiction in Internet content controversies is employed most broadly, capable to justify almost anything”).
recent criminal prosecution of Google executives in Italy, the Italian court held that, because at least some of the processing of information (a video of a child with Down’s Syndrome being abused by other youths) took place in Italy, the court could properly exercise jurisdiction. Thus, if “processing” of “personal data” through EU “equipment” includes a user’s downloading of Internet content somewhere in Europe, the European Union theoretically could exercise world-wide jurisdiction over Internet actors.

The disparity of views on the reach of jurisdiction over Internet-related activities can produce uncertainty, additional cost (in responding to varying standards) and unnecessary barriers to trade (as firms may be deterred from activities that place them at risk of regulation in unfavorable jurisdictions). In addition, the risk that judicial and administrative orders in one jurisdiction may not be enforced in other countries may tend to deter effective implementation of rules.

---


122. Christopher Kuner, Data Protection Law And International Jurisdiction On The Internet Part 2, at 3, 2009, www.ssrn.com [Note: the link is temporarily down. I would say just leave it as this]. (EU concepts of “personal data” and “data processing” are “interpreted very expansively, which “increases their jurisdictional scope”).


124. See Kuner, supra note 122, at 3 (noting that EU use of “equipment” as basis for exercise of jurisdiction “most controversial,” because connection to the European Union may be very limited).

125. See Brummer, supra note 119, at 112 (“extraterritorial regulation, even when justifiable, generates costs,” as foreign firms “must adjust to new standards or move to other jurisdictions to avoid a law’s regulatory effect;” extraterritorial regulation also “often erodes [a country’s] reputation in the international community;” as a result, other regulators and courts “may decide to refrain from cooperating with [the regulating country] or helping it achieve its strategic objectives,” as by refusing to enforce judgments); Kuner, supra note 122, at 4 (jurisdictional uncertainties about data protection law “may dissuade individuals and companies from engaging in electronic commerce,” and may “impose burdens” on commerce); Timofeeva, supra note 88 (prospect for assertion of worldwide jurisdiction “contributes to legal uncertainty”); Chen, supra note 107, at 423 (“The unpredictability of jurisdiction makes it difficult for companies with web sites to limit their legal liability and inhibits the growth of e-commerce.”); Adam Thierer & Clyde Wayne Crews, Jr., Everybody Wants To Rule The Web (Dec. 17, 2003), http://www.cato.org/pub_display.php?pub_id=3343 (noting that “patchwork” of international law may be “confusing, costly, and technically impossible for all but the most well-heeled firms” to navigate).

126. See Chris Reed, Think Global, Act Local: Extraterritoriality in Cyberspace (2010), papers.ssrn.com/sol3/papers.cfm?abstract_id=1620129 (“A state is unlikely to attempt to enforce its laws against foreign defendants where it is known that their home country will probably refuse to enforce a judgment”). In such instances, however, both jurisdictions have an interest in developing a solution, to promote “comity” (equal treatment of laws) between nations. See RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES, § 403(1) (1987) (noting that comity...
Ideally, the United States and the European Union could develop some form of an agreed international standard on jurisdiction. But even if agreement upon a general standard is impossible, or at least unlikely to develop in the immediate future, the United States and the European Union could still achieve less ambitious improvements in international understanding. Recently, for example, the EU Article 29 Working Party issued an opinion on “applicable law” under the EU Directive, which suggested that “additional criteria” should be developed to determine when EU data protection law applies to a “controller established outside the EU. . . .” The Working Party suggested that these criteria could include the “targeting” of individuals within the European Union.

The notion of “targeting” as a standard for the exercise of jurisdiction has relatively wide support among commentators. US courts have developed exceptions of “the recognition which one nation allows within its territory to the legislative, executive, or judicial acts of another nation, having due regard both to international duty and convenience, and to the rights of its own citizens or of other persons who are under the protections of its law”).

127. See Burke T. Ward & Janice C. Sipior, Where In The World Is Internet Jurisdiction: A US Perspective, 4 INT’L J. VALUE CHAIN MGMT. 5 (2010) (noting need to develop “globally agreed” standard for jurisdiction related to Internet activity); Timofeeva, supra note 88 (international agreement on jurisdiction standards for Internet-related issues “certainly” best choice, if possible to achieve); Chen, supra note 107, at 447 (“obvious solution” to divergence in approaches to Internet jurisdiction is for the United States and the European Union to “cooperate and develop an international framework”); see also Julia Marter, When and Where Does an Internet Posting Constitute Publication? Interpreting Moberg v. 33T LLC, 21 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 495 (2011) (“Surprisingly, the legal community has not yet answered this question”).


129. See id. at 24 (“A more specific connecting factor, taking the relevant ‘targeting’ of individuals into account, as a complement to the ‘equipment/means’ criteria could be useful in terms of legal certainty[,] Such a criterion is not new and has been used in other contexts in the EU, and by the United States’ legislation on the protection of children on-line.”); see id. at 31 (“The following examples illustrate what targeting could consist of: the fact that a data controller collects personal data in the context of services explicitly accessible or directed to EU residents, via the display of information in EU languages, the delivery of services or products in EU countries, the accessibility of the service depending on the use of an EU credit card, the sending of advertising in the language of the user or for products and services available in the EU”).

130. Pollicino & Bassini, supra note 52, at 9-10 (suggesting that “common denominator” in authorities is that “as long as websites do not target nor produce harm to certain individuals or entities, a domestic jurisdiction cannot be asserted on the sole ground that website contents do not comply with the laws of that state”); Lorna E. Gilles, Addressing the “Cyberspace Fallacy”: Targeting the Jurisdiction of an Electronic Consumer Contract, 16 INT’L J. L. & TECH. 242 (2008) (suggesting use of “intentional targeting” standard as basis for jurisdiction); Timofeeva, supra note 88 (“Targeting-based analysis could be a solution to the unlimited application of the effects principle if the courts could agree to accept it in a consistent form.”); Thomas Schultz, Carving Up The Internet: Jurisdiction, Legal Orders, And The Private/Public International Law Interface, 19 EURO. J. INT’L L. 799 (2008) (noting risk that Internet will be “fragmented” into discrete legal spheres, by local law, and suggesting a “principle of targeting” and “effects doctrine” as solution); Brian D. Boone, Bullseye! Why A “Targeting” Approach To Personal Jurisdiction In The E-Commerce Context Makes Sense Internationally, 20 EMORY INT’L L. REV. 241 (2006); Greenberg, supra note 46, at 1253-54 (suggesting that agreement on international convention on jurisdiction may be “impossible,” but “targeting” test can be developed with existing standards); Michael A. Geist, Is There A There
Bennett: The "Right to Be Forgotten": Reconciling EU and US Perspectives

2012] THE “RIGHT TO BE FORGOTTEN”

perience in applying such a standard and, to a lesser extent, the notion is recognized in the European Union. Furthermore, both US and EU courts are likely to agree that, on an intuitive level, a country almost certainly enjoys jurisdiction over the "territory" of its top-level Internet domain. Courts can also recognize other indicia of "targeting," such as language, specific content and references to the particular country. The International Organization of Securities Commissions adopted just such a test for national jurisdiction over securities offerings. In the United States, the FTC applies similar standards in determining whether websites are addressed at children. A further example appears in the EU Convention on Cybercrime. Such standards, of course, rely to some


131. Haynes, supra note 100, at 159-60 (noting that many US courts have added a "targeting" element to the Zippo test, which provides "a means of focusing on a defendant’s action—those directed toward a particular jurisdiction, rather than actions directed at all jurisdictions simultaneously"); Timofeeva, supra note 88 (noting that targeting-based analysis is "not a novel doctrine," although, in the Internet setting, "the United States alone favors its application"); Reidenberg, supra note 88, at 1955 (US courts "have looked to online targeting and to deleterious effects within the forum to determine if personal jurisdiction is appropriate"). The issue of targeting has also arisen in US First Amendment jurisprudence. See United States v. Playboy Entertainment Grp., 529 U.S. 803, 804 (2000) ("The Government cannot ban speech if targeted blocking is a feasible and effective means of furthering its compelling interests").


133. Uerpmann-Wittzack, supra note 55, at 1258 (suggesting that countries may "assert full jurisdiction" over matters within their own "top level domain," which "becomes a state’s territory in cyberspace"); Timofeeva, supra note 88 (noting "authority of a country to administer its own Country-Code Top-Level Domain") ("The manager of the German ccTLD ‘.de’ would not register the domain name http://www.heil-hitler.de or the like").

134. See supra notes 128-129 (Article 29 Working Party suggestions for targeting criteria); see generally Matthew L. Perdoni, Revising The Analysis Of Personal Jurisdiction To Accommodate Internet-Based Personal Contacts, 14 U.D.C. L. REV. 159 (2011).

135. Factors include: whether the offeror accepts orders from or provides services to residents in the jurisdiction, whether the offeror uses email or other media to “push” information to residents, and (contrariwise), whether the offeror clearly states that it does not intend to make an offering in specific jurisdictions. See IOSCO, Securities Activities On The Internet (1998), http://www.iosco.org/library/pubdocs/pdf/IOSCOPD83.pdf; IOSCO, Securities Activities On The Internet II (2001), http://www.iosco.org/library/pubdocs/pdf/IOSCOPD120.pdf.

136. See FTC, What Determines Whether Or Not A Website Or Online Service Is Directed To Children?, http://www.ftc.gov/privacy/coppaflow.htm (FAQ section) (factors may include subject matter, language, use of animated characters, and whether advertising appeals to children).


Published by Berkeley Law Scholarship Repository, 2012
degree on subjective and ambiguous factors and may introduce new complications. But a workable standard is at least theoretically possible.

Similarly, the development of “geo-location” technologies for the Internet potentially opens the way to development of new standards for jurisdiction. Website purveyors often tailor their content to specific markets and use geolocation technology to assist them in delivering targeted messages. Indeed,
technologies for location identification, based on Internet usage, mobile device usage, or both, may offer tremendous opportunities for “personalization of services and contextualization of information.” These geo-location technologies have their limits, of course. Such technologies, sometimes used to establish “content zoning,” present serious privacy concerns, may adversely affect innovation, may unduly place burdens on Internet intermediaries and have

formation on the Internet becoming steadily weaker, as technology advances); Timofeeva, supra note 88, at 220 (“Various tools exist to identify the geographical location of the user and many companies routinely employ these tools for targeted advertising purposes.”).


144. See Justice S. Muralidhar, Jurisdictional Issues In Cyberspace, 6 INDIAN J. L. & TECH. 1, 3 (2010) (“Even while it was thought that one could fix the physical location of the computer from where the transaction originates and the one where it ends, that too can be bypassed or ‘masked’”).

145. See Yulia A. Timofeeva, Establishing Legal Order in the Digital World: Local Laws and Internet Content Regulation, 1 J. INT’L COMMERCIAL L. 41, 43 (2006) (content “zoning” consists of technical procedures to “direct information flows to particular users only; “[m]etaphorically, it can be described as creating zones in cyberspace that are open for some categories of users and closed for others”).


148. Etienne Montero & Quentin Van Enis, Enabling Freedom of Expression in Light of Filtering Measures Imposed on Internet Intermediaries: Squaring the Circle?, 27 COMPUTER L. & SEC. REV. 21 (2011) (“It is not always simple to identify the authors of illegal or harmful content in an open digital environment, global in scale, where it is easy to operate from abroad and/or anonymously. On the other hand, intermediary providers involved in transmitting or storing the disputed content
been used in some instances in the service of government repression. At least in theory, such concerns could be addressed by a jurisdictional standard that also considers whether the use of geo-location technology is mandatory or merely permissible. Concerns regarding geo-location technology may also be addressed via the development of appropriate technology.

V. CONCLUSION

Despite cultural divisions between the European Union and the United States on the substance of privacy rights and the reach of jurisdiction over Internet intermediaries are known and clearly identified, close to the victim, and generally solvent.”); Mark MacCarthy, What Internet Intermediaries Are Doing About Liability and Why It Matters, 25 BERKELEY TECH. L.J. 1039 (2010) (noting examples of implementation of Internet regulations through payment intermediaries).

149. Laura DeNardis, The Emerging Field of Internet Governance, Yale Info. Soc. Working Paper at 11, Sept. 17, 2010, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1678343 (“Freedom of expression and association are increasingly exercised online and institutional, governmental, and private decisions about Internet architecture can determine the extent of these freedoms as well as the degree to which online interactions protect individual privacy and reputation. . . . Technical measures such as content filtering, digital rights management techniques, and blocking access to web sites are techniques that repressive governments can use to ‘govern’ the flow of information on the Internet.”); Jessica E. Bauml, It’s a Mad, Mad Internet: Globalization and the Challenges Presented by Internet Censorship, 63 FED. COMM. L.J. 697 (2010) (reviewing censorship problems associated with Internet technologies that permit identification of location of Internet users). But see Patricia Moloney Figliola, Casey L. Addis & Thomas Lum, U.S. Initiatives to Promote Global Internet Freedom: Issues, Policy and Technology, CRS Rep. 7-5700 at Appendix B, Apr. 5, 2011, http://fpc.state.gov/documents/organization/140637.pdf (listing and explaining technical means to circumvent government censorship of web-based communications, including use of proxy servers).

150. Reidenberg, supra note 88, at 1956 (suggesting that website purveyor may be “purposely avoiding” itself of entry into jurisdiction “whenever content is posted without geolocation filtering”); Adam D. Thierer & Clyde Wayne Crews Jr., Internet Libel Ruling: Talk About a Kangaroo Court, Dec. 16, 2002, http://www.cato.org/publications/techknowledge/internet-libel-ruling-talk-about-kangaroo-court (suggesting that Internet vendors and publishers may be able to avoid confrontations with foreign courts and regulators by “using new geographic location technologies to better target their services instead of just blasting their materials out to the planet”).

151. On one view, if a website purveyor makes use of geo-location technology to target specific countries, it may be subject to jurisdiction. See Henn, supra note 140, at 175 (“[A] web site that uses software technology to target advertising toward the specific user should also be considered to have submitted to the jurisdiction of the specific user”). On another view, if a website purveyor does not use geo-location technology to exclude specific countries, then it might be subject to general jurisdiction in all countries. See Reidenberg, supra note 88, at 1953 (“[M]ore sophisticated computing enlists the processing capabilities and power of users’ computers. This interactivity gives the victim’s state a greater nexus with offending acts and provides a direct relationship with the offender for purposes of personal jurisdiction and choice of law”).

net-related activities, a process of “convergence” in views seems almost inevi-
table.153 The means for implementing consensus views on subjects such as the
to be forgotten may vary according to the perceived needs and political
practicalities of the two regions. 154 Often, consensus is best developed, at least
in the first instance, through “soft law” guidelines.155 Such guidelines may per-
mit experimentation, feedback, and revision to respond to developments in tech-
nology and business practices.156

The Internet and regulations surrounding it have matured greatly over the
past generation.157 The wide range in types of data transfers across international
borders that occur daily might give rise to different problems that require diffe-

153. See Reed, supra note 126, at 6 (noting “natural process of convergence” in international
law, supported by a “desire to achieve the benefits of global communication and commerce,” and
developed in part through comparisons between laws, where one country may use another’s law as a
“template” for parallel legislation). The “convergence” process, however, can be lengthy, “particu-
larly for new and fast-moving areas of technology like cyberspace.” Id. at 7; see also, Gehan
INFO. TECH. 147, 149 (2007) (“Despite significant convergence in global information privacy
norms, the difficulties resulting from trans-border data flows represents a further concern insuffi-
ciently dealt with by existing privacy norms”).

154. See Ivana Deyrup, Responses to Questions Posed by CNAS on International Law & Inter-
net Freedom, Mar. 2011, http://www.law.harvard.edu/students/orgs/nsrc/CNAS-Final%20Draft-
3.pdf (“There are many different potential models for international normative regimes to promote
Internet freedom, including existing international human rights norms, international treaties and or-
organizations, industry self-regulation, and domestic legislation that directs the conduct of American
corporations internationally.”); see also, Sean Flynn, ACTA’s Constitutional Problem: The Treaty
That Is Not a Treaty (Or An Executive Agreement), Mar. 1, 2011,
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1982091 (noting variety of international agree-
ment forms that US may constitutionally use).

155. See Andrew Power & Oisin Tobin, Soft Law for the Internet, Lessons from International
Law, 8 SCRIPT-ED 1, 32, 35 (2011) (“soft law consists of those informal rules that are non-binding,
but, due to cultural norms or standards of conduct, have practical effect;” “[s]oft law offers an effec-
tive way to deal with uncertainty, especially when it initiates processes that allow actors to learn
about the impact of agreements over time,” areas in the international arena where a soft law ap-
proach has worked include forestry, labor rights and sustainable development); François Nawrot,
Katarzyna Syska & Przemyslaw Switalski, Horizontal Application of Fundamental Rights: Right to
guidelines, outlining “core principles to protect privacy and personal data,” which are considered
“soft-law”).

156. See Charles F. Sabel & Jonathan Zeitlin, Experimentalism in Transnational Governance:
Emergent Pathways and Diffusion Mechanisms, Mar. 2011,
http://www2.law.columbia.edu/sabel/papers.htm (noting that “experimentalist regimes” for transna-
tional regulation appear to be emerging and suggesting means to encourage this phenomenon).

157. See John Palfrey, Four Phases of Internet Regulation, 77 SOC. RES. 3 (2010) (tracing de-
development of Internet regulation, from “open” Internet to today’s structure, where “access [may be]
contested” by regulators and private parties); Debora L. Spar, Ruling the Waves: Cycles Of Discov-
ery, Chaos, and Wealth from the Compass to the Internet (2001) (noting sequence of innovation,
commercial exploitation, creative anarchy, and eventual government regulation, in systems of new
technology, including the Internet).
ent solutions at different times.\textsuperscript{158} In the end, an approach to regulation based on careful attention to technology and business developments,\textsuperscript{159} coupled with genuine respect for cultural differences, is most likely to produce satisfactory, workable international solutions.\textsuperscript{160} Inaction, however, is not an option as the conflict has already manifested itself in the tensions that exist between the approach to regulation taken in the European Union and the approach taken in the United States.\textsuperscript{161}


\textsuperscript{159} Cass R. Sunstein, \textit{Constitutional Caution}, 1996 U. CHI. LEGAL F. 361, 374-75 (1996) (“In [period[s] of rapid change and technological uncertainty, in which those schooled in law are likely to be ignorant, there is much room for tentative, narrow judgments.”)

\textsuperscript{160} Nicola Lucchi, \textit{Access to Network Services and Protection of Constitutional Rights: Recognizing the Essential Role of Internet Access for the Freedom of Expression}, 19 CARDOZO J. INT’L & COMP. L. 1, 3 (2011) (attempts to regulate Internet content are “often criticized for the inability to reconcile technological progress, protection of economic interests and other interests that might conflict”); Evgeny Morozov, \textit{Whither Internet Control?}, 22 J. DEMOCRACY 62 (2011) (Internet control schemes may threaten both privacy and freedom of expression; caution appropriate); Philip A. Wells, \textit{Shrinking the Internet}, 5 N.Y.U. J. L. & Libr. 531, 532 (2010) (noting “strong temptation to fill the enforcement vacuum [on the Internet] with enhanced government intervention;” that approach may be “misguided” in producing “costly, oppressive” regulations).

\textsuperscript{161} See Frayer, \textit{supra} note 10 (“These problems will absolutely continue to come up, until one of two things happens: either the technology companies begin to build architectures that enable compliance with existing law, or the law begins to change.”) (quoting Joel Reidenberg of Fordham Law School).
VI. POST SCRIPT

As this Article went to press, the European Commission (the executive body within the European Union) issued a proposal to revise the 1995 EU Data Protection Directive. The proposal included a provision for recognition of the “right to be forgotten.”

Valuation In Investor-State Arbitration: Toward A More Exact Science

Joshua B. Simmons

Recommended Citation

Valuation in Investor-State Arbitration:
Toward A More Exact Science

By
Joshua B. Simmons*

INTRODUCTION

The stakes of international arbitration are rising. In cases between investors and states,1 at least seven arbitral awards have topped one hundred million dollars in the past five years,2 while several pending cases involve claims for billions of dollars.3 The arbitrators who preside over these high-stakes cases typically confront complex and divergent calculations of damages, which they may be ill equipped to reconcile. A long-standing refrain for determining damages is

*Josh Simmons is an Associate at Covington & Burling LLP. The views expressed in this Article are his own; they do not represent the views of Covington or its clients. Thanks to Professor John Norton Moore for the invitation to present this Article at the Sokol Colloquium held on April 5, 2011, at the University of Virginia School of Law.

1. Arbitration between foreign investors and host states (“investor-state arbitration” or “international investment arbitration”) is distinct from international commercial arbitration between two private parties.


that it is not an “exact science.” But increasingly the legitimacy of international arbitration depends on the well-explained and financially sound resolution of valuation disputes.

The damages phase of investor-state arbitration presents a variety of challenges, particularly when fair market value applies as the standard for calculating damages. Under customary international law, a fundamental principle of reparation is to “wipe out all the consequences of the illegal act.” This Article focuses on compensation of fair market value as a means of achieving such reparation, as opposed to restitution, contractual formulas, or moral damages. That is, the focus is on determining “the price that a willing buyer would pay to a willing seller in circumstances in which each had good information, each desired to maximize his financial gain, and neither was under duress or threat.”

---


6. See Draft Articles on Responsibility of States for Internationally Wrongful Acts, in Report of the International Law Commission on the Work of its Fifty-Third Session 56 U.N. GAOR Supp. (No. 10) at art. 35, U.N. Doc. A/56/10 (2001) [hereinafter ILC Articles on State Responsibility] (“A State responsible for an internationally wrongful act is under an obligation to make restitution, that is, to re-establish the situation which existed before the wrongful act was committed, provided and to the extent that restitution: (a) is not materially impossible; (b) does not involve a burden out of all proportion to the benefit deriving from restitution instead of compensation.”). In the event that restitution cannot be made, a state must provide compensation for the damage caused by its internationally wrongful act. See id. at art. 36.

7. Damages in many international commercial arbitration cases are determined pursuant to a provision in the underlying contract. Still, the challenges of determining damages in international commercial disputes are often analogous to those addressed in this Article, because international law allows for the recovery of lost profits in breach of contract cases. See infra nn.150-153 and accompanying text; see also John Y. Gotanda, Recovering Lost Profits in International Disputes, 36 Geo. J. Int’l L. 61, 63, 86, 94 n.180 (2004) [hereinafter Gotanda, Lost Profits].


The question of fair market value poses notable challenges for arbitrators because it relates more closely to finance than law. These challenges loom large because arbitrators frequently must determine fair market value in investor-state arbitration.10

The negative consequences of inaccurate and opaque valuations can extend to the entire arbitral system, beginning with lost confidence in awards and unreliable expectations about future cases. For example, unexplained large awards may strain the budget of a developing country and give rise to significant domestic political opposition.11 When facing inconsistent damages awards across similar arbitrations, state governments may also “find themselves in an untenable position of explaining to taxpayers why they are subject to damage awards for hundreds of millions of U.S. dollars in one case but not another.”12 Conversely, investors would not submit claims to arbitration if they could not expect an outcome worth the risk and the costs of pursuing arbitration.13 Without reasonable investor confidence in the system, the benefits of investor-state arbitration would begin to disappear.

Predictable, accurate valuations are necessary not only for the confidence of parties in arbitration proceedings, but also for the efficiency of international investment law. Before adopting measures that harm an investment, states should be able to weigh the benefits of such measures against their expected costs. Compensation of fair market value deters inefficient state actions.14 By the
same logic, states would have an incentive for “efficient breaches” only if investors do not receive more than fair market value.\textsuperscript{15} An expropriation will be efficient only if the state (and its people) gain more from the taking than the cost of fully compensating for the value of the expropriated investment.\textsuperscript{16}

Scholarly guidance regarding the calculation of damages in arbitration has developed quickly in recent years,\textsuperscript{17} highlighting the growing awareness of “one of the least understood and most unpredictable areas of international investment law.”\textsuperscript{18} The emerging literature has principally addressed the mechanics of quantifying damages, rather than the effect of damages determinations on the perceived legitimacy of international arbitration. Likewise, the literature regarding the legitimacy of international arbitration has not targeted the issue of valuation.\textsuperscript{19} This Article begins to fill this gap in the literature. In doing so, this Article takes on what has been described as the “urgent matter” of “encouraging the convergence of methods of calculating awards toward uniform and appropriate unlawful expropriation cases (and sometimes for drawing the line between the two, see id. at 255-57).

\textsuperscript{15} This assumes that “efficient breaches” would be desirable in some circumstances. For an argument in support of that position, see Louis Wells, Double Dipping in Arbitration Awards? An Economist Questions Damages Awarded to Karaha Bodas Company in Indonesia, 19 ARB. INT’L 471, 478 n.23 (2003) (“[E]xcessive awards discourage government takings, or breach of contract, when such actions are in fact efficient and thus desirable.”) (citing RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW (3d ed. 1986)). Such “efficiency” concerns are particularly important when states do not have greater protections of property under domestic law than under international law.

\textsuperscript{16} In other words, Pareto optimality depends on arbitrators accurately determining (and states paying) the fair market value of an expropriated investment. See id. at 473, 480. Admittedly, the value of a state’s “gain” may be more difficult to predict than the costs of compensation. But the more accurately a state predicts costs, the more able it will be to act efficiently. States’ actual payment of damages awarded in international arbitration is outside the scope of this Article.

\textsuperscript{17} As recently as 2005, an arbitrator might find relatively limited guidance on damages in treatises, and damages were said to be the “neglected aspect.” See Geoffrey Beresford Hartwell et al., Assessing Damages—Are Arbitrators Good At It? Should They Be Assisted by Experts? Should They Be Entitled to Decide ex aequo et bono? Some War Stories, 6 J. WORLD INV. & TRADE 7, 17 (2005) (comments by Serge Lazareff); see also Thomas R. Stauffer, Valuation of Assets in International Takings, 17 ENERGY L.J. 459, 460 (1996) (noting that the “economic dimension—the ‘quantification of the quantum’—[had been] given short shrift”). In the few years since then, three books have been published on the subject of damages in international arbitration. See IRMGARD MARBOE, CALCULATION OF COMPENSATION AND DAMAGES IN INTERNATIONAL INVESTMENT LAW (2009); MARK KANTOR, VALUATION FOR ARBITRATION (2008); RIPPINSKY & WILLIAMS, supra note 5. The increasing focus on damages includes a “new interest in interest.” John Y. Gotanda, A Study of Interest, in INTEREST, AUXILIARY AND ALTERNATIVE REMEDIES IN INTERNATIONAL ARBITRATION, DOSSIER V OF THE ICC INSTITUTE OF WORLD BUSINESS LAW 169 (ICC Publication No. 684, 2008).

\textsuperscript{18} NOAH RUBINS & NORMAN KINSSELLA, INTERNATIONAL INVESTMENT POLITICAL RISK AND DISPUTE RESOLUTION: A PRACTITIONER’S GUIDE 258 (2005).

\textsuperscript{19} See, e.g., Franck, Legitimacy Crisis, supra note 12, at 1586-87 nn.326-29 (summarizing scholarship regarding the larger problem of “the ability to determine with certainty the respective rights and obligations of investors and Sovereigns in a given situation”). One complaint lodged against investor-state arbitration is that awards are “unpredictable . . . and extremely generous to foreign corporations.” Asha Kaushal, Revisiting History: How the Past Matters for the Present Backlash Against the Foreign Investment Regime, 50 HARV. INT’L L.J. 491, 510 (2009).
Part II sets forth the framework of challenges facing investor-state arbitration, within the context of its continuing expansion. Part III links those challenges to questions about the legitimacy of valuation. The limited financial and economic experience of most arbitrators plants a seed of doubt regarding valuation. That doubt grows because of perceptions that arbitrators merely “split the baby” between the parties’ proposed valuations, particularly when awards are poorly explained. This Article’s study of recently published decisions involving valuation corroborates those perceptions, because the ratio of the amount awarded to the amount claimed usually falls within the range of one-fifth to one-half. In addition, several annulment petitions demonstrate parties’ frustrations with opaque (“black box”) determinations of fair market value.

Part IV addresses two fundamental aspects of valuation. First, it discusses the issue of awarding interest. The recent trend toward awarding compound interest illustrates how convergence on damages methodologies furthers the legitimacy of arbitral awards. Second, Part IV describes perhaps the most prominent method of determining fair market value: discounted cash flow (DCF) analysis. Despite theoretical agreement on the DCF method of valuation, some tribunals exhibit lingering reluctance to apply the DCF method in practice.

Part V suggests that tribunals should not reject a well-pleaded DCF analysis simply on the basis of “uncertainty,” “speculation,” and “going concern” tests. Several components of DCF analysis can address uncertainty and allow for transparent resolution of the parties’ competing positions. In connection with more frequent usage of the DCF method, tribunals should be more willing to appoint an independent financial expert. As a few recent decisions show, tribunal-appointed experts are helpful guides for valuation and can augment the accuracy and transparency of arbitral awards. When billions of dollars are at stake, it is worth the cost of an independent financial expert to enhance the legitimacy of investor-state arbitration.

I. FRAMEWORK OF THE LEGITIMACY DEBATE

The past decade has witnessed a well-documented growth spurt of international investment arbitration. States typically consent to such arbitration through investment treaties, which designate the rules that will govern dispute resolution. Empowered by thousands of investment treaties—mostly in the form of bilateral investment treaties (BITs)—investors have increasingly submitted disputes against states to arbitration. By the end of 2010, the number of total known investor-state disputes submitted to arbitration was 390, at least twenty-five more than in 2009.21

21. See United Nations Conference on Trade and Development, Latest Developments in Inves-
Most recent, publicly available awards in investor-state disputes have been decided under the auspices of the International Centre for Settlement of Investment Disputes (ICSID). ICSID has witnessed a particularly sharp increase in filings over the past decade, spreading across most regions of the world and economic sectors. Investor-state arbitration has grown under many other rule systems as well, although growth is difficult to measure because, unlike in ICSID cases, the existence of such arbitrations may remain confidential and unknown to the public.

A consequence of this growth is an increased scrutiny of the legitimacy of international investment arbitration. The numerous cases brought against Argentina have been a source of criticism, and the recent withdrawals of Ecuador and Bolivia from the ICSID Convention demonstrate further concern about the system. Challenges to legitimacy include inconsistent decisions, preferential

---

22. ICSID cases constituted 245 of the 390 known disputes by the end of 2010. See id. at 1-2; see also Jeffrey P. Commission, A Citation Analysis of Developing Jurisprudence, 24 J. INT’L ARB. 129, 130 n.10 (2007).

23. See ICSID – INTERNATIONAL CENTRE FOR SETTLEMENT OF INVESTMENT DISPUTES, THE ICSID CASELOAD — STATISTICS 7 (2011) [hereinafter ICSID Caseload Statistics] (showing the number of cases registered each year). From a mere handful of cases in the 1970s, 1980s, and early 1990s, over 250 cases have been filed in just the past ten years. Although ICSID filings seemed to peak in 2007 (thirty-seven cases registered), the number of filings has remained steadily above twenty per year since then.

24. See id. at 11-12.

25. Other arbitral institutions with independent rule systems include the United Nations Commission on International Trade Law (UNCITRAL), the International Chamber of Commerce (ICC), the International Centre for Dispute Resolution of the American Arbitration Association (AAA/ICDR), the London Court of International Arbitration (LCIA), and the Stockholm Chamber of Commerce (SCC).


27. See, e.g., Burke-White & von Staden, supra note 26, at 284-85, 297-301 (“The divergent decisions in these cases, their often strained legal reasoning, and the exceptional sums awarded to claimants against Argentina have called into question, at least in the eyes of some states, the legitimacy of the ICSID system.”); Kathryn Khamsi, Compensation for Non-Expropriatory Investment Treaty Breaches in the Argentina Gas Sector Cases: Issues and Implications, in THE BACKLASH AGAINST INVESTMENT ARBITRATION 165 (Michael Waiibel et al., eds., 2010).

treatment of foreign investors over domestic investors, bias toward investors, arbitrator independence, and intrusions on the regulatory sovereignty of states. Despite these challenges, the pace of investor-state arbitration has shown little sign of slowing.

A. The Endurance of Investor-State Arbitration

States and investors both stand to benefit from international arbitration. For states, the benefits of investment treaties have differed historically between developed and developing nations. Developed, capital-exporting nations can secure protection for their investors and avoid the challenges of diplomatic protection by ensuring that their investors have a private right of action to arbitration. Thanks in part to the enforcement mechanism of arbitration, investment treaties also contribute to broad-scale liberalization. At the domestic level, investment treaties support liberalization because their protections limit government interference in markets. Similarly, by protecting property rights and providing mechanisms of dispute resolution, investment treaties can substitute for poor in-

CONVENTION AND BITS: IMPACT ON INVESTOR-STATE CLAIMS 1, 1 (2010); see also Ignacio A. Vincentelli, The Uncertain Future of ICSID in Latin America, 16 LAW & BUS. REV. AM. 409, 410 (2010).


30. The geopolitical implications of those differences have been a source of much debate. See, e.g., Jeswald W. Salacuse, The Emerging Global Regime for Investment, 51 HARV. INT’L L.J. 427, 434-35 (2010); Garcia, supra note 26, at 314-17. In the 1960s and 1970s, developing countries took a strong position against the Hull Rule (the requirement of “prompt, adequate, and effective” compensation for a taking of property) such that it could no longer be relied on as customary international law. See Andrew T. Guzman, Why LDCs Sign Treaties that Hurt Them: Explaining the Popularity of Bilateral Investment Treaties, 38 VA. J. INT’L L. 639, 646-51 (1998). Paradoxically, by entering into investment treaties that require prompt, adequate, and effective compensation, developing countries have reestablished the fundamental principle of the Hull Rule. See id. at 666-69; see also Kaushal, supra note 19, at 501.

31. See, e.g., Salacuse, supra note 30, at 439-40, 459-60; see also id. at 462-63 (“Prior to the institution of investor-state arbitration, governments had to deal with their nationals seeking diplomatic protection and other forms of interventions [which] entailed significant diplomatic, political, and economic costs for home governments . . . .”).


33. See Salacuse & Sullivan, supra note 32, at 92-94.
stitutional quality and a weak rule of law in a host state.34 To the extent such advances limit state sovereignty, they may prove controversial and ultimately require states to pay an unpredictable price to exercise their sovereignty.35

Developing countries agree to these limitations because they reasonably believe that investment treaty protections signal to foreign investors that the country will provide a measure of stability and protection. Econometric studies suggest that such signaling has promoted foreign investment, although the evidence is somewhat mixed.36 Regardless of the actual effectiveness of BITs, developing countries seek to attract foreign investment when they enter into investment treaties.37 For example, developing nations may perceive that investment treaties are necessary for making credible commitments to investors, particularly in light of growing competition for foreign direct investment among developing nations.38 Developing countries enhance the credibility of their commitments by consenting in investment treaties to arbitration of their disputes with foreign investors, and by establishing fair market value as a standard of compensation.

As a result of this international bargain, investors have gained standing to pursue arbitration against states. Setting aside the nuanced question of investors’ considerations when deciding whether to invest in a particular state, the growing number of arbitration claims shows that investors are becoming savvier to the

34. UNCTAD, Attracting FDI, supra note 32, at 16-17. Investment treaties “may contribute to the coherence, transparency, predictability and stability of the investment frameworks of host countries.” Id. at 25-26; see also Kaushal, supra note 19, at 517 (describing how “the international nudges the national toward convergence on a high level of investment protection”).


36. “Whereas the findings of early empirical studies on the impact of BITs on [foreign direct investment (FDI)] flows were ambiguous, with some showing weak or considerable impact (and one or two no impact at all), more recent studies published between 2004 and 2008—based on much larger data samples, improved econometric models and more tests—have shifted the balance towards concuring that BITs appear to have an impact on FDI inflows from developed countries into developing countries.” UNCTAD, Attracting FDI, supra note 32, at 29-55; see also Jason W. Yackee, Do Bilateral Investment Treaties Promote Foreign Direct Investment? Some Hints from Alternative Evidence, 51 VA. J. INT’L L. 397, 405-14 (2010) (describing reasons why “analysts have had great difficulty reliably demonstrating a statistically significant, substantively meaningful correlation between BITs and FDI”); Susan D. Franck, Empiricism and International Law: Insights for Investment Treaty Dispute Resolution, 48 VA. J. INT’L L. 793 n.116 (2008); Salacuse & Sullivan, supra note 32, at 96, 111-12; Kenneth J. Vandevelde, The Economics of Bilateral Investment Treaties, 41 HARV. INT’L L.J. 469, 489, 498 (2000).

37. See UNCTAD, Attracting FDI, supra note 32, at 29 (“Developing countries have concluded BITs as part of their desire to improve their policy framework in order to attract more FDI and benefit from it.”).

38. See Guzman, supra note 30, at 669-71.
possibility of international arbitration. Evidence of investors’ heightened awareness of international arbitration can also be seen in the growing and accepted practice of restructuring an investment in order to acquire investment treaty protection.40

Today, if a government measure harms a foreign investment, a foreign investor will likely have greater confidence in international arbitration than in domestic courts of the host state.41 Yet the confidence of investors in international arbitration cannot be taken for granted. A recent survey suggests that major corporations are skeptical about the protections provided by investment treaties.42 Particularly with respect to compensation, investors may have a growing sense that international arbitration is unlikely to result in adequate compensation for the harms caused by host states.43

B. Determinacy, Transparency, and Consistency

The legitimacy of international arbitration depends in large part on investors and states having reliable expectations and confidence in the resolution of their disputes through arbitration. That confidence can develop or deteriorate in a variety of ways. Two critical factors are the determinacy and coherence of investment arbitration decisions, which help ensure that the rules governing arbitration “convey clear and transparent expectations” to the parties.44 Arbitral tri-

39. For a summary of “signs that investor awareness about BITs is increasing,” see UNCTAD, Attracting FDI, supra note 32, at 53.

40. See Barton Legum, Defining Investment and Investor: Who Is Entitled To Claim?, ICSID, OECD, UNCTAD Symposium on Making The Most Of International Investment Agreements (Dec. 12, 2005), at 5 (“T]he emergence of th[e] [BIT-shopping] industry suggests that, perhaps for the first time, BITs really are beginning to encourage and promote foreign investment in the way they were intended to do.”); see also Anthony Sinclair, ICSID’s Nationality Requirements, in INVESTMENT TREATY ARBITRATION AND INTERNATIONAL LAW 85, 116 (T.J. Weiler ed., 2008).

41. Two principal reasons for this confidence are the belief in a (more) fair proceeding and the enforceability of an award in international arbitration. Under the ICSID Convention, for example, awards become immediately enforceable in the host state’s courts. See Convention on the Settlement of Investment Disputes Between States and Nationals of Other States, art. 54(1) (Mar. 18, 1965) 17 U.S.T. 1270, 575 U.N.T.S. 159 [hereinafter ICSID Convention] (“Each Contracting State shall recognize an award rendered pursuant to this Convention as binding and enforce the pecuniary obligations imposed by that award within its territories as if it were a final judgment of a court in that State.”).

42. Jason W. Yackee conducted a survey of the general counsels of the top 200 U.S. corporations on the Fortune 500 list. The results of the survey “indicate a low level of familiarity with BITs, a pessimistic view of their ability to protect against adverse host state actions, and a low level of influence over FDI decisions.” Yackee, supra note 36, at 429; see also id. at 429-30 (reporting that representatives of major corporations “did not view BITs as particularly effective at protecting against expropriation” and had “skepticism about the ability of BITs to protect against regulatory change”); Salacuse & Sullivan, supra note 32, at 96 (“Local economic conditions and government policies are probably more important than BITs in influencing the investment decision.”).

43. See infra Part III; Yackee, supra note 36, at 434-35; Franck, Empirically Evaluating, supra note 35, at 49-50.

44. See Franck, Legitimacy Crisis, supra note 12, at 1584 (“Legitimacy depends in large part
bunals should be consistent in their interpretation and application of those rules.\textsuperscript{45} Before turning to the detrimental impact of inconsistencies, the overarching trend of convergence should be explained.

Investment treaties do not provide substantial determinacy; they typically set forth only brief and basic rules.\textsuperscript{46} For example, most investment treaties say nothing at all about the appropriate damages for non-expropriatory treaty violations.\textsuperscript{47} The standard of compensation for expropriation—while referring to the payment of fair market value—also lacks detail in most investment treaties.\textsuperscript{48} A common provision for lawful compensation is payment of “prompt, adequate, and effective” compensation of “fair market value.”\textsuperscript{49} That sample provision includes no instruction on numerous calculation issues, including the appropriate methodology for determining fair market value or whether to compound interest.

Despite the indeterminacy of investment treaties, the publication of more and more investment arbitration decisions has furthered transparency and improved the reliability of parties’ expectations. Transparency can come in a number of forms.\textsuperscript{50} The parties in investor-state arbitration usually consent at least to the publication of decisions, which are available on a number of websites.\textsuperscript{51} When a state is a party to arbitration, the public has a strong interest and expect-
The trend toward publishing awards in investor-state arbitration (and particularly ICSID arbitration) is therefore well established, and it has enabled international lawyers and academics to contribute to the development of international investment law.

The swell in publicly available decisions has contributed to the influence of precedent as a de facto reality in international arbitration. As a matter of law, arbitral tribunals are not bound by prior decisions, and international arbitration is not a system of common law. Yet parties frequently cite to, and arbitrators often rely on, previously published decisions. This has led to the emergence of “case law” and “jurisprudence” in international investment law, as well as to the characterization that investment treaties constitute an international “regime.” As Jeffrey Commission explains, “[g]iven that international investment law now principally develops through case law, the precedential value of each decision, award, and order, is, rightly or wrongly, tremendously significant.”

International arbitration therefore serves as an accelerated, if disaggregated, forum for international investment law. A recognized goal of that forum is to build a coherent body of law.

Notwithstanding the influence of precedent, tribunals have reached conclusions that conflict directly with prior decisions. Criticisms of inconsistency have focused on the most-favored-nation principle, the excuse of necessity, and regulatory.

52. See Coe, Jr., supra note 50, at 1353.
53. See, e.g., Salacuse, supra note 30, at 466-67; Franck, Legitimacy Crisis, supra note 12, at 1614-16; see also Coe, Jr., supra note 50, at 1356-57; J. Anthony VanDuzer, Enhancing Procedural Legitimacy of Investor-State Arbitration through Transparency and Amicus Curiae Participation, 52 McGill L.J. 681, 706-08 (2007).
56. See Alec Stone Sweet, Investor-State Arbitration: Proportionality’s New Frontier, 4 LAW & ETHICS HUM. RTS. 47, 60-61 (2010); Commission, supra note 22, at 129; see also W. Mark C. Weidemaier, Toward A Theory of Precedent in Arbitration, 51 WM. & MARY L. REV. 1895, 1908 (2010) (“Through this engagement with past awards, ICSID tribunals have gradually fashioned what has been called an investment treaty ‘case law or jurisprudence.’”).
57. See Salacuse, supra note 30, at 436 (“[L]awyers and arbitrators . . . implicitly treat investment treaties as constituting a regime in that they regularly refer to prior decisions applying one treaty in order to interpret a wholly separate treaty.”).
58. See Commission, supra note 22, at 131.
In the context of damages, there have been fewer studies of inconsistencies, despite a “common perception [of] a lack of a coherent systematic approach to compensation issues.” As discussed in Part IV, one prominent inconsistency in the realm of damages is the appropriate methodology for valuation.

Inconsistent decisions can be debilitating to the legitimacy of investor-state arbitration. As Susan Franck has explained, “[i]nconsistency tends to signal errors, lends itself to suggestions of unfairness, creates inefficiencies, and generates difficulties related to coherence, most notably a lack of predictability, reliability, and clarity.” In the words of Nigel Blackaby, the co-existence of “diametrically opposed decisions . . . shock[s] the sense of rule of law or fairness.” Despite those potentially dire consequences, it seems a stretch to claim that “chaos reign[s]” in investor-state arbitration because of inconsistent decisions. A more accurate understanding is that, despite several conflicting decisions, international investment law is converging around the growing body of precedent reflected in public decisions. The legitimacy critiques arising from inconsistencies are a healthy sign of pressure toward continuing harmonization. Such convergence is critical in the area of valuation, given the notable challenges that valuation presents in investor-state arbitration.

60. See e.g., Karl, supra note 26, at 236-37; Gabriel Egli, Don’t Get Bit: Addressing ICSID’s Inconsistent Application of Most-Favored-Nation Clauses to Dispute Resolution Provisions, 34 PEPP. L. REV. 1045 (2007); Franck, Legitimacy Crisis, supra note 12, 1559-82 (describing inconsistencies in the Lauder arbitration, the SGS cases, and three NAFTA cases); see also Burke-White & von Staden, supra note 27, at 297 (“ICSID arbitrations have generated a contradictory jurisprudence that lacks theoretical coherence and remains tied to the private law origins of international arbitration. The Argentine cases are illustrative of the problematic jurisprudence to date.”).

61. RIPINSKY & WILLIAMS, supra note 5, at xxxv (noting that the lack of coherence “contributes to the uncertainty of the legal environment and the unpredictability of outcomes of disputes.”). As another commentator put it, there “appears to be strikingly little uniformity in the calculations of awards in international arbitrations.” Wells, supra note 15, at 478.

62. Franck, Do Investment Treaties Have a Bright Future, supra note 51, at 63-67 (citations omitted). See also Karl, supra note 26, at 236-37; Franck, Legitimacy Crisis, supra note 12, at 1558 (“Inconsistency creates uncertainty and damages the legitimate expectations of investors and Sovereigns.”).

63. Franck, Legitimacy Crisis, supra note 12, 1583 (quoting Nigel Blackaby of Freshfields Bruckhaus Deringer).

64. Garcia, supra note 26, at 350-51.

65. See Karl, supra note 26, at 237 (“[A]s case law develops, future arbitration tribunals will have more precedents at hand, which should have a certain harmonising effect.”); Salacuse, supra note 30, at 467 (“Despite the decentralized and private decisionmaking processes of the [international investment] regime, the resulting decisions by arbitral tribunals demonstrate a surprisingly high degree of uniformity and consistency.”); Weidemaier, supra note 56, at 1944 (“[T]he system of precedent can be understood as a response by arbitrators to external critics whose objections threatened ICSID’s viability as a forum for resolving investment disputes,” and “ICSID arbitrators are remarkably well positioned to foster norms concerning their role as producers of law.”).
II. THE LEGITIMACY CHALLENGES OF VALUATION

Multiple valuation challenges threaten to undermine the legitimacy of investor-state arbitration, including the financial competency of arbitrators, awards perceived to “split the baby,” and poorly explained and inconsistent methodologies. By entering into treaties that give foreign investors a private right of action and establish fair market value as a basis for compensation, host states arguably have accepted that economics will prevail over politics in matters of compensation.\(^{66}\) Arbitrators seeking to determine fair market value in investor-state cases do not have the unfettered discretion to decide damages \textit{ex aequo et bono}, i.e., on equitable principles.\(^{67}\) They must engage in the task of valuation.

\textbf{A. The Financial Competency of Arbitrators}

The perceived competency of arbitrators impacts the legitimacy of their conclusions. For most of the challenging issues that are raised in investor-state arbitration, arbitrators are among the most experienced and adept persons in the world at settling the parties’ disputes.\(^{68}\) Arbitral tribunals decide, for example, on complex and dispositive questions of jurisdiction, liability, and standards of compensation. These legal issues are the bread and butter of international arbitrators. If a tribunal proceeds past those questions—that is, concludes that it has jurisdiction, that the state is liable, and that the appropriate standard of compensation is fair market value—then the tribunal must determine the value (or diminished value) of the investment in question. For many arbitrators, this is the “harder part”\(^{69}\) and “dangerous territory.”\(^{70}\)

\(^{66}\) This can be viewed as part of a move in the international order in which “economics replac[es] politics as law’s sidekick and nemesis.” See Burke-White & von Staden, \textit{supra} note 26, at n.14 (quoting David Kennedy, \textit{The International Style in Postwar Law and Policy}, 1994 \textit{Utah L. Rev.} 7, 63 (1994)). The measure of fair market value will nonetheless rarely limit a state’s freedom to exercise its sovereignty. See, e.g., id. at 288-89. The payment of fair market value could affect sovereignty if it pushed a state toward bankruptcy. See \textit{CME v. Czech Republic}, supra note 11; Ripinsky & Williams, \textit{supra} note 5, at 356. Of course, states remain free to opt out of value-based compensation standards (or to revise investment treaties accordingly).

\(^{67}\) “The term ‘value’ is an ‘objective concept with an economic content’ and . . . therefore, where the law prescribes compensation to be equivalent to the value of the asset taken, there is little room for the exercise of an equitable discretion.” Ripinsky & Williams, \textit{supra} note 5, at 128 (quoting E. Lauterpacht); see also Hartwell et al., \textit{supra} note 17, at 21 (comments of B. Hanotiau).

\(^{68}\) Many commentators have remarked upon the small, elite pool of arbitrators who preside over major international investment disputes. See, e.g., Weidemaier, \textit{supra} note 56, at 1950 (describing ICSID arbitrators as an “elite group”); Commission, \textit{supra} note 22, at 137-38.

\(^{69}\) Markham Ball, \textit{Assessing Damages in Claims by Investors Against States}, 16 ICSID REV.—FOR. INV. L.J. 408, 417 (2001); see also Christer Söderlund et al., \textit{The Valuation of Lost Profits—Finding it Right}, 6 \textit{J. World Inv. & Trade} 23, 31 (2005) (“Damages are one of the most challenging topics in arbitration.”).

In most modern investor-state arbitration proceedings, the complexity of valuation is beyond the traditional legal training of arbitrators. Nearly all arbitrators in investor-state proceedings hail from legal backgrounds, whether in private practice, government, or academia. A survey of the publicly available biographies of leading arbitrators reveals that none of those arbitrators have post-graduate degrees in the fields of finance, economics, or mathematics. Because of their backgrounds, arbitrators may be reluctant to immerse themselves in the detailed formulas and spreadsheets submitted by the parties. Legal training and analysis do not align well with the task of assessing fair market value. Even if arbitrators were fully able to acquire the necessary financial competency, they could not overcome the fact that they are neither economists nor financial analysts.

Arbitrators’ legal backgrounds lead to tainted perceptions of the quality of some damages awards. In other words:

Whether or not modern arbitrators are good at assessing damages is not, I suggest, territory by attempting to enter its own economic valuation into the findings of the respective economic experts’ opinions . . . .

The existence of an early investor-state decision with apparent “flawed economic reasoning” has not helped the case of arbitrators. See RIPINSKY & WILLIAMS, supra note 5, at 208-10 (summarizing criticisms of the valuation analysis in Amoco Int’l Fin. Corp. v. Iran, Partial Award, 15 Iran-U.S. Cl. Trib. Rep. 189 (1987)); see also id. at 190 (“Valuation can be a sophisticated exercise going beyond the expertise of the legal profession.”).

Arbitrators are very much a part of an international epistemic community with similar training and, in many cases, comparable backgrounds.”).


Arbitrators are very much a part of an international epistemic community with similar training and, in many cases, comparable backgrounds.”).

73. See infra Part V.B; Hartwell et al., supra note 17, at 11-12 (comments by Nicolas Ulmer) (“Are [a]rbitrators [g]ood at [a]ssessing [d]amages? Answer: no. . . . [W]hat lawyers like particularly, and what they are trained to like, is putting damages in categories, cutting them out, and deciding whether they are allowable, which is not the same thing as assessing them.”); see also J. Brian Casey et al., Arbitration and the Valuator, J. BUS. VALUATION 105, 112-13 (2007).

Legal backgrounds in no way foreclose arbitrators from having significant expertise in deciding valuation disputes. Experienced arbitrators have likely confronted a wide array of damages calculations, set forth in well-argued briefs and detailed expert reports. That experience could be as important as formal education. In addition, economics is far from a stranger to law, and arbitrators may well have studied economics because of the growing connection between the fields. For example, of the surveyed prominent arbitrators, at least one taught “business law and economics” (Lalonde) and another has a Ph.D. in international law from the London School of Economics (Orrego Vicuña). See INTERNATIONAL ARBITRATION INSTITUTE, http://www.iaiparis.com (last visited Oct. 21, 2011). There is also little doubt as to the capacity of arbitrators, over time, to learn what is needed for calculating damages in a given case. Yet parties to an arbitration might object to the efficiency of such on-the-job learning.
gest, the point. The question is whether parties can have confidence that the person assessing damages is properly qualified to do so, and I suggest that in general, with the very greatest respect to all my friends, the modern legal arbitrator is not so qualified for self-evident reasons. Legal and economic reasoning are different.76

Perceptions of arbitrators’ limited skills with respect to valuation are likely to persist, particularly if arbitrators do not properly apply and explain appropriate valuation methodologies.

B. The Perception of “Splitting the Baby”

Imagine a simple formula to resolve a disputed fair market value in international arbitration. The investor claims a valuation of X (hundreds of millions) dollars. The state asserts that Y (near-zero) dollars would be an appropriate award. The arbitral tribunal applies the following formula: Damages = (X + Y) / 2. Such a “split the baby”77 approach would be consistent and predictable but would wholly fail to inspire confidence. Fair market value is a fact-intensive (and often unique) determination, so the parties do not expect formulaic awards.78 Rather, parties expect consistent, well-founded methodologies and accurate valuations.

Yet baby-splitting is perceived as the reality in international arbitration.79 As explained below, evidence of this perception can be seen in parties’ failure to converge on valuation, and in the gulf between states’ non-quantified or zero-

76. Hartwell et al., supra note 17, at 8 (comments by Hartwell); see also, e.g., Thierry J. Sénéchal & John Y. Gotanda, Interest as Damages, 47 COLUM. J. TRANSNAT’L L. 491, 494 (2009) (describing arbitral tribunals as “unfamiliar with modern economic and financial principles”). If arbitrators can be criticized as “lack[ing] critical expertise in public law adjudication,” see Burke-White & von Staden, supra note 26, at 286, 330, parties have all the more justification for questioning arbitrators’ relative inexperience with modern principles of finance.

77. The phrase “split the baby” is an awkward fit to damages determinations. It is based on the Old Testament story of King Solomon threatening to split a baby to identify the true mother of a child. See Christopher R. Drahozal, Busting Arbitration Myths, 56 U. KAN. L. REV. 663, 673 (2007) (citing 1 Kings 3:16-28). The typical task of arbitrators is determining fair market value of property that has been taken by a state, not who gets to keep the property. Despite the imperfect fit, “splitting the baby” is a well-recognized characterization of tribunals seeking a middle ground between the positions of the parties.

78. See Franck, Do Investment Treaties Have a Bright Future, supra note 51, at 78.

damages positions and investors’ claims. One reason for this perception is the competency of arbitrators: “Given that assessment of damages . . . may be a complex exercise requiring knowledge of financial analysis and economic models, ‘splitting the baby’ may offer itself as an attractive option when tribunals get lost in the intricacies of valuation techniques.” Such a simplistic approach to valuation shows how the competency of arbitrators might cast doubt on the enterprise of valuation, thereby undermining the legitimacy of tribunals’ awards.

Assuming for a moment that tribunals in fact applied the formula above (Damages = (X + Y) / 2), recent published decisions suggest that the denominator should be higher than two. From mid-2006 to mid-2011, there were fifteen published investor-state decisions in which a determination of fair market value comprised a large portion of the tribunal’s damages analysis. Those fifteen cases are drawn from approximately thirty decisions published during that time in which tribunals reached the damages phase of the proceedings. This is a

80. In theory, a check on this type of divergence is the weighting of opposing damages figures by their plausibility. See Richard Posner, An Economic Approach to the Law of Evidence, 51 STAN. L. REV. 1477, 1539 (1999). Of course, that check depends on a good understanding of the plausibility of damages figures. Another proposal for limiting such divergence is for states to “put cost-shifting guidelines into investment treaties to reward investors whose claimed damages are in line with the ultimate award or provide deterrence for inflating claimed damages.” Franck, Empirically Evaluating, supra note 35, at 63.

81. RIPINSKY & WILLIAMS, supra note 5, at 122 (citing Ball, supra note 69, at 425-27). Compromise between arbitrators is another reason why awards might seem to “split the baby.” Such compromise awards can also undermine the legitimacy of valuations, because an award “not supported by articulated reasons does disservice to the credibility of the outcome.” Ball, supra note 69, at 427.

82. Canvassing awards prior to June 1, 2006, Susan D. Franck observed that the ratio of average amounts awarded to average amounts claimed was less than one-tenth. See Franck, Empirically Evaluating, supra note 35, at 58-60 (stating that, in twenty-one cases prior to June 1, 2006, in which tribunals awarded cash to an investor, the average amount awarded was $25.6 million and the average amount claimed was $345.5 million). That fraction does not distinguish between cases based on fair market value versus other forms of damages. Franck’s study also includes cases in which tribunals were not actually required to quantify damages, thus not addressing tribunals’ calculation of damages. See id. at 24-25, 58-60 (stating that, of the fifty-two cases she studied, there were thirty-one in which tribunals awarded investors nothing).

83. The decisions are Azurix v. Argentina Award, supra note 2; ADC v. Hungary, supra note 4; Siemens v. Argentina, supra note 2; Enron v. Argentina, supra note 2; Compañía de Aguas v. Argentina, supra note 2; Sempra v. Argentina, supra note 2; BG Group v. Argentina, supra note 2; Biwater Gauff (Tanz.) Ltd. v. Tanzania, ICSID Case No. ARB/05/22, Award (July 18, 2008) [hereinafter Biwater v. Tanzania]; Rumeli Award, supra note 2; Nat’l Grid P.L.C. v. Argentina, Award (UNCITRAL Nov. 3, 2008) [hereinafter Nat’l Grid v. Argentina]; Siag and Vecchi v. Egypt, ICSID Case No. ARB/05/15, Award (May 1, 2009) [hereinafter Siag v. Egypt]; Walter Bau v. Thailand, Award (UNCITRAL July 1, 2009) [hereinafter Walter Bau v. Thailand]; Kardassopoulos v. Georgia, ICSID Case Nos. ARB/05/18, ARB/07/15, Award (Feb. 28, 2010) (a revision proceeding is currently pending); RosInvest v. Russia, supra note 70; Alpha Projektholding GmbH v. Ukraine, ICSID Case No. ARB/07/16, Award (Oct. 20, 2010) [hereinafter Alpha v. Ukraine].

84. The selection of a case as a “fair market value” case required the exercise of discretion in some instances. This discretion was necessary because some decisions do not indicate whether the tribunal was determining fair market value. In addition, a few cases involving the valuation of a fungible asset or of real property have not been included.
significant body of awards, given that tribunals awarded damages in only twenty-one such cases between 1990 and June 2006. The average claimed valuation in the fifteen fair-market-value cases was $203 million. The respondents in these cases challenged the claimed valuations but rarely put forward quantifiable counter-valuations. The average valuation awarded was $80 million, roughly two-fifths the average amount claimed.

Averages serve as "blunt statistical instrument[s]" and hide important outliers. For example, in ADC Affiliate Ltd. v. The Republic of Hungary, ICSID Case No. ARB/03/16, Award of the Tribunal (Oct. 2, 2006) and Kardassopoulos v. Georgia, ICSID Case Nos. ARB/05/18, ARB/07/15, Award of the Tribunal (Feb. 28, 2010), the tribunals awarded exactly or almost exactly the amount claimed by the investors. In BG Grp. PLC v. Argentina, Final Award (UNCITRAL Dec. 24, 2008) and Rumeli Telekom A.S. v. Kazakhstan, ICSID Case No. ARB/05/16, Decision of the Ad Hoc Committee on the Application for Annulment, (Mar. 25 2010), the tribunals awarded over half of the valuation amount claimed. Yet in RosInvest Co. U.K. Ltd. v. Russian Federation, SCC Case No. 075/2009, Final Award, (Sept. 12, 2010) and Biwater Gauff (Tanz.) Ltd. v. Tanzania, ICSID Case No. ARB/05/22, Award (July 18, 2008), the tribunals awarded little to nothing compared to the amounts claimed. Still, in nine of the fifteen cases studied, tribunals awarded between one-fifth and one-half of the amounts claimed. This elementary empirical evidence corroborates the perception that arbitrators continue to “split the baby,” which is difficult to refute without published decisions that reveal a rigorous quantification of fair market

85. Franck, Empirically Evaluating, supra note 35, at 58.

86. For some cases, the average amount claimed is based on the mean of multiple claims, because investors used different valuation approaches and economic models to propose a range of valuations. See, e.g., Enron v. Argentina, supra note 2, ¶¶ 348-51. Where investors proposed valuations at different dates, the higher amount claimed was used. If decisions clearly delineated damages based on fair market value from other damages (such as historical losses, interest, and costs), the average amounts claimed and awarded are based only on the fair market value portion. See, e.g., Siag v. Egypt, supra note 63, ¶¶ 519, 584. The highest and lowest amounts claimed were, respectively, $553 million (in Azurix v. Argentina) and $9 million (in Alpha v. Ukraine). The median amount claimed was $183 million (in RosInvest Co. v. Russia).

87. Two exceptions are Kardassopoulos v. Georgia and RosInvest Co. v. Russia. Notably, the tribunal in RosInvest Co. v. Russia arrived at a valuation equal to the highest amount proposed by Russia. See RosInvest v. Russia, supra note 70, ¶¶ 657, 660, 675-76.

88. This average does not include the interest or costs and fees awarded in each case. The median amount awarded was $76 million (in ADC v. Hungary). The highest and lowest amounts awarded were, respectively, $199 million (in Siemens v. Argentina) and zero (in Biwater v. Tanzania).

89. See Franck, Empirically Evaluating, supra note 35, at 60.

90. The amounts awarded were $76 million (in ADC v. Hungary) and $30 million (in Kardassopoulos v. Georgia). The tribunal in ADC v. Hungary declined to award certain lost development opportunities, seemingly because the claimants did not put forward a quantification of their value. See supra note 4.

91. The amounts claimed were $183 million (in RosInvest v. Russia) and $20 million (in Biwater v. Tanzania).
value. The results are summarized in the chart below.

**Fair Market Value in Recent Investor-State Cases:**

*Amounts Awarded as a Percentage of Amounts Claimed*  

---

**C. The Failure to Explain Calculations**

Arbitrators bolster the legitimacy of their awards by thoroughly explaining their resolution of complex issues of valuation. By contrast, vague and ambiguous decisions invite criticism and undermine the legitimacy of an award. Parties will rarely be satisfied with the excuse of valuation being “inherently uncertain.” Valuation may not be an “exact science,” but it can nonetheless be described with precision.

Well-reasoned decisions benefit the parties, other arbitrators, and third parties. Potential benefits include promotion of settlement between the parties and broader jurisprudential development. From all perspectives, international arbit...
tation is more legitimate when parties and nonparties understand the issues and reasoning underlying a decision. Because awards against states require the use of public funds, tribunals may have an added responsibility to quantify damages transparently. Investors, too, will benefit from detailed, publicly available explanations that “neutralize the repeat player advantage” of states and alert investors to the practices of states with respect to foreign investors. Arbitrators also have self-interest in publishing well-reasoned decisions, which support the arbitral system and an arbitrator’s personal reputation, both of which are necessary for future appointments. In short, “[a] reasoned judgment contributes to ensuring not only that justice is done but that it is perceived to be done.”

Although investor-state decisions are moving toward better explanations of valuation, deficient discussions of specific calculations remain a common exception to the trend. The failure to explain calculations in detail is perhaps justified in rare cases in which investors claim relatively small amounts. In most cases, however, the failure to explain valuation adequately hints at a failure to address the issue methodically, thus exposing an award to greater skepticism.

2009) (“No doubt frivolous, perfunctory or absurd arguments by a tribunal would not amount to ‘reasons’ for purposes of annulment). Because of the many sources that arbitrators may rely on for their decisions, detailed (and of course intellectually honest) reasoning is critical to legitimacy. See Garcia, supra note 26, at 342.

96. H. Perezcano Diaz, Damages in Investor-State Arbitration: Applicable Law and Burden of Proof, in EVALUATION OF DAMAGES IN INTERNATIONAL ARBITRATION 129 (Yves Derains & Richard H. Kreindler eds., 2006). Those explanations are critical because, if a state does not perceive a damages award to be legitimate, it will be less likely to voluntarily comply with the award. See Weidemaier, supra note 56, at 1918-19 (“Whether the reputational costs of noncompliance provide a substantial inducement to pay depends in part on whether parties in a position to impose these costs—perhaps including investors, international financial institutions, and even the borrower’s own citizens—perceive the award and the arbitration process that produced it as legitimate.”).

97. Coe, Jr., supra note 50, at 1358-59, n.108; see also Franck, Do Investment Treaties Have a Bright Future, supra note 51, at 86-88 (discussing the importance of “the equality of arms”).

98. See Weidemaier, supra note 56, at 1946 (“[A]rbitrators may use the award to communicate information to appease ICSID’s many critics. For example, the award may discuss past awards explicitly and in depth . . . . This kind of direct engagement signals that the decision resulted from a deliberative, systematic process, rather than from an ad hoc balancing of the equities in a particular case.”).

99. See SCHREUER, supra note 95, at 996 (citing Lucchetti v. Peru, ICSID Case No. ARB/03/4, Decision on Annulment, ¶ 98 (Aug. 13, 2007)).

100. See infra Part V.2; RIPINSKY & WILLIAMS, supra note 5, at 191 (listing “examples of awards where arbitral tribunals treated valuation matters quite thoroughly”). Fifteen years ago, by contrast, one commentator explained that “[m]ost . . . tribunal awards are parsimonious in the economic detail which is presented. Whatever financial data is offered by the court has been filtered through a jurist’s prism and typically is not amenable to economic analysis. The terminology is either too casual—confusing income with cash flow, for example—or the pieces of the financial puzzle are too few.” Stauffer, supra note17, at 480.

101. See, e.g., Funnekotter v. Zimbabwe, ICSID Case No. ARB/05/6, Award, ¶¶ 128, 135 (Apr. 22, 2009) [hereinafter Funnekotter v. Zimbabwe]; Bogdanov v. Moldova, SCC Case No V/114/2009, Award, ¶¶ 84-86 (Sept. 22, 2005) (the tribunal devoted no more than a couple sentences to its “estimate” of damages, which was in the $100,000 range).

102. Taking an example from international commercial arbitration, the decision in Karaha
At a minimum, tribunals in ICSID arbitration must state the reasons for their decision to protect awards from annulment under the ICSID Convention. 103 Annulment is a limited exception to the principle of finality of ICSID awards. 104 It should not be confused with an appellate procedure in the United States, in part because of the limited, non-substantive scope of annulment. 105 Annulment seeks to balance concerns of correctness and finality, with a thumb on the scale of finality. 106

In principle, the “failure to state reasons” annulment standard is a bulwark for legitimacy. 107 The annulment decision in Mar. Int’l Nominees Establishment v. Guinea, ICSID Case No. ARB/84/4, Decision on Annulment, (Dec. 22, 1989) (MINE) serves as an early example of how this standard deters arbitrators from poorly explained valuations. 108 In that case, the annulment committee upheld a challenge to the tribunal’s failure to state reasons with respect to its calculation

Bodas v. Pertamina has been attacked in part because the final damages figure looks as if it “was pulled out of the air.” See Wells, supra note 15, at 476; Kantor, supra note 17, at 82-87, n.281 (explaining that a reader could not “recreate the calculation” of the tribunal in Karaha Bodas and that “[t]he ambiguity in the panel’s computation of damages has invited criticism by commentators”). A similar criticism has been lodged against awards of lost future profits “by a rough assessment.” See CME v. Czech Republic, supra note 11; see also Cheng, supra note 35, at 497; Wells, supra note 15, at 473 (finding it “frustratingly difficult to determine exactly how the arbitrators calculated [lost] profits”). The copying and pasting of the parties’ arguments is also no substitute for providing detailed analysis, particularly with respect to damages. For example, in Rumeli Award, the tribunal laid out the parties’ arguments in detail but devoted only a few paragraphs to its determination of the actual amount to be awarded. Compare Rumeli Award, supra note 2, ¶¶ 752-784 with id. ¶¶ 805-818. The inadequate explanation in the decision surely contributed to the state’s decision to seek annulment, as discussed below.

103. See ICSID Convention, art. 52(1)(e) (listing as a ground for annulment “that the award has failed to state the reasons on which it is based”). The failure to state reasons has been a common ground for challenging the damages portions of awards. See infra n.116.

104. Schreuer, supra note 95, at 899.

105. As compared to a U.S. appeal, “annulment is only concerned with the legitimacy of the process of the decision: it is not concerned with its substantive correctness.” Id. at 901. This distinction also applies with respect to damages. See, e.g., Azurix Annulment, supra note 116, ¶ 362 (“The Committee recalls that it is not a court of appeal, and that it is not the function of the Committee to pass judgment upon the substance of the Tribunal’s decision with respect to the quantum of damages.”).

106. See Schreuer, supra note 95, at 903 (“In international arbitration the principle of finality is often seen to take precedence over the principle of correctness.”); see also Franck, Legitimacy Crisis, supra note 12, at 1548 (“Ultimately, because legal errors cannot be corrected in ICSID awards, the possibility of inconsistent awards is an accepted reality at ICSID, and the correctness of decisions has been sacrificed for the sake of finality.”). For a criticism of the limited scope of ICSID annulment, see Garcia, supra note 26, at 344-45.

107. As one annulment committee explained, this standard “aims at ensuring the parties’ right to ascertain whether or to what extent a tribunal’s findings are sufficiently based on the law and on a proper evaluation of relevant facts.” Lucchetti v. Peru, ICSID Case No. ARB/03/4, Decision on Annulment, ¶ 98 (Aug. 13, 2007).

of lost profits. The claimant in MINE had proposed two methodologies for calculating lost profits (theories “Y” and “Z”). The tribunal rejected those proposals because their results were too speculative, and instead used a third method that it found more realistic. The annulment committee held that “[h]aving concluded that theories ‘Y’ and ‘Z’ were unusable because of their speculative character, the Tribunal could not, without contradicting itself, adopt a ‘damages theory’ which disregarded the real situation and relied on hypotheses which the Tribunal itself had rejected as a basis for the calculation of damages.” The annulment decision in MINE demonstrates how poorly explained valuations can lead parties (and others) to challenge the legitimacy of an award.

The annulment in MINE is an exception, however, because the “failure to state reasons” standard has not been used in any other case to overturn a decision based on an inadequate explanation of valuation. The standard may have become watered down by its repetitive invocation: the “failure to state reasons” has been an alleged ground for annulment in every publicly available annulment decision. Annulment committees, while sometimes criticizing decisions’ opaque rationales, have generally accepted “implicit” reasoning and given deference to tribunals’ modest explanations. Parties have also infrequently pursued annulment of ICSID awards: out of one hundred twenty-seven awards rendered, parties sought annulment of thirty-two awards and prevailed in eleven of those cases. Moreover, parties typically pursue annulment on a wide variety of grounds, such that the “failure to state reasons” for valuation is rarely a standalone challenge. For those reasons, avoidance of annulment, while important, is an insufficient and sub-optimal incentive for more thoroughly explained valuations in international arbitration. Decisions can survive an annulment proceeding with a bare explanation of damages, which undermines the confidence of par-

109. Id. ¶ 6.107.
110. Id.; See also Mar. Int’l Nominees Establishment (MINE) v. Guinea, ICSID Case No. ARB/84/4, Award (Jan. 6, 1988), 4 ICSID REP. 54, 75-76 (1997).
111. MINE v. Guinea Annulment, supra note 108, ¶ 6.107. This decision has proved controversial with respect to the appropriate application of the “failure to state reasons” standard for annulment. See SCHREUER, supra note 95, at 1012-13 (“The ad hoc Committee’s arguments on this point are not convincing and have been criticized by several commentators.”).
112. As one prominent scholar stated, “[t]he only possible criticism that may be levelled against the Tribunal is that the Award could have explained in more detail why it found the method for the calculation of damages that it adopted more realistic than the theories that it dismissed.” See SCHREUER, supra note 95, at 1013.
113. Id. at 998.
114. Id. at 999-1003 (noting numerous instances of annulment committees “reconstructing reasoning” that was not apparent in the tribunal’s award).
116. States have challenged damages for “failure to state reasons” in a number of cases, including: Duke Energy Int’l Peru Invs. No. 1, Ltd. v. Peru, ICSID Case No. ARB/03/28, Decision of the Ad Hoc Committee, ¶ 258 (Feb. 28, 2011) [hereinafter Duke Energy v. Peru]; Azurix Corp. v. Argentina, ICSID Case No. ARB/01/12, Decision on Application for Annulment (Sept. 1, 2009) [hereinafter Azurix Annulment]; MTD Equity Sdn. Bhd. v. Chile, ICSID Case No. ARB/01/7, Decision on
ties and stifles the legitimacy-building effect of greater transparency.

Nevertheless, parties continue to pursue annulment challenges because of inadequate analyses of damages. A clear, recent example is the annulment proceeding in *Rumeli Telekom*. There, the state challenged the award of the arbitral tribunal principally because its “decision to award damages of $125 million was inexplicable, being based on inconsistent, illogical or nonexistent reasons.”[^117] The state put forward a laundry list of complaints about the tribunal’s determination of damages. One basis for the state’s challenge was “that the DCF approach required an actual calculation, not a ‘shot in the dark.’”[^118] In other words, “when a tribunal adopted a DCF analysis, it was required to provide full reasons for its decision to reject or adopt certain factors . . . .”[^119] The state’s challenge essentially was that the tribunal needed to show its math:

> [I]f that [$125 million] figure was reached as the product of a DCF analysis, it was not possible to see how the figure was reached. No inputs were given by the Tribunal, and the methodology was not described. Rather than being “extremely succinct,” the [state] contended that the reasons were nonexistent and that tribunals are obliged to properly reason their awards to avoid deciding *ex aequo et bono*.[^120] The Annulment Committee denied the state’s challenge even though the “[t]he figure of US$125 million is baldly stated in the Award, without an explanation of the mathematical calculation undertaken by the Tribunal in arriving at it.”[^121]

Similarly, in *Azurix Corp. v. Argentina*, ICSID Case No. ARB/01/12, Award (July 14, 2006), the tribunal stated without much support that only “a fraction” of the claimed value was recoverable, and therefore concluded that the value should be $60 million.[^122] The state sought to annul the award for many reasons, including that the tribunal had come up with a damages value without providing “any formulae or principles in the Award as to how that figure was calculated or otherwise obtained.”[^123] The investor countered that “[i]t is not necessary to prove the exact damage suffered in order to award damages; determining damages is not an exact science, and a certain amount of independent judgment is required.”[^124] The Annulment Committee seemed to agree. It held that

[^117]: *Rumeli Annulment*, supra note 93, ¶ 118.
[^118]: Id. ¶ 124.
[^119]: Id. ¶ 126.
[^120]: Id. ¶ 129.
[^121]: Id. ¶ 178.
[^122]: *Azurix v. Argentina Award*, supra note 2, ¶ 429.
[^123]: *Azurix Annulment*, supra note 116, ¶ 297(j).
[^124]: Id. ¶ 298(h). The investor also argued that tribunals have broad discretion in calculating damages. Id. ¶ 298(g, m).
the tribunal’s limited discussion overcame the “failure to state reasons” threshold: “[a]lthough the Tribunal in this case may not have said so expressly, the Committee considers it clear from the Award that the figure of USD 60 million was an approximation that the Tribunal considered to be fair in all the circumstances.” But the state obviously did not agree.

These annulment petitions are more than thorough advocacy; they show that the states were frustrated with tribunals’ discussions of quantum. Indeed, annulment petitions premised on damages highlight the central role of damages for the legitimacy of international arbitration. If an arbitrator awarding damages to an investor wanted to draft an “annulment-proof” decision, the arbitrator would rightly seek to explain all the building blocks for the award, such as jurisdiction and liability. Yet the arbitrator must also focus on the pinnacle of the award—the amount of compensation due.

III.
CONVERGENCE AROUND COMPOUND INTEREST
AND THE DCF METHOD

Jurisprudential convergence provides parties with more accurate expectations. The developing international investment law on awards of interest shows the benefits of such convergence, as it enables parties to better predict the damages at stake. By contrast, parties will be uncertain in many cases about which methodologies a tribunal will endorse for valuation. That uncertainty, as explained above, can be debilitating to the investor-state arbitration system. While there has been promising convergence around the DCF method in theory, inconsistencies persist with respect to tribunals’ application of the DCF method in practice. This is particularly true in cases involving limited evidence.

A. The Trend in Interest

The trend toward compound interest demonstrates the benefits of convergence. This convergence has significant consequences, because interest awards may involve stakes as high as valuation itself. The basic principle of awarding interest is well settled. As stated in the Draft Articles on State Responsibility, “[i]nterest on any principal sum . . . shall be payable when necessary in order to

125. Id. ¶ 351.

126. The following discussion assumes that investors, in seeking to satisfy the burden of proving their case, will adequately develop and argue for the appropriate methodology. That assumption is reasonable in light of the increasing stakes of investor-state arbitration and the sophistication of most lawyers and party-appointed experts involved.

ensure full reparation. The interest rate and mode of calculation shall be set so as to achieve that result.”

128. This established principle of international law is consistent with principles of finance—the time value of money means that full compensation requires an award of interest to compensate for the loss of the use of money between the time of the alleged harm and the award.129 Despite widespread acceptance of interest as generally appropriate, the amount of interest to award had long been in a state of flux.130 Two fundamental questions for determining interest are (1) whether to apply simple or compound interest; and (2) which rate of interest to apply. This Article focuses on the former question, because it is the clearest example of convergence.131

1. Convergence Around Compound Interest

The monetary difference between simple and compound interest can be substantial for high-stakes claims in investment arbitration, particularly because of the increasingly long periods of time between harmful state action and an award.132 Simple interest is calculated only on the principal owed and is never added to the principal. Compound interest is “interest on interest”—meaning that interest is added to the principal.133 Under compound interest, therefore, the amount of interest in later periods will be higher than the amount of interest in earlier periods. Despite the potentially large impact of the simple versus compound interest distinction, that binary question is left unanswered under most

128. ILC Articles on State Responsibility, supra note 6, art. 38(1); see also Sénéchal & Gotanda, supra note 76, at 508, n.73, 516.

129. See Sénéchal & Gotanda, supra note 76, at 495-96, n.15 (explaining that awards of interest also prevent unjust enrichment and promote efficiency).

130. See, e.g., RIPINSKY & WILLIAMS, supra note 5, at 361, n.1 (“International investment tribunals have routinely awarded interest, although frequently without much consistency or even explanation.”); Gotanda, Compounding Interest, supra note 127, at 261 (“In recent years, perhaps no area of private international law has undergone more significant changes than the awarding of interest.”); Sénéchal & Gotanda, supra note 76, at 493.

131. There has also arguably been some convergence with respect to rates of interest in recent years. In exercising their typically wide discretion in determining the rate of interest, see RIPINSKY & WILLIAMS, supra note 5, at 366-67, tribunals tend to be shifting away from “fair” approximations and toward the “investment alternatives” approach to interest, based on floating market rates (i.e. U.S. Treasury Bills and the London Inter Bank Offer Rate (LIBOR)). See Aaron Xavier Fellmeth, Below-Market Interest in International Claims Against States, 13 J. INT’L ECON. L. 423, 431-37 (2010); Sénéchal & Gotanda, supra note 76, at 493-94, 508, n.72 (listing cases); Gotanda, Compounding Interest, supra note 127, at 278-79, 282.

132. See Sénéchal & Gotanda, supra note 76, at 532-33 (“[C]ompounding will have greater impact for high interest rates and longer periods of time.”); RIPINSKY & WILLIAMS, supra note 5, at 380. As with valuation, determinations of interest often involve millions of dollars, sometimes as large as the principal claim itself. See id. at 492-93, n.2.

133. See Sénéchal & Gotanda, supra note 76, at 504, n.59 (“Compound interest is calculated through the use of the following formula: \( FV = PV (1+i)^n \), where FV is the future value of the total award, including interest, PV is the present value of the award (i.e., not including interest), i is the interest rate per compounding period, and n is the number of compounding periods.”).
Throughout most of the twentieth century, simple interest was the norm in international arbitration. According to a leading mid-century treatise, “there are few rules within the scope of the subject of damages in international law that are better settled than the one that compound interest is not allowable.” That rule eroded in subsequent decades. In 2000, the tribunal in Compañía del Desarrollo de Santa Elena S.A. v. Costa Rica, ICSID Case No. ARB/96/1, Final Award, (Feb. 17, 2000) issued a seminal decision awarding compound interest and noting the shifting jurisprudence from simple to compound interest. Still, as late as 2001, one scholar observed that “[w]hile there is little consensus on approaches to awarding interest generally in international arbitration, the issue of compound interest is especially problematic.” In other words, the arbitration community began converging around compound interest, but it had not yet been firmly established.

Ten years later, little to no uncertainty remains with respect to awarding compound interest in investor-state arbitration. Tribunals in the vast majority of published investor-state cases of the past decade have applied compound interest. Indeed, claimants have used empirical data to support the frequent application of compound interest. In Siag v. Egypt, ICSID Case No. ARB/05/15, Award, (May 11, 2009) (hereinafter Siag & Vecchi), for example, the claimants “submitted that since 2000, no less than 15 out of 16 tribunals have awarded compound interest on damages in investment disputes.” Commentators have described this trend as compound interest “com[ing] to be treated as the default

134. RIPINSKY & WILLIAMS, supra note 5, at 365, n.12 (stating that the ILC Articles of State Responsibility, by “refraining from setting out specific rules on the award of interest,” confirmed that as of 2001 there was an “absence of any consistent practice on the matter”).

135. MARJORIE M. WHITEMAN, DAMAGES IN INTERNATIONAL LAW 1997 (1943); RIPINSKY & WILLIAMS, supra note 5, at 382; but see Natasha Affolder, Awarding Compound Interest in International Arbitration, 12 AM. REV. INT’L ARB. 45, 71-73 (2001) (“The authorities cited by Ms. Whiteman . . . fail to support the existence of a general principle of international law against the awarding of compound interest. At most, it can be said that the question of whether compound inter-
est can be awarded is an unsettled question before international tribunals.”).

136. See Starrett Hous. Corp. v. Iran, supra note 9, at 237 (Holtzmann, J., concurring); F.A. Mann, Compound Interest as an Item of Damage in International Law, in FURTHER STUDIES IN INTERNATIONAL LAW 381 (F.A. Mann ed., 1990) (noting that “the general principles of law recog-
nized by civilized nations do not yield an unequivocal guidance” on the question of compound inter-
est; see also RIPINSKY & WILLIAMS, supra note 5, at 383-84 (surveying criticisms of the simple interest rule).

137. See Compañía del Desarrollo de Santa Elena S.A. v. Costa Rica, ICSID Case No. ARB/96/1, Final Award, ¶ 96-107 (Feb. 17, 2000) (“W]hile simple interest tends to be awarded more frequently than compound, compound interest certainly is not unknown or excluded in interna-
tional law.”).  

138. Affolder, supra note 135, at 45-46 (“This lack of uniformity mean[t] that it may be entirely impossible to predict in advance whether an arbitral tribunal will award compound interest.”).

139. See, e.g., RIPINSKY & WILLIAMS, supra note 5, at 384-87; Sénéchal & Gotanda, supra note 76, at 508-09.

140. Siag v. Egypt, ICSID Case No. ARB/05/15, Award, ¶ 595 (May 11, 2009).
solution.” 141 In sum, international investment law has converged around the principle of compound interest.

2. The Benefits of Convergence

Before the convergence around compound interest, parties had limited ability to predict the amount of interest that would be awarded. The negative consequences of that uncertainty included reduced chances of settlement, unnecessary delay in proceedings, and increased litigation costs. 142 Parties also had reason to doubt the competence of arbitrators with respect to interest determinations. 143 Now, however, parties have little doubt with respect to compounding. The convergence around compound interest leads to a more predictable jurisprudence, largely because tribunals turn to arbitral precedent when determining interest. Parties considering settlement, for example, can reliably expect that adequate compensation will include interest at a compound rate. Over time, parties may even avoid incurring the costs of disputing whether interest should be simple or compound. 144

A broader legitimizing benefit of the trend toward awarding compound interest is that it aligns with economic reality. 145 As one tribunal explained, quoting Gotanda, “almost all financing and investment vehicles involve compound interest . . . . If the claimant could have received compound interest merely by placing its money in a readily available and commonly used investment vehicle,

141. RIPINSKY & WILLIAMS, supra note 5, at 387.

142. See Affolder, supra note 135, at 46 (“Uncertainty as to whether compound interest will be awarded is problematic. The fact that parties are unable to ascertain their liabilities (or the amount they may possibly gain) may reduce chances of settlement. Parties may further delay the arbitral process if they believe the cost of interest which they will eventually pay is below the market rate. This lack of uniformity means that parties in similar situations are treated differently so that considerable resources are spent in litigating interest issues.”); see also David J. Branson & Richard E. Wallace Jr., Awarding Interest in International Commercial Arbitration: Establishing a Uniform Approach, 28 VA. J. INT’L L 919, 921 (1988).

143. See Fellmeth, supra note 131, at 435.

144. Of course, as with all damages issues, the convergence around interest does not mean that tribunals will necessarily take up the issue on their own accord. As illustrated by Enron v. Argentina, a tribunal faces the concern of exceeding its powers if it were to award interest that the claimant has not specifically requested. Enron v. Argentina, ICSID Case No. ARB/01/3, Decision on Claimants’ Request for Rectification and/or Supplementary Decision of the Award, ¶¶ 41-42, 56 (Oct. 3, 2007). That remains true notwithstanding an “extensive arbitration practice” of awarding interest to update compensation in light of the time value of money. Id. ¶ 41. But see Wena Hotels v. Egypt, supra note 116 ( awarding compound interest at a rate of 9% even though the claimant claimed interest “but neither specified a rate nor whether interest should be compounded).

145. See Sénéchal & Gotanda, supra note 76, at 505 (explaining “that a loss of value incurred by a company, active in normal trading operations, implies the loss of use of that value”); see also Affolder, supra note 136, at 90-91. Alignment with economic reality also reinforces the convergence around compound interest, because it provides “strong theoretical support.” RIPINSKY & WILLIAMS, supra note 5, at 387.
it is neither logical nor equitable to award the claimant only simple interest.”146 Simply put, “compound interest is a closer measure to the actual value lost by an investor.”147 The convergence around compound interest is therefore consistent with the principle of full reparation under international law.

B. The Prominence of the DCF Method

The fundamentals of DCF analysis are well established and well known in the investment arbitration community. In short, forecasting and discounting the cash flows generated by that enterprise determine the value of an enterprise. Cash flows in the future are worth less than cash flows today.148 Accordingly, “forecasted cash flows are discounted to obtain present value. The appropriate discount rate is the opportunity cost of capital, that is, the expected rate of return from investing in other assets of equivalent risk. . .”149 The principal grounds of dispute regarding a DCF valuation are the amount of projected cash flows and the appropriate discount rate.

The future cash flows valued in a DCF analysis are akin to “lost profits,” but it is important to distinguish the two concepts. In breach of contract cases, lost profits (lucrum cessans) are the future gains that a party would have earned if not for the actions of the other party, as opposed to actual losses suffered (damnum emergens).150 Awarding damnum emergens is more straightforward and less controversial than awarding lucrum cessans.151 DCF, as a method of determining fair market value, is analytically distinct from lucrum cessans,

146. Wena Hotels v. Egypt, supra note 116, ¶ 129.
147. Siemens v. Argentina, supra note 2, ¶ 399.
148. See RICHARD A. BREALEY ET AL., PRINCIPLES OF CORPORATE FINANCE 16 (8th ed. 2006) (“The first basic principle of finance is that a dollar today is worth more than a dollar tomorrow, because the dollar today can be invested to start earning interest immediately. Financial managers refer to this as the time value of money.”).
149. Ball, supra note 69, at 419 (quoting Report of Stewart C. Myers in Phillips Petroleum Co. v. Iran); see also BREALEY ET AL., supra note 148, at 16 (“To calculate present value, we discount expected payoffs by the rate of return offered by equivalent investment alternatives in the capital market. This rate of return is the discount rate, hurdle rate, or opportunity cost of capital.”). For a more detailed explanation of how the DCF method operates in the context of investor-state arbitration, see William H. Knill et al., Accounting for Uncertainty in Discounted Cash Flow Valuation of Upstream Oil and Gas Investments, 25 J. ENERGY & NAT. RES. L. 268 (2007).
150. Gotanda, Lost Profits, supra note 7, at 65-66. “Where the claimant seeks both damnum emergens and the lucrum cessans, [tribunals] need to be careful to avoid double counting” when applying the DCF method. Id. at 111. Note that an award of interest “compensates for the same general class of injury” as lost profits. See Fellmeth, supra note 131, at 427.
151. See RIPINSKY & WILLIAMS, supra note 5, at 107 (“Tribunals rarely have a problem in awarding damnum emergens because this is a loss that has already occurred and is relatively easy to establish and quantify,” but lucrum cessans claims “are de factor much more difficult to sustain, primarily due to a degree of uncertainty inherent in future profits.”); see also Gotanda, Lost Profits, supra note 7, at 62 (stating that lost profits is “arguably the most complicated issues for a tribunal deciding a transnational contracting dispute”).
which is a component of damages.\textsuperscript{152} Despite their differences, both the DCF method and “lost profits” analysis involve similar evidentiary challenges that require a look into the future.\textsuperscript{153}

1. Other Valuation Methods

Tribunals employ a variety of methodologies to determine fair market value, including the DCF method, comparable transactions, book value, and amount invested. Tribunals in investor-state arbitration have encouraged the application of multiple methodologies to corroborate their results.\textsuperscript{154} The analysis of comparable transactions is widely accepted, because such transactions indicate what a willing buyer has in fact paid a willing seller for an investment similar to the one at issue in the arbitration (that is the “comparable”).\textsuperscript{155} Still, using comparable transactions involves its own share of controversy, as exemplified by debates about the relative similarity of the assets or interests at issue.\textsuperscript{156} Perhaps the greatest difficulty with respect to comparables is their availability—because many investor-state arbitrations involve unique investments, comparable transactions are typically limited or non-existent.\textsuperscript{157}

Two other methodologies, book value and amount invested, can be helpful in some contexts, but come up short as systematic approaches to valuation. By definition, book value is not fair market value.\textsuperscript{158} Nor is the amount invested

\textsuperscript{152} For a more detailed explanation of this distinction and an overview of its implications, see Ripinsky & Williams, supra note 5, at 294-98. See also Söderlund et al., supra note 69, at 36 (“Cash flow is not the same as profit.”)

\textsuperscript{153} The general principle of international law is simply that “compensation shall cover any financially assessable damages including loss of profits insofar as it is established.” ILC Articles of State Responsibility, supra note 6, art. 36(2). Both lost profits and DCF analysis tend to involve complex financial models that generate discomfort for arbitrators. Because of the complexity, John Gotanda has observed that, in international commercial arbitration, the calculation of future profits often results in “different approaches and seemingly arbitrary awards.” Gotanda, Lost Profits, supra note 7, at 88. The following sections show that the same is true with respect to the DCF method in investor-state arbitration.


\textsuperscript{155} KANTOR, supra note 17, at 18 (“The best evidence of fair market value may be the price agreed between a willing buyer and a willing seller, each with knowledge of the relevant facts, in a recent arm’s-length transaction.”).

\textsuperscript{156} KANTOR, supra note 17, at 125-30 (providing a checklist of comparability issues).

\textsuperscript{157} See Kardassopoulos v. Georgia, supra note 83, ¶ 598 (“It is not common in investment treaty arbitrations that a Tribunal has available to it three arm’s-length, contemporaneous transactions (or potential transactions) to assist in valuing an investment, much less three that converge in a narrow range of value . . . .”); see also Paul D. Friedland & Eleanor Wong, Measuring Damages for the Deprivation of Income-Producing Assets: ICSID Case Studies, 6 ICSID REV.–FOR. INV. L.J. 400, 404-05 (1991).

\textsuperscript{158} “Net Book Value” equals the assets minus the liabilities of a company, as recorded in ac-
equal to fair market value, even though it can be appealing as an easy approximation of value for recent investments, since it might show what a willing buyer paid to a willing seller for the interests in question. Yet the amount invested has the potential to overvalue or undervalue an investment. Notwithstanding those limitations, valuations based on book value or the amount invested can corroborate valuations based on comparable transactions and DCF analysis. Only in rare cases, however, will book value or the amount invested be appropriate as the exclusive methodology for determining fair market value. By contrast, the DCF method can serve as an appropriate stand-alone methodology, particularly when comparable transactions are not available.

2. Theoretical Convergence Around the DCF Method

In the business and financial spheres, the DCF method is standard practice. Of the many valuation methods, it is the “most conceptually correct method because it captures the driving principle of valuation: value is the present worth of future benefits.” Financial analysts therefore begin the task of valuation by calculating DCF. Tribunals have followed suit, accepting decades ago the DCF method as a matter of theory and applying it to determine fair market value. Yet the method was described even in 2001 as “relatively new in the histor-
The novelty has worn away by now. Recent investor-state arbitration awards demonstrate theoretical convergence around the DCF method. Tribunals have recognized that the “DCF method is widely endorsed, both by financial institutions and international jurists.” Indeed, “DCF techniques have been universally adopted, including by numerous arbitral tribunals, as an appropriate method for valuing business assets,” because they form “[t]he only method which can accurately track value through time . . .”

The DCF is the most common methodology used in valuation analyses. First, it is widely supported by the professional literature, and its workings are well understood. Indeed, most investors rely on a DCF analysis to determine whether or not to undertake a particular project. Second, the DCF approach is widely accepted by international agencies, such as the World Bank, as a valid method to estimate damages and fair market valuations in international disputes.

Parties’ arguments in arbitration proceedings confirm the broad acceptance of DCF analysis. Some claimants have asserted, for example, that DCF is the most “widely accepted and highly regarded methodology used to calculate the value of cash flows being generated by a business.” Respondent states also increasingly accept that DCF is proper as a matter of theory. Accordingly, ex-

164. Ball, supra note 69, at 419-21 (“It took some decades for the cases to establish that lucrum cessans (lost future profits) is an allowable element of compensation for expropriation. The next step—developing techniques for calculating the value of the lost profits and convincing tribunals of the validity of this method—is a relatively late method.”); see also CAMBELL McLACHLAN ET AL., INTERNATIONAL INVESTMENT ARBITRATION 331 (2007) (“The view of these authors is that the DCF approach is becoming so widely accepted because it is, put simply, the best method for valuing lost profits.”).

165. Enron v. Argentina, supra note 2, ¶ 385, n.118.

166. CMS Gas Transmission Co. v. Argentina, ICSID Case No. ARB/01/8, Award, ¶ 416 (May 12, 2005) [hereinafter CMS v. Argentina].

167. Walter Bau v. Thailand, supra note 83, ¶ 14.12. Another tribunal stated that valuation based on future lost profits “theoretically . . . may even be the preferred method of calculating damages in cases involving the expropriation of or fundamental impairment of going concerns.” Compagnía de Aguas v. Argentina, supra note 2, ¶ 8.3.3.

168. Manuel A. Abdala, Key Damage Compensation Issues in Oil and Gas International Arbitration Cases, 24 AM. U. INT’L L. REV. 539, 548-49 (2009); see also Ball, supra note 69, at 427 (“[T]he DCF method is recognized increasingly as a valid valuation method, even in cases in which tribunals have found the evidence insufficient to support a DCF analysis.”).

169. Rumeli Award, supra note 2, ¶ 722; see also Compañía de Aguas v. Argentina, supra note 2, ¶ 8.3.1 (“Claimants contend that a DCF analysis is recognised as the preferred approach to valuation in modern practice where projected cash flows are reasonably capable of determination and are not speculative.”).

170. Compare Rumeli Award, supra note 2, ¶ 726, and CMS v. Argentina, supra note 166, ¶ 417 (the experts from both sides agreed “that DCF was the proper method in this case for determining losses that extend through a prolonged period of time”), with Rumeli Annulment, supra note 93, ¶¶ 124-26, and CMS v. Argentina, supra note 166, ¶ 398 (“The Respondent also asserts that the DCF method is not appropriate and that it has resulted in gross overvaluation of the shares.”). See also Enron v. Argentina, supra note 2, ¶ 355 (“The Respondent objects to the use of DCF to calculate the value of equity damage as a matter of principle and formulates specific objections to the results obtained by the Claimants.”). For a now dated example of a respondent state’s more theoretical criti-
Simmons: Valuation In Investor-State Arbitration: Toward A More Exact Scie

226   BERKELEY JOURNAL OF INTERNATIONAL LAW   [Vol. 30:1

erts for respondent states have submitted their own DCF analyses in order to
counter the valuations submitted by investors.171 The use of DCF analysis by
state governments and state-owned entities supports the proposition that parties
are moving away from theoretical challenges to the DCF method of valuation.172

3. Arbitrators’ Lingering Reluctance

Despite its growing acceptance, the DCF method continues to face incon-
sistent and hesitant application in investor-state arbitration. In some cases, tribu-
nals avoid DCF analysis for legal reasons, for example, by finding that fair mar-
ket value is not the appropriate standard of compensation in a non-expropriation
case.173 More important is arbitrators’ possible avoidance of the DCF method for
evidentiary or subjective reasons. Tribunals rejected the DCF method for such
reasons in Metaclad Corp. v. Mexico, ICSID Case No. ARB(AF)97/1, Award,
(Aug. 30, 2000), Wena Hotels Ltd. v. Egypt, ICSID Case No. ARB/98/4, Deci-
sion on Annulment Application, (Feb. 5, 2002), Tecnicas Medioambientales
Tecmed, S.A. v. Mexico, ICSID Case No. ARB(AF)/00/2, Award, (May 29,
2003), S. Pac. Props. (Middle East) Ltd. v. Egypt, ICSID Case No. ARB/84/3,
Award, (May 20, 1992), and Amoco Int’l Fin. Corp. v. Iran, Partial Award, 15
analyses in four of the past twelve public decisions in which investors proposed
a fair market valuation principally based on the DCF method.175 To understand
cisms of the method, see S. Pac. Props. (Middle East) Ltd. v. Egypt, ICSID Case No. ARB/84/3,
Award, ¶¶ 186-87 (May 20, 1992) [hereinafter SPP v. Egypt].

171. See, e.g., Sempra v. Argentina, supra note 2, ¶ 407; CME v. Czech Republic, supra note
102, ¶¶ 563-64.

172. See KANTOR, supra note 17, at 135 n.416.

173. This legal question has also reflected tribunals’ inconsistent approach to valuation. See
Khamsi, supra note 27, at 175-78, 183. In LG&E v. Argentina, for example the tribunal rejected
the DCF method because, while appropriate in expropriation and “total loss of investment” cases, DCF
(as a measure of asset value) was not appropriate for the non-expropriation breaches in question. See
LG&E v. Argentina, supra note 47, ¶¶ 35-39; see also PSEG Global Inc. v. Turkey ICSID Case No.
ARB/02/5, Award, ¶ 309 (June 4, 2004) (“The Tribunal accordingly finds that the fair market value
shall not be retained as the measure for compensation in this case and hence it will also not discuss
the many technical aspects raised by the parties in connection with the factors that were taken into
account for assigning a value to the claim and the appropriate method for its calculation.”) [hereinafter
PSEG v. Turkey]. The typical posture of the parties was reversed in LG&E v. Argentina. The
claimants did not propose the DCF method and instead relied on the sale price of their publicly-
traded shares in the investment and comparables. LG&E v. Argentina, supra note 47, ¶¶ 13-15. The
respondent’s expert proposed “the use of DCF as a more appropriate and rigorous method to value
the investments,” as compared to the claimants’ stock-price method. Id. ¶ 23. Yet, interestingly, the
respondent did not conduct a calculation based on the DCF method. Id. ¶ 34.

174. See RIPINSKY & WILLIAMS, supra note 5, at 205-10 (summarizing six cases in which tri-
butals rejected the DCF method). See also KANTOR, supra note 17, at 136 n.421 (listing early inves-
tor-state cases in which tribunals declined to award lost profits).

175. The twelve cases considered are Alpha v. Ukraine, Walter Bau v. Thailand, Siag v. Egypt,
Nat’l Grid v. Argentina, Rumeli Award, Biwater v. Tanzania, BG Grp. v. Argentina, Sempra v. Ar-
this persistent and sometimes puzzling reluctance toward accepting the DCF method, an analysis of the four recent cases in which the DCF method was rejected is necessary.

Siemens v. Argentina. In Siemens A.G. v. Argentina, ICSID Case No. ARB/02/8, Award (Feb. 6, 2007), the claimant presented an unusual argument that compensation should be calculated based in part on book value and in part on “discounting an estimate of profits” that were expected from the investment. Commenting on that proposal, the tribunal stated that “the DCF method is applied to ongoing concerns based on the historical data of their revenues and profits; otherwise, it is considered that the data is too speculative to calculate future profits.” The tribunal accepted the book value approach, but rejected the DCF value of lost profits as “very unlikely to have ever materialized” for five reasons: (1) the “excessive” amount of profit assumed by the claimant needed to be reduced; (2) that reduced amount included a value added tax that needed to be subtracted; (3) a discount rate of thirteen percent needed to be applied; (4) the profits depended on uncertain assumptions about possible contract extension; and (5) the profits would have been subject to a corporate profits tax.

The tribunal’s analysis illustrates both quantified and unquantified explanations of the DCF method. For its first two stated reasons, the tribunal gave exact numerical reductions that should have been applied to the proposed DCF analysis (resulting in profits before taxes of AR$81 million). For its next three stated reasons, the tribunal did not quantify how any of those three reasons could reduce the admitted value of AR$81 million (over US$20 million) to zero. The tribunal’s failure to explain the specific adjustments of the DCF analysis that resulted in a zero or negative value gives rise to doubt about whether the tribunal undertook to perform a complete DCF analysis.

Compañía de Aguas v. Argentina. In Compañía de Aguas del Aconquija S.A. v. Argentina, ICSID Case No. ARB/97/3, Award (Aug. 3, 2007), the claimant used the DCF method to value over twenty-seven years of lost profits. The respondent challenged the use of that method, and the claimant provided no alternative in response. The tribunal held that the claimants “failed to establish
with a sufficient degree of certainty” that the investment in question would have been profitable. The tribunal explained that “the net present value provided by a DCF analysis is not always appropriate and becomes less so as the assumptions and projections become increasingly speculative.” While the tribunal acknowledged that “the absence of a history of demonstrated profitability does not absolutely preclude the use of DCF valuation methodology,” it stated that, “to overcome the hurdle of its absence, a claimant must lead convincing evidence of its ability to produce profits in the particular circumstances it faced.” Under that “convincing evidence” standard, the tribunal found the “claimants’ evidence deficient.” The tribunal therefore found that it was unnecessary to further analyze the claimants’ valuation based on the DCF method.

BG Group v. Argentina. In BG Group, the claimant’s damages expert applied the DCF method to determine the reduction in value of its investment. With little discussion, the tribunal concluded that the DCF analysis led “to a result which is uncertain and speculative.” The tribunal did not address the specifics of the claimant’s proposed DCF methodology, such as the projected future cash flows or the discount rate. Instead, the tribunal relied exclusively on two transactions related to the investment in question. The tribunal’s discussion begs the question, at least for non-parties, of why the proposed DCF valuation was “uncertain and speculative.” Similarly, the award leaves one to wonder why the proposed DCF valuation was not a helpful crosscheck for the implied values of the other transactions.

Siag & Vecchi. In Siag & Vecchi, the claimant submitted three methodologies for calculating the fair market value of an expropriated investment: comparable transactions, residual land valuation, and DCF. Egypt countered that on-

183. Id. ¶ 8.3.5.
184. Id. ¶ 8.3.3.
185. Id. ¶ 8.3.8.
186. Id.; see also id. ¶ 8.3.10 (“A claimant which cannot rely on a record of demonstrated profitability requires to present a thoroughly prepared record of its (or others) successes, based on first hand experience (its own or that of qualified experts) or corporate records which establish on the balance of the probabilities it would have produced profits from the concession in question in the face of the particular risks involved, other than those of Treaty violation. This approach was not taken here.”).
187. Id. ¶ 8.3.11.
188. BG Grp. v. Argentina, supra note 2, ¶¶ 415, 438.
189. Id. ¶ 439.
190. Id. ¶¶ 440-44.
191. Siag v. Egypt, supra note 83, ¶¶ 519, 549-52. Perhaps reading cues from the tribunal, the claimants shifted their focus to the comparables analysis by the time of closing submissions in the proceeding. Id. ¶ 571.
ly an analysis of comparables was appropriate. The tribunal found that the investment did not “lend itself to a robust DCF analysis” and based its valuation exclusively on the comparables analysis. The tribunal expressed unease with the DCF method because of its “numerous ‘moving parts’ . . . whether at the front end in terms of building up the model of revenue and operating costs and capital expenditure, or in terms of the Weighted Average Cost of Capital (WACC) used to discount future cash flows back to a present value.” With respect to the discount rate, the tribunal stated that it was “not necessary to attempt the impossible exercise of determining which figure is ‘right’ to realise that the DCF analysis in such a case is attended by considerable uncertainty.” Accordingly, the tribunal agreed with the wisdom in the established reluctance of tribunals . . . to utilise DCF analyses for ‘young’ businesses lacking a long track record of established trading. In all probability that reluctance ought to be even more pronounced in cases such as the present where the business is still in its relatively early development phase and has no trading history at all.

The decisions discussed above reflect two common, interrelated circumstances that lead arbitrators to reject DCF as a method of determining fair market value. First, arbitrators appear willing to reject the DCF method when there is clear, reliable evidence of comparable transactions. That approach, while reasonable in some cases, overlooks the benefit that the DCF method provides as a crosscheck for valuations based on comparable transactions. For example, in CME Czech Republic B.V. v. Czech Republic, Final Award (UNCITRAL Mar. 14, 2003) the tribunal used an “adjusted DCF calculation”—which the tribunal had explained in great, numerical detail—as a confirmation of the Tribunal’s findings based on a comparable transaction. Second, and more troubling, arbitrators reject the DCF method outright when there is a basis for deeming it uncertain and speculative. The degree of speculation and uncertainty varies in

192. Id. ¶¶ 526-28.
193. Id. ¶ 566.
194. Id. ¶¶ 572-73. The tribunal’s preference for valuation based on comparables seemed to derive from its appreciation of the expert who presented that analysis. See id.
195. Id. ¶ 568 (describing the cost projections of the DCF analysis as “necessarily a sketch or rough estimate”).
196. Id. ¶ 569 (emphasis added).
197. Id. ¶ 570 (emphasis added).
198. CME v. Czech Republic, supra note 102, ¶ 604. The DCF method is a particularly helpful tool for corroborating the value suggested by comparable transactions, because the parties involved in those transactions likely utilized DCF analysis for their purchase and sale decisions. See id. ¶¶ 514-17 (explaining how the potential purchases in a comparable transaction had relied on the “budget numbers” of the seller and adjusted “projections for market growth” and other factors).
199. See Thomas Wildé & Borzu Sabahi, Compensation, Damages, and Valuation, in THE OXFORD HANDBOOK OF INTERNATIONAL INVESTMENT LAW 1075 (Peter Muchlinski et al. eds., 2008) (“Almost every tribunal now repeats the mantra that ‘speculative profits’ or ‘speculative elements’ should be discounted in valuation.”); see also RIPINSKY & WILLIAMS, supra note 5, at 210 (stating that a tribunal’s “skepticism of DCF calculations owing to their speculative character . . . is
every case, but tribunals often fail to explain such variances when dismissing the DCF method. It is therefore hard to reconcile theoretical convergence around the DCF method with arbitrators’ reluctance to apply it.

The confused standard of evidence helps explain this inconsistency toward the DCF method. Tribunals have broad discretion to decide the evidentiary threshold for future cash flows and they “are split on what constitutes sufficient evidence.” Some tribunals employ a “certainty” standard in DCF analysis that seems stricter than the more common “reasonable certainty” standard for proving damages. For example, several investor-state decisions suggest that arbitrators have applied a “certainty” requirement to not only the showing of damages per se, but also to the amount of damages. Yet, in general, “it is well not a sensible basis for rejecting the DCF method as a valuation technique.”

200. See Schreuer, supra note 95, at 1012 (“The speculative character of damages theories in the calculation of lost profits is a matter of degree.”).

201. See, e.g., Ripinsky & Williams, supra note 5, at 201 nn.71-72 (citing seven cases in which the DCF method was applied and seven cases in which it was rejected); Rubins & Kinsella, supra note 18, at 249 et seq.; Cheng, supra note 35, at 497 (noting the inconsistency between the willingness to make a “rough assessment” of lost profits in CME v. Czech Republic and the refusal to award compensation “for the earning capacity of an expropriated license in the absence of ‘proof of concrete contracts missed and of the profits lost from them’” in SPP v. Egypt).


203. See ILC Articles on State Responsibility, supra note 6, at 36(2). Investment treaties and applicable arbitral rules provide little to no guidance on the standard of evidence.

204. Weisburg & Ryan, supra note 202, at 175-77; see also Khamsi, supra note 27 (explaining the divergent approaches to compensation in cases against Argentina, and particularly “the divergent approaches to certainty”).

205. Kantor, supra note 17, at 77 (noting a “somewhat stricter test” in international investment law); see also id. at 80 (“[A] tendency on the part of tribunals to decline to award future-looking damages on the basis of insufficient certainty often shows up in claims by investors against States — claimants face a high burden of proof requirement [as compared to] U.S. breach of contract cases . . . .”); Gotanda, Lost Profits, supra note 7, at 87 (“In general, the claimant must prove lost profits with reasonable certainty. In many countries, though, the certainty rule applies only the fact that the breach resulted in claimant’s loss of future revenues and not to the amount of profits it lost. The UNIDROIT Principles require that lost profits be established with a reasonable degree of certainty.”).

206. See, e.g., LG&E v. Argentina, supra note 47, ¶ 51 (“[L]ost future profits . . . . have only been awarded when ‘an anticipated income stream has attained sufficient attributes to be considered legally protected interests of sufficient certainty to be compensable.’ Or, in the words of the Draft Articles, ‘in so far as it is established’. The question is one of ‘certainty’. ‘Tribunals have been reluctant to provide compensation for claims with inherently speculative elements.’”) (citations omit-
settled that the fact that damages cannot be assessed with certainty is no reason not to award damages when a loss has been incurred."\textsuperscript{207} The standard of evidence is particularly important for entities with little to no track record. Such entities are not "going concerns,"\textsuperscript{208} a categorization that arbitrators frequently employ when rejecting a DCF analysis.\textsuperscript{209} The "going concern" label thus serves as states’ primary weapon for opposing DCF valuations.\textsuperscript{210} Whether a company satisfies the definition of a "going concern" is essentially an evidentiary question of the certainty of cash flows.\textsuperscript{211} Business plans also demonstrate tribunals’ division on evidence of future cash flows. For example, one tribunal has stated that "[i]t is always difficult to assess lost profits. One cannot simply rely on a
business plan." By contrast, another tribunal recognized that business plans "constitute the best evidence before the Tribunal of the expectations of the parties at the time of expropriation for the expected stream of cash flows." Such evidentiary disagreements suggest a deeper cause of inconsistency.

The varying competence and unease of arbitrators with respect to the DCF method is critical. Some arbitrators and commentators have made clear their discomfort with DCF’s requirements of forecasting and detailed financial analysis. More generally, tribunals have insisted that DCF “be used with caution,” or even “great caution.” The ILC Articles on State Responsibility disparage the DCF method as relying on “a wide range of inherently speculative elements, some of which have a significant impact upon the outcome (for example discount rates, currency fluctuations, inflation figures, commodity prices, interest rates and other commercial risks).” Tribunals appear to appreciate an easier, more certain method of calculating damages.

Another possible reason for arbitrators’ rejection of the DCF method is political sensitivity to large awards against states and a perception of the DCF method “as putting too much of a burden on the respondent state.” That sensi-

---

213. ADC v. Hungary, supra note 4, ¶ 507; see also Walter Bau v. Thailand, supra note 83, ¶ 14.21; CME v. Czech Republic, supra note 102, ¶ 59 (separate opinion of Ian Brownlie) (stating that a business plan was “a reliable guide to the business expectations of the investors”).
214. See, e.g., Söderlund et al., supra note 69, at 24 (“The forward-projecting assessment methods . . . strike us as being very uncertain, very speculative, and not agreeable at all to apply.”) (emphasis added); Thomas Wälde, Introductory Note to SVEA Court of Appeals: Czech Republic v. CME Czech Republic B.V., 42 I.L.M. 915, 918 (2003) (“Damages in complex businesses relying on calculations of future cash flows (quite speculative) discounted to present value by applying a specific discount rate (itself very uncertain as the risk factor added to the risk-free discount rate is inevitably highly subjective) can be reasonably and plausibly determined within a very wide range.”); see also supra Part III.A. This discomfort has been highlighted by some in the arbitral community who find the DCF method to be “extremely difficult.” See Weisburg & Ryan, supra note 202, at 174; see also, e.g., Wälde & Sabahi, supra note 199, at 1073 (“The difficulty with [the DCF] method . . . is that while it may look objective and scientific when presented by experts using spreadsheet models, it does not provide objective and predictable outcomes. The DCF method is in essence a speculation about the future dressed up in the appearance of mathematical equations.”); Wells, supra note 15, at 474-75; Cheng, supra note 35, at 497.
216. ADC v. Hungary, supra note 4, ¶ 502 (noting “the Respondent’s admonishment that ‘international tribunals have exercised great caution in using the [DCF] method due to its inherently speculative nature’”).
217. ILC Articles on State Responsibility, supra note 6, art. 36, ¶ 26.
218. See Siag v. Egypt, supra note 83, ¶ 583 (“[T]he Tribunal prefers to apply a simple analysis to what is on its face a fairly simple contractual term.”); Azurix Award, supra note 2, ¶ 425 (“[T]he Claimant] has asserted in addition that the argument in support of using actual investment is compelling as the investment is recent and highly ascertainable. The Tribunal agrees that the actual investment method is a valid one in this instance.”).
219. RIPINSKY & WILLIAMS, supra note 5, at 231. Sergey Ripinsky provided this explanation in a presentation to the Investment Treaty Forum of the British Institute of International and Compara-
Activity might give arbitrators considerable pause when an award based on fair market value would “entail catastrophic repercussions for the livelihood and economic well-being” of a state’s population.\textsuperscript{220} Regardless of the validity of these considerations, if political or equitable factors in fact drive a tribunal’s award, the tribunal should not purport to award damages based solely on an accurate valuation. Nor should the relative uncertainty of the DCF method be a scapegoat for arbitrators who seek to avoid fair market valuation for political or equitable reasons.

IV. RECOMMENDATIONS FOR A MORE EXACT SCIENCE

Arbitrators could enhance the legitimacy of valuation in investor-state arbitration through more consistent willingness to apply the DCF method and more frequent appointments of independent valuation experts. Those two beneficial shifts would face few obstacles. Most investment treaties and arbitral rule systems give tribunals broad discretion to employ the DCF method and enlist tribunal-appointed experts. Parties will likely continue to propose the DCF method and are not likely to oppose a tribunal’s appointment of an expert. Indeed, particularly in high-stakes cases, parties may welcome the involvement of such experts, who would likely contribute to financially sound, well explained, and thus legitimate valuations.

A. Utilize the DCF Method to Address Uncertainty

The general trend in favor of the DCF method is consistent with principles of modern finance. Yet tribunals and commentators have continued to decry the complexity and inexact components of DCF analysis. Except in cases of notably deficient evidence, tribunals should not dismiss a detailed DCF analysis of fair market value with the magic wands of “uncertainty,” “speculation,” and “not a going concern.” This is not to say that “uncertainty” or “speculation” do not exist. Rather, to the extent that “uncertainty” and “speculation” have become legal bars that limit the precision of damages awards, those legal bars should be abandoned.\textsuperscript{221} Arbitrators should employ the DCF method as a tool to help pinpoint...
and reduce sources of uncertainty and speculation. At least three related benefits would result from this usage of the DCF method: closer alignment with accepted business practices, increased transparency, and an unbiased approach to valuation.

1. Alignment with Established Business Practices

Unlike arbitral tribunals, analysts in the modern business world are very unlikely to reject the DCF method simply because of “uncertainty” or “speculation.” Because tribunals are seeking to determine market value, their decisions should be informed by the real-world practices of willing buyers and willing sellers. Such an approach would be consistent with the legitimate expectations of investors. Even where there are questions related to the historical data of a company, “a DCF valuation would likely have formed one of the measures which would have informed a discussion between a willing seller and a willing buyer . . . .” In other words, even when lost profits are “very unlikely to have ever materialized,” those lost profits have a value. Perhaps that value will be minimal, but it will not be zero. If there is any probability of future profits, a willing buyer can determine a price at which it would purchase the rights to those profits.

Financial analysts account for uncertainty in their forecasts by adjusting inputs to their models and evaluating a range of outcomes. Arbitrators would benefit from doing the same. The DCF method can help weed out claims that have no evidentiary basis (as opposed to general “uncertainty”). The DCF method “has the advantage of forcing the parties to articulate the various factors which enter into their calculations and, where some individual items are too

---

222. See, e.g., Walter Bau v. Thailand, supra note 83, ¶ 14.22 (“If value and damages must be computed on the basis of what was legitimately expected at any given time, then the DCF method is the most reasonable one to apply.”).

223. Rumeli Award, supra note 2, ¶ 810 (“But the discussion would certainly not have ended there. It is well known that DCF values are to a greater or lesser extent sensitive to the validity of the data on which they are based, such as the inflation rate, the discount rate, the assumptions underlying the predicted cash flows. Claimants’ expert’s report contains a number of sensitivity analyses which demonstrate that quite small changes in input can materially affect the outcome. . . . The Tribunal is aware that the sensitivity analyses are used as a cross check on the figure adopted by the expert, and not to invalidate the figure. Nevertheless, they demonstrate that the method must be understood as an approximation which is dependent on the validity of the assumptions, and not as a mechanical calculation which will yield a value whose validity is not open to question.”). Another tribunal, while rejecting the claimant’s proposed DCF approach, explained that “implicit in the valuation of” comparable transactions “is an expectation of cash flow to equity.” BG Grp. v. Argentina, supra note 2, ¶ 452.

224. Siemens v. Argentina, supra note 2, ¶ 379.

speculative to properly constitute damages, they may be excluded on an item-by-item basis.” 226 For example, in *ADC v. Hungary*, the tribunal accepted the DCF method for the bulk of the claim, but the tribunal rejected a portion of the claim for future opportunities because “the Claimants had no firm contractual rights to those possible projects.” 227

Like financial analysts, arbitrators will of course be called upon to make reasonable approximations, particularly regarding projected cash flows. These approximations are all the more necessary and acceptable when evidence is limited or the projections extend far into the future. 228 But approximation does not undermine the validity of a tribunal’s application of the DCF method. 229 As one tribunal explained:

> There is no reason to apologise for the fact that [the modified DCF] approach involves approximations; they are inherent and inevitable. Nor can it be criticised as unrealistic or unbusinesslike; it is precisely how business executives must, and do, proceed when they evaluate a going concern. The fact that they use ranges and estimates does not imply abandonment of the discipline of economic analysis; nor, when adopted by arbitrators, does this method imply abandonment of the discipline of assessing the evidence before them. 230

The challenges of forecasting are most acute for investments with limited track records, 231 including enterprises that would not be characterized as “going concerns” for purposes of investor-state arbitration. In such cases, it may be particularly appropriate to corroborate and supplement a DCF analysis with other valuation methods. 232 Yet arbitrators should abandon the practice of dismissing the DCF method simply because an entity being valued is not a “going concern.” Financial analysts employ the method even for enterprises that arbitral tribunals would not deem “going concerns.”

226. CHRISTOPHER DUGAN ET AL., INVESTOR-STATE ARBITRATION 586 (2008); see also MCLACHLAN ET AL., supra note 164, at 331-32. In addition, the discount rate accounts for much of the macroeconomic risk or “uncertainty” that tribunals fear. As explained two decades ago but still under-appreciated, “[t]hrough the risk factor in the discount rate, the DCF method explicitly recognizes the uncertainty which is inherent in valuing an income-producing asset.” Friedland & Wong, supra note 157, at 408 (emphasis added).

227. *ADC v. Hungary*, supra note 4, ¶ 515 (noting that the claimants were “unable to quantify, with any fair degree of precision, the damages that would have resulted from the loss of those alleged opportunities”).

228. See RIPINSKY & WILLIAMS, supra note 5, at 121-22, 170-72.

229. “[A]pproximations are inevitable; the settling of damages is not an exact science.” *Compañía de Aguas v. Argentina*, supra note 2, ¶ 8.3.16.

230. *Himpurna v. PLN*, supra note 4, ¶ 376; see also Sénéchal & Gotanda, supra note 76, at 521.


232. Cf. Wells, supra note 15, at 473-74 (“Projecting the stream of earnings for 30 years requires some heroic assumptions, especially for a project that has not yet been completed and thus has no track record; in some cases, such projections are essential, as uncertain as they might be; but there are advantages in seeking another approach when another is feasible.”).
One reason for this disconnect is terminology. Arbitrators’ usage of “going concern” incorporates the evidentiary question of demonstrable future earning power.\textsuperscript{233} By contrast, in the business community, a “going concern” is “a business enterprise that is expected to continue to operate into the future,”\textsuperscript{234} without requiring evidence of future earnings. Mark Kantor aptly suggests avoiding the phrase “going concern” and points to the example of Google Inc.: the current “going concern” test in international arbitration would not have valued the expected future cash flows of Google in 2004, even though Google’s stock market capitalization was over $50 billion after an initial public offering in 2004.\textsuperscript{235} That valuation depended on “uncertain” and “speculative” future cash flows based in part on conjectural sources of revenue.\textsuperscript{236} John Gotanda similarly recommended in the context of international commercial arbitration that “[t]he rule prohibiting the recovery of lost profits whenever the injured business is not a going concern is inappropriate and should be discarded.”\textsuperscript{237} The same is true in investor-state arbitration. Just as a party in a breach of contract case is entitled to the benefit of its bargain,\textsuperscript{238} investors in arbitrations against states are entitled to full reparation.

In order to “wipe out” the consequences of unlawful state action, arbitrators will often be required to determine the fair market value of an investment. When a tribunal denies an accepted method for valuation simply because of uncertainty, justice is impeded.\textsuperscript{239} For financial analysts, the DCF method is hardly a “fig leaf” or an “illusion of scientific analysis [masking] the reality of subjective approximations.”\textsuperscript{240} The opposite is true: the DCF method removes the fig leaf and the illusion, revealing the approximations inherent in valuation. To realize that benefit, it is paramount that tribunals thoroughly explain their application of the DCF method.

\begin{references}
\item 233. See Weisburg & Ryan, supra note 202, at 172.
\item 234. KANTOR, supra note 17, at 95 (quoting International Glossary of Business Valuation Terms).
\item 235. KANTOR, supra note 17, at 100, 102; see also Weisburg & Ryan, supra note 202, at 175 ("[T]he DCF method can be applied to value both new and established going concerns.").
\item 237. See Gotanda, Lost Profits, supra note 7, at 100; RIPINSKY & WILLIAMS, supra note 5, at 283 ("The approach of rejecting lost profits in respect of enterprises which at the time of breach were not going concerns with a profitable record has been criticized as leaving the injured party less than whole and failing to achieve the goal of full compensation.").
\item 238. See Gotanda, Lost Profits, supra note 7, at 101 ("Denying lost profits simply because the injured business is new would leave the injured claimant less than whole and would fail to achieve the goal of full compensation.").
\item 239. Himpurna v. PLN, supra note 4, ¶ 237 ("In this case as in so many others, it is impossible to establish damages as a matter of scientific certainty. This does not, however, impede the course of justice.").
\item 240. Id. ¶ 373.
\end{references}
2. Increasing Transparency: Getting Outside the Black Box

Advances in computer technology have enabled unprecedented precision and clarity for valuations, including better explanations of the adjustments made in determining fair market value. Tribunals increasingly have provided detailed, line-item summations of damages, often based on Excel spreadsheets provided by the parties.\(^{241}\) DCF analysis has undoubtedly benefited from these technology advances. Before today’s ubiquitous spreadsheets, tribunals might explain their consideration of a party’s inputs to a DCF analysis, but fail to quantify any of their adjustments, resulting in a non-verifiable award.\(^{242}\) Of course, the mere involvement of spreadsheets and the DCF method will not guarantee a transparent decision. As discussed above, the attempted annulment of the black-box award in _Rumeli Telekom_ shows that the inputs to DCF analysis should be explained.\(^ {243}\)

The DCF method is amenable to transparent explanations because it has two readily observable components: projected cash flows and a discount rate. Tribunals may also adjust the period of projected cash flows to be valued. If a tribunal explains those inputs, then both parties and nonparties will be in a good position to understand the tribunal’s valuation. The analysis of the tribunal in _Alpha v. Ukraine_ is illustrative. In that case, the claimant’s expert submitted a spreadsheet setting forth the basis for its calculation of the NPV components of the damages claim.\(^ {244}\) The tribunal accepted several inputs from the claimant’s

---

241. See, e.g., _S.D. Myers, Inc. v. Canada_, supra note 231, ¶ 175 (UNCITRAL Oct. 21, 2002) (“[T]he accounting experts provided to the Tribunal spreadsheets which stated amounts claimed by categories. The Tribunal found the spreadsheets to be a useful tool, which assisted its analysis.”). Still, only the “courageous arbitrator” is likely request such computerized financial models. _Kantor, supra note_ 17, at 301. Indeed, the failure to submit a spreadsheet has been found to indicate a failure of proof. See _BG Grp. v. Argentina_, supra note 2, ¶ 448 (finding no support for a claim of historical losses in part because a spreadsheet “used as a source for . . . calculations [was] not on the record”); Zhinvali v. Georgia, ICSID Case No. ARB/00/1, Award, ¶¶ 35-37 (Jan. 24, 2003) (finding that if an electronic version of the claimant’s financial model was not produced, the tribunal would not take the model into account in determining damages).

242. An early and prominent investor-state decision involving DCF analysis and such rough adjustments is _Phillips Petroleum v. Iran_. See _Ripinsky & Williams, supra note_ 5, at 205 (noting the tribunal’s “lengthy examination of the variables affecting the claimant’s proposed DCF valuation” but the lack of any indication of the “precise quantitative effect” of each of those variables). This has also been true in national courts. See, e.g., _Kantor, supra note_ 17, at 91 n.295 (“Once financial advisors had access to the computational power of such programs as EXCEL to develop future earnings forecasts, courts in the U.S. soon followed suit.”).

243. See _supra_ Part III.C; _Rumeli Award, supra note_ 2, ¶¶ 724-36. Ironically, the tribunal in the _Rumeli Award_ provided a good explanation of the DCF method’s role in valuation and took the claimants’ DCF base case valuation as its starting point. _Id._ ¶¶ 810, 813. The tribunal then explained why it moved from that starting point (US$227 million) to the valuation awarded (US$125 million)—for example, because of the value of shares implied by the subsequent purchase of the entity (US$210 million). _Id._ ¶ 813-14. The tribunal failed, however, to quantify or even mention how the DCF “starting point” was modified because of those reasons. If it had, Kazakhstan would have had little basis for pursuing annulment of that portion of the award for “failure to state reasons.” See _id._

244. _Alpha v. Ukraine, supra note_ 83, ¶ 478.
expert, such as the discount rate,245 and explained the precise portions of the spreadsheet that it used in quantifying damages.246

Computerized financial models do not solve the challenges of forecasting in DCF analysis. Rather, because the DCF method makes clear the effects of adjustments to the projections proposed by parties, arbitrators should use it to evaluate those projections and to exercise their discretion. Arbitrators should also heed the principles of business and finance underlying the method, such as accounting for certain risks by adjusting forecasted cash flows instead of arbitrarily increasing the discount rate.247

3. An Unbiased and Fair Method

Arbitrators’ more frequent and improved application of DCF analysis does not favor investors or states at the expense of the other. The typical posture of a case involves an investor proposing the DCF method and a state opposing it, which suggests that the DCF method is more likely to benefit investors than states.248 On the one hand, this perception is correct because rejection of the DCF method can work to the advantage of respondent states in several ways. First, if investors rely exclusively on the DCF method in support of their valuations, tribunals might have no other evidentiary basis for awarding fair market value. Second, if denial of the DCF method results in denial of full market value, arbitration awards would not adequately deter against states breaching their obligations under international law. Third, because some of the “uncertainty” of the DCF method stems from the allegedly unlawful actions of a respondent state (for example, because of expropriation), states have an incentive to limit investors’ ability to put forward a well-documented valuation based on the DCF method.249

On the other hand, the DCF method is not a tool for inflating claims. It is a

245. Id. ¶¶ 482-83.
246. Id. ¶¶ 489-90.
247. BREALEY ET AL., supra note 148, at 223-24. Some commentators have confused this point. See, e.g., Smutny, supra note 10, at 13 (“Where there is a high degree of uncertainty as to what future revenues and costs would be, the discounted cash flow method simply calls for application of a higher discount rate factor to reflect the greater risk that the predicted level of profits in fact would be achieved.”); Weisburg & Ryan, supra note 202, at 178.
248. Coe & Rubins, supra note 154, at 629 (noting that respondent states “may push for a method based upon ‘book value’ or ‘sunk costs,’ which tend to yield a lower result than DCF”); Stauffer, supra note 17, at 478 (discussing the possibility of a “Cinderella effect” by which investors use the DCF method to overvalue assets for purposes of their claims in arbitration).
249. See Funnekotter v. Zimbabwe, supra note 101, ¶ 124 (“[U]nder general international law as well as under the BIT, investors have a right to indemnities corresponding to the value of their investment, independently of the origin and past success of their investment, as well as of the number and aim of the expropriations done”); see also Gotanda, Lost Profits, supra note 7, at 102 (“[B]ecause the respondent’s wrongful act caused the difficulty in proving damages with certainty, from a policy standpoint, the respondent should not be able to escape liability on the ground that lost profits are inappropriate because they are uncertain.”).
tool for determining and corroborating accurate valuations. Even if parties use the DCF method to compute “vastly different damages amounts,”\(^{250}\) the DCF method is not to blame for the divergent calculations. Parties will usually employ whatever tools are available to put forward a damages amount in their favor. The DCF method has the advantage of clarifying whether and how the parties have inflated or deflated valuations, such that arbitrators can adjust the assumptions and calculations as they deem appropriate.

Several arbitration decisions demonstrate the neutrality of the DCF method. For example, in CMS Gas Transmission Co. v. Argentina, ICSID Case No. ARB/01/8, Award, (May 12, 2005) and Enron Corp. v. Argentina, ICSID Case No. ARB/01/3, Award (May 22, 2007), tribunals employed the DCF method and step-by-step explanations of the relevant inputs to reduce the amount of the valuation proposed by investors.\(^{251}\) Another example of where the parties’ positions on the DCF method defied the norm is Biwater Gauff. The claimant asserted that the DCF approach was “too speculative on the facts” because no profits had been made as of the date of expropriation.\(^{252}\) Tanzania countered that the net present value of future cash flows from the claimant’s investment was negative.\(^{253}\) In other words, Tanzania argued, “[n]o prospective purchaser would have been credulous enough to pay anything for [the claimant’s] investment” as of the valuation date.\(^{254}\) The tribunal agreed with Tanzania and held that the fair market value of the expropriated investment “was nil.”\(^{255}\) The DCF method thus worked in favor of the state.

### B. Enlist An Independent Financial Expert

Given the enormous stakes and complicated valuations in many investor-state cases, one might expect that arbitral tribunals frequently turn to independent experts, or perhaps that parties would recommend that tribunals do so. Yet this practice has not been commonplace. Although tribunals have long had the authority to appoint independent experts under nearly all existing arbitration rules,\(^{256}\) their use of financial experts has been quite rare. For example, the Iran-

---

\(^{250}\) See Weisburg & Ryan, supra note 202, at 174; Ripinsky & Williams, supra note 5, at 201 (discussing the tendency of experts using DCF analysis to arrive at diverging results).

\(^{251}\) See, e.g., CMS v. Argentina, supra note 166, ¶¶ 434-67 (applying “a number of changes to [the] assumptions” of the claimant’s expert’s evaluation of damages); Enron v. Argentina, supra note 2, ¶¶ 405-07 (finding that “a number of variables [in the DCF analysis proposed by the claimants] require adjustment”); see also Ripinsky & Williams, supra note 5, at 334, 337.

\(^{252}\) Biwater v. Tanzania, supra note 85, ¶ 749.

\(^{253}\) Id. ¶ 766.

\(^{254}\) Id. Of course, aside from political motivations, this argument begs the question of why Tanzania seized the assets at issue if they had no economic value.

\(^{255}\) Id. ¶¶ 793-97.

\(^{256}\) See Meg N. Kinneer et al., Investment Disputes Under NAFTA: An Annotated Guide to NAFTA Chapter 11, 1133-1-2 (Supp. No. 1 2008) (noting that, although arbitral tribunals have the authority, no tribunal had appointed its own expert in NAFTA arbitration); Claus von
US Claims Tribunal appointed an independent expert in less than 1% of the claims brought before it. Some commentators have suggested that there is also a trend away from tribunal-appointed experts in international arbitration. Despite a growing body of literature on the mechanics of using tribunal-appointed experts, there has been little commentary regarding whether tribunals should appoint an expert on damages. The circumstances of many recent investor-state arbitrations indicate that tribunals should appoint valuation experts more often.

1. The Results of Recent Appointments

The rare cases in which tribunals have appointed an independent financial expert demonstrate the potential benefits of the practice. Four recent decisions are exemplary: CMS v. Argentina, Sempra Energy v. Argentina, Enron v. Argentina, and Nat’l Grid P.L.C. v. Argentina, Award (UNCITRAL Nov. 3, 2008).

CMS v. Argentina. In CMS, the tribunal enlisted, and “was ably assisted” by, its own experts. Perhaps because of this expert assistance, the tribunal explained its decisions regarding numerous approaches to the DCF analysis well,
including the proper capital structure and whether to use “the indirect equity value” or the “direct equity value.”  \(^\text{262}\) More impressively, the tribunal “built its own model” for DCF analysis with the help of its experts.  \(^\text{263}\) With that model in hand, the tribunal was able to systematically modify the assumptions of the claimant’s expert.  \(^\text{264}\)

*Sempra Energy v. Argentina.* The tribunal in *Sempra Energy* also presented a thorough, number-intensive explanation of its valuation.  \(^\text{265}\) Although the decision does not reveal the extent to which the tribunal relied on its appointed financial expert, the careful reasoning of the decision suggests that the expert played an important role.  \(^\text{266}\)

*Enron v. Argentina.* In *Enron*, the tribunal-appointed expert helped the tribunal address the parties’ competing positions. The tribunal accepted the expert’s recommendations and explanations regarding the tariff base,  \(^\text{267}\) the discount rate,  \(^\text{268}\) and the basis for attributing earnings.  \(^\text{269}\) It is not surprising that the tribunal found the approach of its expert to be “more balanced and realistic” than the approaches of the parties’ experts, given that the tribunal’s expert had a nonpartisan role.  \(^\text{270}\) One commentator used the tribunal’s decision as an example of how the DCF method is properly applied in investor-state arbitration.  \(^\text{271}\)

*National Grid v. Argentina.* Most recently, in *National Grid*, the tribunal appointed an independent expert pursuant to criteria provided by the parties.  \(^\text{272}\) Appointing an expert was particularly useful in this case because the tribunal did not have the benefit of a competing valuation from the state.  \(^\text{273}\) Roughly four months after his appointment, the independent expert submitted his final report to the parties and the tribunal, which included the expert’s identification of “manifest errors.”  \(^\text{274}\) In its decision, the tribunal expressed its gratitude to all of the experts “for their contribution to [the tribunal’s] understanding of this matter.”  \(^\text{275}\) The tribunal followed several recommendations by its appointed expert,

\(^{262}\) *Id.* ¶¶ 424-33.

\(^{263}\) *Id.* ¶ 435.

\(^{264}\) *Id.* ¶¶ 439-63.

\(^{265}\) See *Sempra v. Argentina*, supra note 2, ¶¶ 407-82.

\(^{266}\) The decision notes that the expert produced two reports, and that the tribunal gave “due consideration” to the parties’ comments on those reports. *Id.* ¶ 399.

\(^{267}\) *Enron v. Argentina*, supra note 2, ¶¶ 408-10.

\(^{268}\) *Id.* ¶¶ 411-12.

\(^{269}\) *Id.* ¶¶ 418-19.

\(^{270}\) *Id.* ¶ 435.

\(^{271}\) *RIPINSKY & WILLIAMS*, supra note 5, at 203-04.

\(^{272}\) *Nat’l Grid v. Argentina*, supra note 83, ¶ 46.

\(^{273}\) *Id.* ¶ 267 (“The Respondent did not present its own model or methodology attempting to evaluate [damages], submitting instead an expert report purporting to show at least four serious conceptual errors and four methodological errors in the analysis presented by Claimant’s expert.”).

\(^{274}\) *Id.* ¶¶ 47-49.

\(^{275}\) *Id.* ¶ 271.
including (i) recourse to comparable transactions (as a check on DCF)\textsuperscript{276} and (ii) an increase in the discount rate proposed by the claimant’s expert.\textsuperscript{277}

In contrast to the positive examples in those four cases, there has been one clear statement that demonstrates the detriment to legitimacy that might result from not appointing an independent expert. An arbitrator wrote separately in \textit{Siemens v. Argentina} about the tribunal’s refusal to appoint an expert:

It should be noted that the present case comprises complex valuation and financial issues, which were amply argued and discussed by the parties and their respective experts, with very complicated opinions and data. In light of the above, a report from an independent expert is necessary in order to calculate and fully support the amount of damages to be awarded, for all of which I find reasonable the request of its appointment and unjustified its refusal, as such a request never seemed impertinent or untimely to me, but rather reasonable, which acceptance would not have implied any inconveniences.\textsuperscript{278}

The tribunal’s decision not to appoint an independent expert in the face of that arbitrator’s clear desire and need for such assistance does not inspire confidence. Even if only a subset of arbitrators on a tribunal would benefit from a tribunal-appointed expert, such an appointment could bolster the legitimacy of the tribunal’s awarded valuation.

2. The Benefits of Tribunal-Appointed Experts

As the cases above show, tribunal-appointed experts can serve as guides to arbitrators, both in understanding and adjusting financial models and in explaining how those adjustments result in the ultimate award.\textsuperscript{279} Independent expert guidance thus helps alleviate concerns that arbitrators lack financial expertise. Over time, the involvement of independent financial experts might also serve an educational function for individual arbitrators and more generally for the proper application of the DCF method.\textsuperscript{280} At the very least, financial experts lend the technical ability to understand and analyze complex valuation formulas and programs.\textsuperscript{281}

\textsuperscript{276} Id. ¶ 285.

\textsuperscript{277} Id. ¶ 289.

\textsuperscript{278} \textit{Siemens v. Argentina}, supra, note 2, ¶¶ 4-5 (separate opinion of Domingo Bello Janeiro).

\textsuperscript{279} The cases therefore confirm the general observation of Ripinsky and Williams that such experts “can assist the tribunal in evaluating the reports of the parties’ experts, in understanding complicated financial models and, ultimately, to be able to render a well-reasoned decision.” RIPINSKY & WILLIAMS, supra note 5, at 176.

\textsuperscript{280} See id. at 177-78 (explaining that the reports of independent experts may “be useful in developing the law relating to the award of damages where a tribunal endorses an expert’s findings and conclusions on the assessment of compensation,” largely because “judicial endorsement of valuation techniques used in circumstances that arise frequently can provide useful guidance to parties and their experts in valuing similar claims”).

\textsuperscript{281} Such benefits of expert assistance were recognized two decades ago, in connection with the rising importance of computerized expert evidence. Arthur L. Marriott, \textit{Evidence in International Arbitration}, 5 ARB. INT’L 280, 284 (1989).
Thanks to rapidly advancing technology, today’s financial models are often far beyond the traditional training of arbitrators. These complexities are perhaps most pressing in the context of DCF analysis. When valuing a sophisticated enterprise, the DCF method requires technical skills akin to other “scientific” areas. Billions of dollars can hinge on the many interlinked formulas typically embedded in an expert’s computerized financial model. In most investor-state cases, the models “are a riot of numbers, value, and shorthand identifiers for computer calculations, where each printed spreadsheet page (itself highly complex) represents many hidden underlying, inter-linked equations.” For example, a thorough valuation might require multiple simulations to evaluate a broad distribution of expected future cash flow outcomes.

In addition to serving as guides, tribunal-appointed experts provide a useful independent voice in the typical investor-state case in which the parties submit detailed but widely divergent expert reports on damages. The “battle of experts” has become the norm, and the battle can be hardest to judge when the experts submit drastically different valuations. The parties’ experts often take instructions from counsel and therefore are perceived to have questionable independ-

282. See supra Part III.A; see also KANTOR, supra note 17, at 302 (noting that “arbitrators are often unfamiliar with the intricacies of manipulating EXCEL spreadsheets or the detailed building blocks of an Income-Based [e.g., DCF] forecast” and that “[a]rbitrators cannot intuit these relationships”).

283. Of course, tribunal-appointed experts can aid tribunals in any method of determining fair market value. For example, if the parties argue only on the basis of book value, expertise in “the application of ‘complex accounting principles to determine the quantum of damages to be awarded’ . . . ensures that the arbitral tribunal has a proper understanding of the facts and their relation to the applicable law, and increases the prospects that decisions regarding liability and damages will be fully informed, accurate, and, most of all, just.” DAVID D. CARON ET AL., THE UNCITRAL ARBITRATION RULES 665-66 (2006).

284. There are indications of more willingness to use tribunal-appointed experts in “scientific” fields than in finance. For example, the US Model BIT explicitly considers tribunal-appointed experts on “any factual issue concerning environmental, health, safety, or other scientific matters,” but does not mention the possibility of tribunal-appointed experts on damages issues. See U.S. MODEL BIT, art. 32; but see id. art. 20(3)(c)(i), 20(5) (encouraging parties in disputes related to financial services to “take appropriate steps to ensure that the tribunal has expertise or experience in financial services law or practice”).

285. See KANTOR, supra note 17, at 302. Financial experts would likely identify and address mistakes in such models that arbitrators would not catch. See id.

286. Id. at 133; see also, e.g., Duke Energy v. Peru, supra note 116, ¶¶ 463-64 (noting that the claimant’s expert had constructed a financial model that covered over 6,000 assets and had twenty-three steps accounting for the annual cash flow impact of the government measures in question).

287. BREALEY ET AL., supra note 148, at 253-57; see also Knell et al., supra note 149, at 24-25. For such complex simulation models, even business managers “may delegate the task of constructing the model to management scientists or consultants.” BREALEY ET AL., supra note 148, at 257. There are nonetheless practical limits to the complexity of a model prepared by an expert, including the ability of the decision maker to understand the model. See id.

ence.\textsuperscript{289} Even if the parties’ experts are independent, in many cases they will be subject to cross-examinations that seek simply to undermine their credibility and obfuscate the relevant issues.\textsuperscript{290} Moreover, as commentators have noted, there is an “unwillingness of experts on opposing sides of a dispute to reach any meaningful consensus in many cases . . . .”\textsuperscript{291} The problem has also been described as “two ships passing in the night,” based on the perception of expert evidence as “highly-paid advocacy from a credentialed witness.”\textsuperscript{292}

Although some commentators are skeptical that tribunal-appointed experts will assist in solving the battle of the experts,\textsuperscript{293} other methods of managing opposing experts come up short compared to independent experts. Expert conferencing, for example, has become one popular practice in the battle of the experts. Also referred to as “hot-tubbing,” such conferencing involves experts from opposing sides sitting together for questions from the tribunal and, in some instances, the parties. These “hot tubs” can yield benefits such as reducing tensions, highlighting differences in opinion, and exploring those differences and the credibility of the experts.\textsuperscript{294} However, several factors limit the effectiveness of hot-tubbing, including the tendency of experts to focus solely on avoiding hurting their party’s case, rather than genuinely seeking agreement or guiding


\textsuperscript{290} See KANTOR, \textit{supra} note 17, at 134 (“[A]rbitrators are regularly left with two widely different valuations, each having suffered heavy damage to its credibility from litigation artillery duels.”).

\textsuperscript{291} Jacobs & Paulson, \textit{supra} note 79, at 399. There is a risk that a tribunal-appointed expert will be just another opinion that does not help the tribunal decide between the parties. See, e.g., \textit{Act II: Pre-Hearing Advocacy}, 21 ARB. INT’L 561, 569 (2005) (comment by Rob Smit). That risk will likely be minor (or at least a risk worth taking) with respect to the complex valuations at issue in investor-state arbitration.

\textsuperscript{292} See Frances P. Kao et al., \textit{Into the Hot Tub . . . A Practical Guide to Alternative Witness Procedures in International Arbitration}, 44 INT’L L. 1035, 1035-36 (2010) (describing in particular the issues of expert evidence in United State litigation, and stating that international arbitration “frequently looks like the usual morass found in the U.S. courts, particularly where American lawyers are involved. . . . This posture may result in polarized, intractable positions between the parties’ experts. Most importantly, lawyer control over the examination process means that the important questions—typically the thorniest ones—can go unanswered or are glossed over, either because they are intentionally sidestepped or because counsel does not have sufficient facility with the particular topics at issue to elicit clear, relevant testimony.”).

\textsuperscript{293} See CARON ET AL., \textit{supra} note 283, at 668 (arguing against the use of tribunal-appointed experts because of avoiding a “battle of experts” in which the parties feel “compelled to seek expert advice for purposes of evaluating and possibly challenging the conclusions of the tribunal-appointed expert); Allison & Holtzmann, \textit{supra} note 257, at 281 (“[A]n economic analysis of such matters as the future level of prices for a given commodity or the prospects for a particular trade or business may, perhaps, be as readily determined by the application of common sense and careful analysis by the arbitrators of the evidence before them as by an expert opinion will, predictably, be subjected to attached leveled against it by other experts marshaled by the parties.”).

\textsuperscript{294} See Kao et al., \textit{supra} note 292, at 1042-43.
the arbitrators.295 Having an independent expert join the hot tub could help overcome these limitations by moderating the competing experts’ positions and helping identify common ground. One might therefore think of independent experts as specialized ad hoc scholars on expert valuation battles: they enter the fray, offer analysis, and provide clarity.296

3. The Costs of Tribunal-Appointed Experts

Two principal objections to a tribunal’s appointment of an expert are cost and overreliance on (or delegation to) the expert. The parties are responsible for paying the direct costs of a tribunal-appointed expert,297 and will likely incur even greater costs in reviewing and potentially challenging the conclusions of such an expert.298 The parties therefore may believe that a tribunal-appointed expert will only delay and increase the costs of the arbitration.299 That would likely be true when the parties’ experts have submitted calculations that are within a close range of, or that rely on, similar methodologies and present detailed, comparable analysis.300

Arbitrators must therefore consider the amount at stake and complexity of the valuation dispute in deciding whether to appoint an expert.301 There will be

295. The hot tub might place undue emphasis on the general appeal of an expert, rather than the solidity of the positions he or she is endorsing. See id. at 1044 (listing the important traits of an expert for the hot-tubbing process, including teaching ability, likeability, and skills with “cross-examining” the opposing expert); KANTOR, supra note 17, at 300-01.


297. See, e.g., INTERNATIONAL BAR ASSOCIATION RULES ON THE TAKING OF EVIDENCE IN INTERNATIONAL ARBITRATION art. 6(8) (“The fees and expenses of a Tribunal-Appointed Expert, to be funded in a manner determined by the Arbitral Tribunal, shall form part of the costs of the arbitration.”) [hereinafter IBA RULES]. The financial burden of a tribunal-appointed expert will be typically be relatively less imposing for states than for investors. See Wälde, supra note 12, at 23. Yet, because tribunal-appointed experts are most appropriate in high-stakes cases, it is reasonable to expect that in those cases investors will have adequate funds to cover the costs of a tribunal-appointed expert.

298. Allison & Holtzmann, supra note 257, at 281 (noting that an independent expert’s “opinion will, predictably, be subjected to attached leveled against it by other experts marshaled by the parties”).

299. See Hunter, supra note 289; see also Hartwell et al., supra note 17, at 20 (comments of B. Hanotiau).

300. See, e.g., Walter Bau v. Thailand, supra note 83, ¶¶ 14.5, 14.23-14.24. An ideal outcome is that both sides’ experts submit damages calculations that “reflect a high degree of professionalism, clarity, integrity and independence,” such that those calculations can be weighed without requiring supplemental financial analysis. See ADC v. Hungary, supra note 4, ¶ 516.

301. See Michael McIlwrath & John Savage, The Conduct of Arbitration, in INTERNATIONAL ARBITRATION AND MEDIATION: A PRACTICAL GUIDE ¶ 5-222 (2010) (“Whether a tribunal will choose to appoint an expert will depend on a variety of factors: the technical complexity of the dispute, the existence of relevant expertise within the tribunal, the assistance provided by the parties’ experts [if any are appointed], and the amount at stake. A tribunal that counts lawyers among its members will be unlikely to appoint an expert to assist it on issues of law.”); Allison & Holtzmann, supra note 257, at 271 (“[T]he Iran-US Claims Tribunal generally has focused upon two considera-
relatively small or simple cases in which the costs of a tribunal-appointed expert will not be justified. The trend in investor-state arbitration is not, however, toward small and simple disputes. Rather, the high stakes of recent investor-state cases involve complex valuations and require several years to resolve. A pressing reality of this trend is that parties will continue to submit calculations with such drastic differences that arbitrators will face great difficulty in reconciling them. Arbitrators will face similar difficulty in resolving disputed valuations when investors submit detailed expert reports, but respondent states do not provide such detail.

One perhaps counterintuitive method for limiting the monetary costs associated with a tribunal-appointed expert is to raise the issue early in the proceedings. As one commentator observed, "as soon as the arbitral tribunal realizes the importance or necessity of appointing an expert, it should initiate the appropriate steps and inform the parties as soon as possible in order to receive their perspectives." In an ICSID arbitration, for example, tribunals will often learn from the request for arbitration that a complex valuation is likely to ensue. Accordingly, a tribunal could pose to the parties the question of whether to appoint an expert at the first procedural session. While unlikely, the parties may even agree that a single, tribunal-appointed expert would be preferable to the high costs of engaging competing party-appointed experts.

The second potential disadvantage of a tribunal-appointed expert is that arbitrators could defer too much of their authority to such experts, or that a perception of such deference would undermine parties’ confidence in an award. Ar-
bitral rule systems and general principles of international arbitration provide an important defense against such deference. It is the tribunal’s responsibility
to determine the final damages amount. As stated in the IBA Rules, “[a]ny Expert
Report made by a Tribunal-Appointed Expert and its conclusions shall be as-
signed by the Arbitral Tribunal with due regard to all circumstances of the
case.”

Tribunal-appointed experts also may not investigate and develop the
underlying facts, which would aid a party in proving its case.

Despite such established principles, parties might still view arbitrators as
dependent on their experts’ analyses. Ironically, one reason for doubting a tri-
bunal’s independence in assessing damages might be the same limited financial
competency that makes a tribunal-appointed expert advisable. There is an
important distinction, however, between useful reliance on an expert’s guidance
and complete deference. For example, if an independent expert submits a report
that is “whole, meticulous and comprehensive,” it is reasonable for a tribunal
to rely on that report. In each of the positive examples discussed in this Article,
the tribunals followed the guidance of their experts, but also undertook their
own analysis.

Perceptions of over-reliance on an independent expert are more likely to
arise from parties with common law backgrounds because of their relatively lim-
ited familiarity with independent experts. In contrast to adversarial common

308. IBA RULES art. 6(7); see also Starrett Hous. Corp. v. Iran, supra note 9, at 197.
309. See von Wobeser, supra note 256, at 805 (“[I]t does not serve to liberate either of the par-
ties from the burden of proving its case.”). Such an expansion of a tribunal-appointed expert’s role
will prove controversial, as illustrated by Behring International, Inc. v. Islamic Republic Iranian Air
Force. In that case, one arbitrator dissented to the decision in part because he perceived that the tri-
bunal’s expert was “in reality aiding one Party to engage in what amounts to a ‘fishing expedition.’”
Behring Int’l, Inc. v. Islamic Republic Iranian Air Force, 15 Iran-U.S. Cl. Trib. Rep. 89, 93–95 (J.
Richard M. Mosk dissenting).
310. See, e.g., Smutny, supra note 10, at 23 (“[M]any parties would object to a tribunal hiring a
‘neutral’ expert to perform the necessary calculations, as it may be seen as too great a delegation of
the arbitrators’ decision-making authority . . . .”); Hartwell et al., supra note 17, at 18 (comments of
Serge Lazareff) (“It is very difficult for a tribunal not to follow the expert it has itself appointed. It
then must find technical reasons to go against its own expert, and I think it becomes sort of vi-
cious.”); id. at 9 (comments of G. Hartwell).
311. Starrett Hous. Corp. v. Iran, supra note 9, ¶¶ 265-73.
312. See supra Part IV.B.1. In National Grid, for example, the tribunal selected a discount rate
that was closer to the Claimant’s proposed rate than to the high end of the tribunal-appointed ex-
313. See Voser & Mueller, supra note 259, at 73-74, 80; JANE JENKINS & JAMES STEBBINGS,
INTERNATIONAL CONSTRUCTION ARBITRATION LAW 204 (2006) (“While certain parties [particularly
those from civil law jurisdictions, who are accustomed to inquisitorial-style proceedings] may be
comfortable with this arrangement, most parties from common law jurisdictions are not.”); Ruth
Fenton, A Civil Matter for a Common Expert: How Should Parties and Tribunals Use Experts in
International Commercial Arbitration, 6 PEPP. DISP. RESOL. L.J. 279, 288-91 (2006); YVES
DERAINS & ERIC A. SCHWARTZ, GUIDE TO THE ICC RULES OF ARBITRATION 278 (2005) (“[T]he
practice of appointing such experts in ICC arbitration is still much more prevalent among civil law
lawyers than their common law counterparts, who are more accustomed to weighing expert evidence
presented by each of the parties.”); Garcia, supra note 26, at 362-63; Allison & Holtzmann, supra
law practices, civil law systems might not even grant arbitrators the task of “decid[ing] between the opinions of two conflicting technical experts.” Thus it is thus all the more critical that parties from common law traditions have full opportunities to question a tribunal-appointed expert, under the tribunal’s procedural oversight.

4. The Procedures for Tribunal-Appointed Experts

Tribunal-appointed experts must be selected and managed with care. Tribunals should ensure that they appoint an expert who is well suited for the task and unlikely to be challenged. While the parties are unlikely to agree on who should serve as an independent expert, tribunals might turn to the party-appointed experts in seeking consensus. Four essential qualifications for a tribunal-appointed expert are: (1) requisite level of expertise in the relevant field; (2) independence and impartiality; (3) availability; and (4) an ability to perform the necessary function within the financial constraints imposed by the arbitral tribunal. Consistent with these suggested requirements, the recently revised IBA Rules on the Taking of Evidence in International Arbitration require that a tribunal-appointed expert submit to the tribunal and the parties not only a statement of independence, but also “a description of his or her qualifications.” Once appointed, an expert may be challenged under the IBA Rules only for “reasons of which the Party becomes aware after the appointment has been made.”

Creating a single list of qualified experts on valuation would be unnecessary and potentially counterproductive. Rather than artificially limiting available
experts, tribunals should utilize their own experience with experts and seek the recommendations and approval of the parties. Over time, the arbitration community will likely establish an unofficial set of tribunal-appointed “repeat player” experts. There is already a pool of qualified experts who have offered opinions on behalf of both investors and states in a variety of cases. For example, published investor-state decisions from the past five years reveal that Brent Kaczmarek of Navigant Consulting has worked with states in at least two cases (Walter Bau v. Thailand and Rumeli Telekom) and with investors in at least three (Duke Energy v. Ecuador, Duke Energy v. Peru, and Pey Casado v. Chile). Such experts have begun to represent a “cottage industry” in international arbitration, which reinforces the importance of an expert’s reputation.

Just as an esprit de corps has developed among arbitrators, the same is beginning to develop among financial experts in international investment arbitration. After selecting an appropriate expert, tribunals must “precisely define the scope of the functions entrusted . . . .” A number of commentators have provided guidance on how tribunals should handle terms of reference and the management of tribunal-appointed experts. For independent financial experts, the terms of reference should at a minimum assign the task of analyzing the valuations proposed by the party-appointed experts. In some cases, arbitrators might be justified in requesting that an independent expert prepare a separate financial model.

Whatever the scope of an independent expert’s involvement, tribunals should ensure that there are adequate procedural restraints. The 1999 version of the IBA Rules outlined the basic procedures for tribunal-appointed experts, including requesting relevant information from the parties, submitting a report in writing to the tribunal, and answering questions from the tribunal and the parties. The revised 2010 version of the IBA Rules keeps these provisions intact and makes several noteworthy additions. For example, the revised IBA Rules set forth a more detailed rubric for written reports by a tribunal-appointed expert, including “a statement of the facts on which he or she is basing his or her expert opinions and conclusions,” “a description of the methods, evidence and infor-

321. For a brief statement in favor of such a list, see Hartwell et al., supra note 17, at 9 (comments of G. Hartwell) (“[M]ore work [should] be done—and this could be done by institutions of various kinds—to inculcate the necessary judicial or quasi-judicial skills in those experts who wish to serve tribunals. Then, perhaps gradually, over the years, the barriers between lawyers and experts might disappear.”).
322. See Jacobs & Paulson, supra note 79, at 383 n. 128.
323. See Commission, supra note 22, at 135 (noting arbitrators’ similar “backgrounds, qualifications, experiences in international law and their regular interactions, both professionally and otherwise”).
324. See von Wobeser, supra note 256, at 807.
325. See Jacobs & Paulson, supra note 79, at 399; von Wobeser, supra note 256, at 801; Schneider, supra note 259, at 448-64; Allison & Holtzmann, supra note 257, at 273-74.
326. IBA RULES art. 6(1)-(6).
information used in arriving at the conclusions,” and the submission of “[d]ocuments on which the Tribunal-Appointed Expert relies that have not already been submitted.” 327 These procedural steps give parties an opportunity to test an independent expert—a process that contributes to the legitimacy of the expert’s analysis.

CONCLUSION

The jurisprudence of investor-state arbitration is evolving quickly. In 1992, Professor Amerasinghe observed that “the assessment of full compensation is at the present time filled with variables and is certainly not a very scientific process.” 328 Much has changed. As the stakes of investor-state arbitration have risen, so too has the “scientific” precision of valuation. Yet several roadblocks stand in the way of further progress toward a more exact science. Poorly explained valuations are at the root of many challenges, including persistent perceptions that arbitrators lack the competence required for valuation and continue to “split the baby.” Arbitrators can also retreat to the legal safe havens of uncertainty and speculation to avoid DCF analysis, even in circumstances where financial analysts would readily employ the DCF method to evaluate and explain uncertainty. To sidestep those roadblocks, tribunals should appoint an independent financial expert. The benefits of such experts outweigh their costs in these high-stakes cases. The experts would help arbitrators make valuation a more exact science, and that is critical for the legitimacy of investor-state arbitration.

327. Id. art. 6(4).
2012

Book Review, Twilight of Impunity by Judith Armatta

Maya Karwande

Recommended Citation
Available at: http://scholarship.law.berkeley.edu/bjil/vol30/iss1/6

This Article is brought to you for free and open access by the Law Journals and Related Materials at Berkeley Law Scholarship Repository. It has been accepted for inclusion in Berkeley Journal of International Law by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.
I. INTRODUCTION

Over the past two decades, international criminal law has expanded its reach to include the prosecution of heads of state. The trial of Slobodan Milošević, the former president of Serbia, marked the beginning of a new era of accountability. After retaining power through years of war in the Balkans, on June 28, 2001, Serbian officials finally handed Milošević over to The Hague where he faced an indictment by the International Criminal Tribunal for the former Yugoslavia (ICTY) for war crimes, crimes against humanity, and genocide. His trial started a year later. Four years later, before a final judgment was rendered, he died. While the Milošević trial was a crucial step forward for international justice, it has been criticized for its cost and length, the problematic nature of the accused’s self-representation, and the ultimately inconclusive ending.

In her book, *Twilight of Impunity: The War Crimes Trial of Slobodan Milošević*, Judith Armatta documents and unpacks many of the common criticisms surrounding this historic trial to extract important lessons for improving international war crimes prosecutions. As a court monitor for the International Coalition for Justice, Armatta was in the courtroom every day. This experience enabled her to provide illuminating detail about the witnesses, the evidence presented, and the court record. Armatta uses her wealth of knowledge to illustrate how Milošević attempted to undermine the successful completion of his trial by delaying the proceedings, purposefully manipulating his health, and using the trial to promote his political agenda.

II. SUMMARY

*Twilight of Impunity* generally follows the chronology of the trial, which
was divided into three parts: Kosova, Croatia, and Bosnia. Although Armatta follows this structure, she also organizes the book around unifying themes such as the type of witness or crime charged. For example, she groups together types of witnesses such as victims and former military officers. She also dedicates a chapter to the charge of genocide. Further, she organizes some of her analysis by the type of evidence provided, such as research reports about conflict and military records of the Yugoslavia People’s Army (JNA). Using this structure, Armatta distills important lessons from the trial that are applicable to future war crimes prosecutions.

A. The War in Kosova

The ICTY gave the prosecution four and a half months to present evidence of Milošević’s crimes in Kosova. With the evidence, the Prosecution intended to prove that Milošević was responsible for Serbian forces’ war crimes and crimes against humanity through his planning, ordering, or aiding and abetting of their commission by virtue of his superior position to the perpetrators. To prove its case, the Prosecution called victims, journalists, foreign diplomats, generals, researchers, experts, Serbian generals and politicians, and Kosovar politicians.

The trial began with the testimony of a series of victims from Kosova, who explained how their families had been killed and how Serbian police had surrounded their homes and forced them to leave. Commentators have criticized the victims’ testimony as a “disappointment” to the public, who were expecting well-known political leaders and other “big names” to start the trial. Armatta rightly questions these reactions to the victim testimony, pointing out that this testimony was both necessary to prove the charges and also supplied the stories and substance at the heart of the trial.

Time constraints limited the Prosecution’s ability to present victim testimony. As a result, the Prosecution submitted the majority of victim testimony as written testimony, which the court allowed so long as it did not directly concern Milošević’s actions and as long as the accused had the opportunity to cross-examine the witnesses in court. This allowance largely defeated the benefit that written testimony provided in saving time because Milošević insisted on personally cross-examining almost every witness. It also had the unfortunate effect of generally denying the victims’ ability to tell their story through direct examination.

In addition to the lack of oral victim testimony, the Kosova-centered part of the trial introduced a difficulty that would plague the rest of the trial—

1. I follow Armatta’s decision to use the Albanian “Kosova” instead of the Serbian spelling, “Kosovo” because of her respect for its status as an independent state.

Milošević’s pro se representation. Although Armatta acknowledges the allowances the court should make for pro se defendants, she uses the first few chapters of *Twilight of Impunity* to convincingly detail the problems created by Milošević’s pro se defense. Milošević employed sarcasm to ridicule witnesses and often substituted commentary for genuine questions. Moreover, he used intentional distortions to attempt to discredit witnesses. Although advised several times on the rules of cross-examination, Milošević ignored any directions on proper examination conduct and acted disrespectfully toward both the Tribunal and witnesses. Armatta argues that Milošević simply was not trying to defend himself, did not believe in the legitimacy of the Tribunal, found no reason to follow its rules, and used the trial as an opportunity to speak whenever possible. Further, within the first few weeks of the trial, Milošević required time off for health issues, an indicator of the many delays his medical condition would cause the trial. The Tribunal frequently responded to the delays caused by Milošević’s health and his abuse of the trial process by making threats to appoint counsel or to cut off his privileges, but the Tribunal never followed through with these threats. These issues occurred time and again throughout the trial.

**B. The War in Croatia**

As the Prosecution switched its focus to Milošević’s crimes in Croatia, it warned the court of two major obstacles before it in prosecuting the former head of state. The first was the Serbian government’s lack of cooperation in producing documents. The second was the risk posed to witnesses by the continued presence of Milošević’s powerful and sometimes violent supporters and allies in the Balkans. The danger posed by these supporters meant that many witnesses, including key insider witnesses, would have to testify in closed sessions with additional protective measures.

After several months of foiled efforts, the Prosecution requested the Tribunal’s help in obtaining documents from the Serbian government. In its petition to the Tribunal, the Prosecution described the Serbian government’s minimal cooperation: the government only delivered documents that the Prosecution could identify with sufficient specificity so as to preclude denial of their existence. Even in these situations, the Prosecution had to continually press the government to produce the documents. However, the Serbian government’s resistance seemed to hinder only the Prosecution. At trial, Milošević often produced documents that the Prosecution had been requesting for months. Armatta highlights this occurrence in the cross-examination of Ante Marković, the former President of the Former Yugoslavia. Milošević surprised both the witness and the Prosecution by producing Marković’s daily appointment calendar from the latter half of 1991: a document both the witness and the Prosecution had been seeking for years. Despite these incidents, when the Prosecution made an appeal for intervention, the Tribunal nevertheless granted Serbia two more months to produce documents before making a decision on the matter. Be-
cause documents had to be introduced through the correct witnesses, the Tribunal’s decision effectively allowed the Serbian government to wait until the very end of the Prosecution’s case to surrender documents, at which point it was too late to introduce them into evidence.

In addition to the lack of state cooperation, the Prosecution had to take imperative measures to protect witnesses, many of whom already had been threatened. The Tribunal used closed sessions, voice and face distortions, and pseudonyms. Despite these measures, the Tribunal still had to remain particularly wary of Milošević, who often attempted to disclose identifying information for protected witnesses during the trial. For example, in an open session with a witness who was testifying under several protective measures, Milošević attempted to reveal the identity of the witness through cross-examination. The witness was a former member of the Arkan Tigers, a paramilitary force that was known for being particularly brutal. Despite warnings from the Prosecution, it was not until the witness himself warned the court he could no longer answer without identifying himself that the court closed the session.

As in the Kosova portion of the trial, the court again faced serious time constraints when trying Milošević’s action in Croatia. The court struggled to balance the rights of the accused with the Prosecution’s initiatives to submit more evidence. The court generally allowed the increased volume of written evidence; however, the court also allowed the defense the right to cross-examine witnesses on written evidence. Following a pattern of behavior established in the early part of the trial, Milošević used his right to cross-examine written evidence frequently, ultimately decreasing the time the Prosecution had attempted to save.

C. The War in Bosnia

After an extension of one hundred days, more than half of which were attributable to Milošević’s illnesses, the Prosecution began its Bosnia case. This phase of the trial brought new legal issues in addition to the continued difficulties of Milošević’s pro se defense and poor health. Unlike the crimes in Kosova and Croatia, the Bosnian case included charges of genocide. To frame the charges, Armatta provides a useful discussion of genocide jurisprudence, advocating for a broader conception of responsibility in general. She notes that in the Milošević case, the Trial Chamber dismissed the amici’s motion to acquit on genocide and found there was sufficient evidence to support the genocide charge. In the decision, the Trial Chamber followed a recent ICTY Appeals Chamber decision and expanded liability for genocide to include actors who may not have genocidal intent but know that their action or inaction was necessary for the resulting genocide. Armatta points out that the Trial Chamber took an expansive view of genocide, indicating that the Chamber, if given the opportunity to make a final ruling on genocide, would have probably favored a broader conception of responsibility for genocide.
As the trial continued, the Prosecution encountered further evidentiary problems. The Prosecution often confronted clearly fraudulent documents that were anonymously delivered to its offices. The Prosecution had to report these documents to Milošević, regardless of their authenticity, and Milošević frequently used them in cross-examination. The court allowed the use of these documents, reasoning that they would be considered for “what they are worth.”

Armatta is critical of this lax admission policy, suggesting that it undermined the truth-seeking function of the Tribunal as well as the Tribunal’s credibility. Armatta argues that although allowing looser standards for a pro se defendant may promote legitimacy by inducing participation in the proceedings and reducing the chance of a “show trial,” the Tribunal also risks sacrificing legitimacy if the public perceives that the Tribunal enables the defendant to defraud and manipulate the suit. Finding this equilibrium is difficult in a war crimes trial, but the rights of the accused must be balanced with the interests of the public.

Milošević’s health problems continued throughout this portion of the trial, and rumors circulated with increasing frequency that he was purposefully manipulating his medications in an effort to derail his health. After one physical and mental examination, Milošević’s doctor suggested limiting the trial to three days a week, which slowed the its pace considerably. Despite these problems, the Tribunal remained resistant to the appointment of counsel or even less drastic measures, such as limiting the time it allowed Milošević for cross-examination. Despite many obvious efficiency issues, the Tribunal feared infringing on the defendant’s rights. By the close of the Prosecution’s case, it was clear that Milošević’s tactics had significantly inhibited its presentation of evidence. Because of time constraints and Milošević’s manipulations, the Prosecution was unable to fully present evidence on the crimes in Sarajevo. Additionally, much of the important genocide testimony was written and unavailable to the public. As the Prosecution’s case came to a close, the problems presented by Milošević’s pro se defense were clear: it had sacrificed efficiency, the legitimacy of the proceedings, and public access to material presented in trial.

D. Milošević’s Defense

After two years, the Prosecution rested its case and the Defense’s case began. Initially, the Defense faced numerous setbacks. The presiding judge on the case fell ill, resigned, and died. After the Tribunal appointed a replacement judge, Milošević’s continuing health problems further delayed the trial. Finally, in September 2004, the court ordered the appointment of counsel for Milošević. As a result, Milošević stopped all cooperation with the court and organized a witness boycott. The Trial Chamber appointed two amici curiae as his defense counsel, but Milošević refused to communicate with them. Defense counsel was

3. Armatta, supra note 2, at 238.
left to defend Milošević without the cooperation of the accused and with only a few witnesses who actually agreed to come to the Tribunal. The Appeals Chamber eventually overruled the amici curiae defense system, which effectively put Milošević back in control and left the appointed defense counsel standing by in case his health deteriorated.

At the time of this Appeals Chamber decision, the law on self-representation remained unsettled. Armatta points out that the Appeals Chamber, like the Trial Chamber, focused exclusively on the impact of Milošević’s health on the trial when deciding whether or not to appoint counsel. However, Armatta argues that the court should have also considered additional factors, such as the integrity of the Tribunal. She reasons that Milošević’s health problems, combined with his disrespect for the Tribunal, apparent manipulation of medication, and obstructionism presented a sufficiently strong case for the appointment of counsel.

Milošević’s defense continued to ignore court rules and focus on purely political defenses. Armatta demonstrates how Milošević’s defense not only wasted time, but it was also ineffective. Points that could have been part of a legitimate legal defense were buried within attempts to characterize the proceedings as a political trial meant to prosecute the Serbian people. Even amidst the convoluted nature of Milošević’s presentation, Armatta identifies some potentially viable defense theories from his statements: (i) Kosovars were forced to leave and often killed by NATO bombings, which were also responsible for property damage to mosques, private homes, and cultural heritage sites; (ii) Milošević did not have control over the Bosnian Serbs and Croats and had nothing to do with any crimes committed by those groups; and, (iii) all supplies Milošević gave to the Bosnian Serb armed forces constituted humanitarian aid and, as such, he is not responsible for the acts of “madmen.” These potential defenses remained unsupported as Milošević continued to introduce fabricated evidence, a parade of military officers who simply agreed with all of his statements, and other witnesses who lacked credibility.

Milošević’s health deteriorated during this time, largely due to his intentional manipulation of prescribed medicine and use of smuggled, unauthorized drugs that raised his blood pressure. As his health worsened, he attempted to get an extension for his defense case and to arrange for a provisional release to a clinic in Moscow for treatment and testing. The court denied both proposals, and a month later, on March 11, 2005, Milošević was found dead in his cell. Medical reports indicated that he died of a heart attack caused at least in part by his refusal to take prescribed medicine and his use of contraband medication designed to counteract the effects of his prescribed medicine. He was able to sneak the non-prescribed medication in through “privileged” visitors who met with him in a private office space, an extra allowance given to him as a pro se defendant. Armatta’s claim is that Milošević’s final manipulation of his pro se status confirmed too late that the Trial Chamber had correctly imposed counsel, and that
the Appeals Chamber should not have reversed that decision. Armatta clarifies that although it appears that the Appeals Chamber facilitated Milošević’s actions by allowing his pro se representation to continue, it was ultimately Milošević who decided to put his own life and health at risk in an attempt to compel a release to Moscow.

III.
ANALYSIS

Armatta effectively illustrates the unique challenges of prosecuting a head of state. In terms of evidence, one difficulty lies in proving a direct connection between the accused and the crimes committed. Such proof requires massive amounts of evidence documenting the actual crimes, knowledge of the crimes, and the level of control or influence over perpetrators. When the accused’s former government controls these documents, it may be difficult for the prosecution to gain access to evidence. Further, despite confinement, the accused may continue to exert influence and intimidation, precipitating a need for advanced witness protection and preventative measures to combat contradictory statements made by insider witnesses. These challenges multiply when the defendant chooses to represent himself, as illustrated by both Milošević and the more recent case of Radovan Karadžić, the former President of the Serbian Democratic Party (SDS).4

The cost, pace, and frequent legitimacy problems faced in Milošević’s trial seem traceable to the decision to allow a pro se defense. Armatta convincingly challenges criticisms that the Prosecution should have narrowed its case and demonstrates that rather than prosecutorial error, Milošević’s tactics, and the Tribunal’s resulting failure to circumscribe his actions, fundamentally slowed the trial’s pace. In making the case for increased control over Milošević, Armatta suggests that in a war crimes trial, the tribunal must balance the interests of the defendant against those of the victims and the legitimacy of the tribunal. Although a war crimes trial focuses primarily on the deeds of the individual, its procedure and outcome profoundly impact victims, serving as an acknowledgment of their suffering and setting precedent for similar trials in the future.

Although Armatta touches on the impact war crimes trials have on victims, more attention to the topic would strengthen her analysis. She readily points out potential benefits the trial may have for victims—especially those who testified—but she does not fully address how or whether these benefits might reach the majority of victims who never had an opportunity to confront Milošević in The Hague. While Armatta discusses the importance of a public record, the impact of various procedures on the creation of a public record, and, in particular,

the use of written evidence and closed sessions, she does not explain how such a public record ought to be made accessible to victims.

Therefore, although sensitive to the interests of victims, Armatta could better explain victims’ receptions of trials, identify outreach efforts made by the ICTY, and demonstrate how victims were able to watch the trial and follow it in a meaningful way. More information from the victims’ perspectives would serve as a useful complement to the convincing explanation of the importance of considering the interests of the victims in a war crimes trial.

IV. CONCLUSION

Armatta’s Twilight of Impunity provides an excellent account of Milošević’s trial while both highlighting the challenges of prosecuting a head of state and unpacking the typical “lessons learned” commentaries that have emerged from Milošević’s trial. As Armatta indicates, although the trial suffered debilitating setbacks, it in no way forecasts the end of international justice. Indeed, the trial has become increasingly relevant as efforts to improve international accountability have multiplied. For example, the International Criminal Court issued an arrest warrant for Omar al-Bashir, the current President of Sudan, and had opened an investigation into the actions of President Muammar Qaddafi in Libya prior to his death. In this context, Armatta’s book is a useful contribution to a growing body of literature that explicitly addresses the prosecution of heads of states. Armatta recognizes the special difficulties of prosecuting these officials and offers a nuanced analysis of the decisions made by the Prosecution and judges in one case. In addition to procedural elements, such as pro se representation, Armatta highlights the practical difficulties of prosecuting someone who once exercised complete power and still retains great influence in his home country.