Eurasia Group President Ian Bremmer
China's Bumpy Road Ahead

Edward Fishman, The Peter and Katherine Tomassi Essay
History as Nemesis: Britain, German Unification, and
the European State System, 1870–1875

Also Inside:
Promoting the Rule of Law in China
Climate Change and the Ramsey Model
Between Global Citizenship and Empire
Recent Developments in Immigration Law
Post-Merger Bundling in the Cable Industry
Toward An Integrated Approach to Microfinance
It's been another exciting year. In this climate of economic and political upheaval, it seems that anything is possible. Received wisdom is challenged daily. The eurozone crisis, the threat of a double dip recession in the United States, and the Occupy Wall Street movement on the one hand, and the Arab Spring and continued growth in emerging economies on the other, suggest that this moment may mark a historic shift in the global balance of power. In such unpredictable times, when we can't rely on textbooks or habit, thoughtful research and analysis of politics and society is of the utmost importance.

§

This issue's essays provide just that. This edition of the Journal of Politics & Society begins with a guest essay by Eurasia Group President Ian Bremmer. Why, he asks, do so many take China's rise for granted? Bremmer points out that a host of little-discussed phenomena, from frequent outbursts of civil unrest to the looming cost of caring for a large retiree population, threaten to undermine China's sought-after superpower status.

China is hardly the only country with domestic challenges ahead. Many of this issue's essays examine the relationship between law and society. In the United States, demographic issues also stand at the center of the country's struggle to maintain its long-term global dominance, but reforming immigration policy is unlikely in the current political climate. Sarah Paige examines how recent developments in immigration law differ from Supreme Court doctrine, and calls attention to the way immigration adjudication increasingly resembles that of the criminal system. With a similar focus on the relationship between law and society, Rush Doshi examines the misuse of China's criminal law Article 306, and makes recommendations for reforms in the Chinese criminal justice system.

Adam S. Sieff, on the other hand, addresses more normative notions of legality. It is a false choice, he argues, between cosmopol-
itan internationalism grounded in the codification of international law and hegemonic liberalism. Sieff offers an original interpretation of Hans J. Morgenthau's political thought that aspires to put the concept of the political back into theories of international order.

It would be naive to evaluate our world without attempting to understand the complicated past. Edward Fishman investigates historical visions of international order in this issue's Peter and Katherine Tomassi Essay. Fishman traces the shifts in British public opinion that led to a reappraisal of German unification and power between 1870 and 1875.

Finally, we have three essays that examine specific policy questions relevant to development and economics. Eva Orbuch, working in Bolivia and Peru, finds that microfinance organizations that provide social services along with their regular credit services can offer distinct advantages over their credit-oriented counterparts. Yangbo Du evaluates the success of the Ramsey economic growth model in accounting for the impact of climate change policies, arguing for an alternative framework that considers problems of intergenerational equity in balancing the costs and benefits of climate policy. Casting a quantitative eye on the cable news industry, Sarah Amanda Levis analyzes the 1996 merger between Time-Warner and Turner Broadcasting. Building a model that takes into account heterogeneous consumer preferences, she counters the view that the bundling of independent goods restricts entry, revealing that bundling can, at times, facilitate entry and boost consumer welfare.

§

In the Helvidius Group's twenty-third year, I am excited to announce that this edition marks the successful transition from an annual to a semiannual format. The Journal will now provide an even bigger platform for the best undergraduate research in the social sciences.

Balancing this transition with the doubled pace of our normal selection and editing process would have not been possible without the incredible work of the Helvidius Group's thirty members. I am grateful for their faith in our ability to grow as an organization, and
for their professionalism, hard work, and dedication to the social sciences. Good social science research asks hard questions, solves difficult and urgent puzzles, and, in doing so, helps to improve human life. As it continues to grow, the Journal pulls together the sometimes-disparate disciplines that share this goal. I am confident this publication will continue to be a fitting platform for younger generations dedicated to this lofty enterprise.

Alex Merchant
Editor in Chief

New York City
December 2011
Guest Essay

CHINA’S BUMPY ROAD AHEAD: UNREST, INFLATION, AND AN AGING POPULACE THREATEN CHINA’S TOUTED DOMINATION

Ian Bremmer

When exactly will China take over the world? A scan of recent headlines suggests the moment of truth is coming closer by the minute. China will become the world’s largest economy by 2050 says HSBC. It’s 2040 say analysts at Deutsche Bank. Try 2030, say the Institute for International Economics and the World Bank. Goldman Sachs and PricewaterhouseCoopers say it’s 2020, and the IMF said a few months ago that with a purchasing power parity calculation, China’s economy will push past America’s in 2016. There’s probably someone out there who thinks China became the world’s largest economy five years ago.

But there’s a lot of local turmoil simmering beneath the surface of China’s miracle. A traffic accident in Henan province sparks a dispute. Within minutes, police are scrambling to contain a riot involving more than a thousand people. In Hunan, farmers pushed off their land by aggressive property developers discover that local authorities are not on their side. A farmer sets himself on fire, and protests spread quickly from town to town. A chemical spill into a Chinese river cuts off water supplies to Harbin, a city of four million people, sparking public fury. In Inner Mongolia, a Han Chinese truck driver kills a local herdsman in a hit-and-run accident, and ethnic unrest flares for days. Rioting in Xinjiang province spins out of control, forcing a state Internet shutdown across an area three times the size of California. Inside China’s coastal cities, security guards sent to break up a protest by migrant workers push a pregnant woman to the ground, igniting a firestorm that only paramilitary forces in armored personnel carriers can handle.

Ian Bremmer is the president and CEO of Eurasia Group, the world’s leading political risk consulting firm. He holds a PhD in political science from Stanford University and is the author of several books, including The End of the Free Market: Who Wins the War Between States and Corporations? This essay was originally published in the Wall Street Journal on July 9, 2011 and is used with permission.
Thousands of people assault government buildings and set police cars ablaze.

Yes, China’s security services are the world’s best at containing large-scale riots. No, these protests are not yet coordinated by any form of coherent opposition to Communist Party control. It’s clear that most protests are still directed at local officials and are fueled by local grievances, and three decades of double-digit growth has earned the Chinese leadership deep reserves of public patience.

But this is a country that measures its annual supply of large-scale protests in the tens of thousands. For 2006, China’s Academy of Social Sciences reported the eruption of about 60,000 “mass group incidents,” an official euphemism for demonstrations of public anger involving at least fifty people. In 2007, the number jumped to eighty thousand. No longer published, a leak of official figures put the number for 2008 at 127,000. Today, the leaks have been plugged, and no reliable data is available.

Western observers have gotten China wrong for years. A generation ago, the crackdown in Tiananmen Square and the collapse of European and Soviet communism persuaded many analysts that the country’s leadership could not embrace globalization’s free flow of ideas and information without significantly relaxing its grip on domestic political power. Others predicted confidently that China’s authoritarians would soon sink beneath the waves of democratic change. More than twenty years later, the sun continues to rise on the Communist Party leadership, which turns out to have learned important lessons from the demise of other authoritarian regimes. Unlike their counterparts in Moscow and the satellite capitals, China’s leaders have delivered on promises of much improved living standards and greater personal (if not political) freedom. Double-digit growth has completed its third decade, and few analysts want to appear naive by again writing the leadership’s obituary.

A new storyline then emerged, one that discounted the importance of political freedom almost completely. Emerging players, led by China, were competing to win, and authoritarian rule offered them advantages that the inefficiencies of democracy simply could not match. Globalization lifted hundreds of millions of con-
sumers, from democracies and authoritarian regimes alike, into the
global economy and onto the global stage. Some analysts began to
talk of a new Asian century, one that would arrive with good news
for America. A rising Asian tide would prove strong enough to lift
all boats, including Western ones.

The financial crisis upended these assumptions, provoking
widespread fear that no boost will prove strong enough to lift
America, Europe, and Japan to new heights. Now that China’s
growth numbers have bounced back nicely, this latest version of the
story holds that China will inevitably become the world’s driving
economic power, with dire implications, both economic and politi-
cal, for the world’s advanced industrial democracies.

There’s a common theme—and a common flaw—in all these
versions of China’s future and its implications for the rest of us:
They assume there is something inevitable about the country’s fu-
ture. But China is the emerging power least likely to develop along
a predictable path. There is no credible evidence that the country
is on the brink of unforeseen crisis. Its political system has sur-
vived the worst of the global meltdown, and the state is strong and
wealthy enough to buy off or put down the many indirect domes-
tic challenges to Communist Party control. Next year’s transfer of
power to the next generation of leaders should go smoothly.

Yet, all that public anger beneath the surface gives us more
than enough warning that China will face enormous challenges in
the road ahead. The twelfth Five Year Plan—the much heralded
reform document which will set China’s policy path for the next
five years—is the Communist Party’s blueprint for correction of
the country’s massive imbalances, but there is little evidence so far
to suggest that the government will actually put the right policies in
place to be even partially successful.

This is why we should treat precise predictions of China’s tra-
jectory with extreme skepticism. Major emerging powers like India,
Brazil, and Turkey can continue to grow for the next ten years with
the same basic formula that sparked growth over the past ten. Chi-
na, on the other hand, must undertake enormously complex and
ambitious reforms over the next generation to continue its drive
to become a modern power, and the country’s leadership knows it.

For example, the financial crisis made clear that China’s dependence for growth on the purchasing power of consumers in America, Europe and Japan creates a dangerous vulnerability. Those who insist that it’s possible to map the precise arc of China’s rise seem to assume that China’s leaders can steadily shift the country’s growth model toward greater domestic consumption, by transferring enormous reserves of wealth from China’s powerful state-owned companies to hundreds of millions of new consumers. That’s quite an assumption. Despite the best efforts of policy architects in Beijing, the share of household consumption in China’s economic growth last year actually moved in the other direction, in part because there are political powerbrokers within the elite who have made too much money off the old growth model to fully embrace a new one.

Further, as China’s gap between rich and poor continues to widen, resulting social unrest will almost certainly force tighter state restrictions on free expression and free assembly. That could promote a violent backlash if rising expectations for material success aren’t met. Most dangerous of all for the ruling party’s future, factions within the government might not agree on how the state should respond to a sudden convulsion of organized unrest. (That’s what happened in the lead-up to the Tiananmen Square crackdown—and at a time before modern tools of communication made it much easier for protesters to coordinate their actions.) How real is this risk? Thirty-year growth projections don’t even try to answer this question.

China’s demographics will provide another serious challenge. The country’s labor force is becoming more expensive as China urbanizes and moves up the value chain for manufacturing. The population is also getting older, as the one-child policy and other factors leave fewer young people to join the labor force. As more Chinese reach retirement age, the need to expand and reinforce a formal social safety net to provide pensions and health care for hundreds of millions of people will add unprecedented costs.

In addition, given that so much of China’s growth is still
coming from infrastructure projects and other state-directed investments, the impact on an already overtaxed environment could shock the system. Land degradation, air quality, and water shortages are urgent and growing problems. China’s capacity to tolerate a deteriorating environment is higher than in most developing markets (to say nothing of the developed world), but the chances for an environmental incident to provoke a dangerously destabilizing event are growing by the day. The water problem alone will keep policymakers busy.

As rising prices undermine China’s low-end manufacturing capacity, a failure to bolster domestic innovative capacity at the higher-end of production could mean slower economic growth. China is outstripping other countries in the numbers of patents, but for parts of products, not systems. Major new innovations continue to come overwhelmingly from outside. Increased state spending on research and development will help, but without a social, educational, and political system that encourages innovation and entrepreneurship, change will come slowly. That’s why most of the next big developments in key emerging technologies—biogenetics, nanotechnology, and probably even alternative energy—aren’t likely to come from China.

Then there’s inflation, which hit a thirty-four-month high in May, according to China’s National Bureau of Statistics. Food prices surged 11.7 percent. An over-expansionary monetary policy, higher transportation costs related to urbanization, and large-scale wage hikes are just a few of the variables that ensure the government will have a harder time containing inflation in months to come. This is a highly personal issue inside China, as it is everywhere, and it will distract policymakers from longer-term challenges.

Finally, as popular demand, expressed online and in China’s blogosphere, plays a larger role in how Chinese policymakers make decisions, unhappy citizens will check the state’s ability to implement strategic economic and development policies. That too could limit China’s longer-term economic growth. For example, widespread discontent with an enormous water diversion project provides one recent example of the ways in which public pressure in
China can distort public policy to a degree that would not have seemed possible just a few years ago. There is also growing resistance to reforms from within the ruling party itself—especially among bureaucratically influential state-owned companies that stand to lose from the changes.

Even if China’s leadership makes major unexpected progress on domestic reform, it will find that the international environment is becoming less conducive to Chinese growth. Higher prices for the oil, gas, metals and minerals China needs to power its economy will weigh on growth. The expansion of all those other emerging market players will only add to the upward pressure on food and other commodity prices, weighing on future growth rates and undermining consumer confidence, the most important source of China’s baseline social and political stability.

What about China’s relationship with the United States. A strong Chinese growth environment, coupled with an unsustainable fiscal environment, higher US unemployment, and weaker US consumption, will generate friction between the world’s two largest economies—in particular, by drastically raising the likelihood of protectionism on both sides. That’s a problem for American companies looking for access to Chinese consumers, but it’s even more troublesome for the Chinese, who rely more on US fiscal stability, investment, technology, and consumption.

Finally, there are “fat tail” risks aplenty—in the international system and within China—that could create major headaches inside the country. Take the problem of nuclear proliferation, especially in North Korea. Beijing continues to engage Pyongyang, because it fears North Korean chaos and the flood of sick and starving refugees it might unleash, more than it fears North Korea’s nuclear weapons. But over time, the inability to coordinate an international approach to the nuclear problem will create less regional stability, not more.

At a minimum, China’s reform challenges will provide plenty of good reasons for us to question multi-decade projections of China’s economic trajectory and the country’s predicted rise to global dominance. As Yogi Berra once said, “It’s tough to make predic-
tions, especially about the future.” That’s truer of China than of any other country in the world.
On February 9, 1871, just over three weeks after the new German Empire was declared at Versailles, Benjamin Disraeli rose to address the House of Commons. At the time Disraeli was the Conservative leader of the opposition, and the Liberal Gladstone Ministry had maintained strict neutrality during the Franco-Prussian War. Although Disraeli also opposed British intervention, he recognized in Prussia’s victory far-reaching consequences for the European state system, stating “It is no common war, like the war between Prussia and Austria, or like the Italian war in which France was engaged some years ago; nor is it like the Crimean War.” Disraeli further pronounced:

This war represents the German Revolution, a greater political event than the French Revolution of last century ... There is not a diplomatic tradition which has not been swept away. You have a new world, new influences at work, new and unknown objects and dangers with which to cope, at present involved in that obscurity incident to novelty in such affairs...but what has really come to pass in Europe? The balance of power has been entirely destroyed, and the country which suffers most, and feels the effects of this great change most, is England.¹

To the modern reader, Disraeli’s speech seems remarkably prescient. As A.J.P. Taylor could reflect in 1954, “Sedan* marked the end of an epoch in European history; it was the moment when the myth of la grande nation, dominating Europe, was shattered forever. The Balance of Power was startlingly altered.”² And indeed, not only was the myth of French predominance destroyed, but a new era of European history, one in which Germany would be the pivotal power, was born.

Edward Fishman is a candidate for the MPhil in international relations at the University of Cambridge. He graduated Phi Beta Kappa with a BA in history from Yale University in 2011, where he won the Winifred Sturley Prize in European history.

* The Battle of Sedan marked a major turning point in the Franco-Prussian War.
But in acknowledging Disraeli’s foresight we ought not assume that either he or any other British statesman was confident that the newly unified Germany would be a menace to England, much less that British diplomacy should aim to counterbalance the fledgling central European empire. It was only after a long series of disturbing German actions, culminating in the “war-in-sight” crisis of 1875, that the British fully came to view the “German Revolution” as hurtful to their interests. In fact, for decades educated English opinion had favored the emergence of a unified German state, and from the outbreak of the Franco-Prussian War British sympathies tended toward the Prussian side. It is not difficult to see why such sentiments prevailed. Reared in the tradition of Britain as a world power, British statesmen of the era were more concerned with Russian advances into Central Asia and French designs in the Mediterranean than alterations in the continental balance of power. Moreover, a strong and united Germany, situated between Russia and France, would distract England’s two great rivals and deter them from extra-European adventures. In the middle of the nineteenth century, Lord Palmerston became the chief British proponent of German unity, believing that Britain and the German states faced common threats in Russia and France. “England and Germany therefore,” Palmerston avowed, “have mutually a direct interest in assisting each other to become rich, united and strong.”

Thus, from a geopolitical standpoint, a long British tradition supported the idea of a strong and unified Germany in the center of Europe. But British sympathy for the cause of German unification was not merely rooted in calculations of geopolitics; rather it extended to the realms of culture, religion, and ideology. The land of Kant, Goethe, and Schiller, Germany was home to an illustrious intellectual tradition very fashionable in nineteenth-century England, and Prussia, the largest German kingdom, was the only European great power that shared Britain’s Protestant religious heritage. Throughout the nineteenth century British intellectuals such as Samuel Taylor Coleridge, Matthew Arnold, George Eliot, John Stuart Mill, and Thomas Carlyle had propagated the image of Germans as “noble, intelligent, peaceful, civilized, and profoundly
Furthermore, by the 1860s, the doctrine of nationalism had become part and parcel of British liberalism, and the British had already enthusiastically advocated Greek and Italian unity. A Germany unified under Prussian auspices, British statesmen believed, would be liberal, peace-loving, and England's "natural ally" on the continent.

One such statesman was Robert Morier, who became British secretary at Darmstadt just after the Austro-Prussian War of 1866. He followed the war with great interest, but as a liberal and a proud Germanophile, he was disturbed that Bismarck, whom he called "one of the most sinister figures that has ever been painted on the canvas of history," had co-opted the cause of German unity. Indeed, most British intellectuals and statesmen, despite their positive feelings for the German people, considered Bismarck reactionary and tyrannical. Nevertheless, steeped as they were in the "Whig interpretation" of history, these same Englishmen assumed that Bismarck, like so many other "enemies of progress" before him, would eventually be cast into the dust heap of history. "When any institutions come directly into contact with the spirit of the time," wrote Liberal politician Mountstuart Grant Duff, "they may resist for five years, or ten, or twenty, but down they must go in the end." Thus Bismarck did not substantially alter English opinion on German unification.

The British Foreign Office welcomed Prussia's victory over Austria. Although Morier and other English liberals had hoped that Austria would be expelled from Germany and Italy by "the natural course of events," they were nonetheless pleased to see "the congenital malformation of Europe" remedied. Bismarck's choice of means for effecting unification were reprehensible, but both the

---

* See Herbert Butterfield, *The Whig Interpretation of History* (New York: W.W. Norton, 1965), p. 5: "[T]he historian tends in the first place to adopt the whig or Protestant view of the subject," writes Butterfield, "and very quickly busies himself with dividing the world into the friends and enemies of progress."

† The "natural course of events" refers to "the ripening of political history in Prussia." Violent war against Austria was considered unnatural—the product of Bismarck's ideas alone. Robert Morier to M. Grant-Duff (June 9, 1866), Rosslyn Wemyss, ed., *Memoirs and Letters of the Right Hon. Sir Robert Morier*, From 1826 to 1876, Vol. II. (London: Edward Arnold, 1911), p. 66. Italics added.
national principle and the enlightened Weltanschauung of the German people ensured that a unified Germany would foster a more peaceful European order. Hence Morier could write to his father, “I entirely concur in your wish to see Bismarck hung,” and nevertheless support the Prussian minister’s war against Austria.*

Despite British inclinations toward Prussia and the cause of German unification in the 1860s, these sentiments never translated into action. British foreign policy toward Europe was decidedly non-interventionist during the decade of Bismarck’s wars. Most historians agree that the first Gladstone Ministry, which came to power in late 1868, was one of the most inward-looking in the annals of Victorian Britain.† And fortunately—at least it seemed—the international situation was calm and clear. A few months after Gladstone entered office, Bismarck wrote him a letter about the European situation. “The political horizon seen from Berlin appears at present so unclouded that there is nothing to report,” professed the Prussian minister. Colonel Walker, the British military attaché at Berlin, reasoned that the “unification of Germany was far distant.” Preoccupied with affairs at home, the Gladstone Ministry readily accepted such reports. When Lord Granville became foreign secretary after Clarendon’s death in June 1870, Edmund Hammond, the under-secretary, told him “he had never during his long experience known so great a lull in foreign affairs, and that he was not aware of any important question that [Granville] should have to deal with.”

Yet, however “unclouded” the horizon appeared, tensions were brewing in Europe. Napoleon III, emperor of France since 1852, had spent his career championing the nationalist cause throughout

---

* Morier believed that “A signal victory on the part of Austria in the present struggle would throw Europe back three generations.” Robert Morier to his father (June 20, 1866). Wemyss, Memoirs and Letters of Morier, pp. 67–68.

the continent, and the British, in light of their fixation on Belgian neutrality, looked upon him with suspicion. Napoleon welcomed Prussia’s rapid victory over Austria in 1866; a Prussian state organized on a national basis, he assumed, would inevitably be liberal and, consequently, estranged from his enemy, Russia. But many Frenchmen were angry, believing that France must achieve its own coup to offset Prussia’s gains. Napoleon thus instructed one of his top diplomats, Count Benedetti, to negotiate a secret accord with Bismarck. A draft agreement was drawn up in which France would ensure the integrity of Prussia’s territorial acquisitions in exchange for carte blanche in Luxembourg and Belgium.* However, the so-called “Benedetti Treaty” was never signed, so Napoleon was forced to look elsewhere to satisfy disgruntled French opinion. His next interlocutor was the king of Holland, with whom he negotiated the annexation of Luxembourg to France. But at the last minute, the Dutch king lost his nerve and refused to sign the treaty without first informing Prussia. The result was predictable: a torrent of nationalist feeling in Germany, followed by a conference of the powers affirming Luxembourg’s neutrality. Although seemingly trifling, the Luxembourg Crisis was a major watershed in the diplomatic history of Europe. Henceforward it seemed no longer possible for France to accept gracefully the emergence of a strong Prussia–Germany on its eastern border.18

With Franco-Prussian relations so bitter, French public opinion calling for revanche for Sadowa, and constant military preparations in both France and Prussia, it is astonishing that the British Foreign Office thought the international landscape so placid in early summer 1870.† Less than three weeks after Hammond spoke of an unprecedented “lull” in foreign affairs, the issue of the Hohenzollern candidature had exploded into a full-fledged Franco–Prus-

---

* Luxembourg and Belgium were considered natural areas in which Napoleon III might seek to augment the French Empire. That such an agreement could even be proposed indicates how far continental perceptions of British power and willfulness had fallen by 1867. A. J. P. Taylor, The Struggle for Mastery in Europe, 1848-1918 (Oxford: Claredon Press, 1954), pp. 173-74.

† Sadowa (Königgrätz) was the site of Prussia’s major battlefield victory of the Austro-Prussian War. During Prussia’s next war, against France, Bismarck said to Count Bernstorff: “Even our victory at Sadowa roused bitterness in France; how much more will our victory over themselves.” Quoted in Ibid., p. 217.
sian war. From the outset of the conflict British opinion across all levels of society regarded Napoleon as the aggressor. Queen Victoria, always partial for things German over French, was a committed advocate of the Prussian side. However, despite her sympathies and the pleadings of her daughter, the Crown Princess of Prussia, the Queen recognized the necessity of British neutrality. “We must be neutral as long as we can, but no one here conceals their opinion as to the extreme iniquity of the war, and the unjustifiable conduct of the French!” wrote the Queen to her daughter, “… the feeling of the people and country here is all with you … And need I say what I feel?”

British liberals were particularly pro-Prussian in their leanings, as the war seemed to them a selfish and aggressive crusade by Napoleon against popular German national feeling. One such liberal, Gladstone, suspected that the French people might rise up against their emperor: “If ever there was a Government-made War, it is this: & France may call the author to account.” Another, Morier, resented the Gladstone Ministry for permitting the continuation of British arms sales to France. “France draws the sword to assert her political preponderance over Europe. Germany draws the sword to assert her national existence. But the result will be that the preponderance of Germany over Europe for centuries to come will take the place of French preponderance,” opined Morier. “We sit by like a bloated Quaker, too holy to fight, but rubbing our hands at the roaring trade we are driving in cartridges and ammunition.” To liberals like Morier and indeed most Englishmen, Germany represented progress, while Napoleon personified reaction; the German people were fighting for liberty, while the French emperor was striving for self-gratification. It is thus understandable that Gladstone thought the French people might topple Napoleon, and that Morier could censure the British government for selling weapons to the French side. Free trade was supposed to be a civilizing principle, not an excuse for profiting off of Napoleon’s unjust war against the German people.

— Morier was concerned that the arms sales were giving Britain a bad name in Germany. Robert Morier to his father (July 31, 1870). Wemyss, Memoirs and Letters of Morier, p. 157.
The only continental issue that really concerned the Gladstone Ministry was Belgian neutrality. When The Times published a leaked copy of the “Benedetti Treaty” on July 25, in which Napoleon’s Belgian ambitions were laid bare, British opinion, both popular and official, moved decisively against France. Meanwhile, defying a great portion of British expectations, the Prussian army won victory after victory over the French in August. Disraeli relished in the irony of Napoleon’s fortune: “This collapse of France has all come from the Emperor’s policy of nationality… The Emperor started this hare in order that he might ultimately get Belgium. Belgium is safe and France is smashed!” On September 1, Prussia’s victory became total on the battlefield of Sedan, where Napoleon surrendered himself and his entire army the next day. The Second French Empire had been driven out of existence in just over a month, and the question of a unified and powerful German state in the center of Europe, for all intents and purposes, ceased to be one. At least in form, the hopes of Palmerston, Morier, and the British liberals had come to fruition.

But less than four years after Sedan, when it seemed that Germany and France were again plunging headlong toward war, Britain’s response was strikingly different. During the “war-in-sight” crisis of 1875, it was not only British public opinion that was against Germany; the Foreign Office went so far as to intervene to stay Bismarck’s hand. How could the British position have changed so drastically in such a short period of time? After Sedan the birth of a German empire was all but an accomplished fact; the character of that new empire, however, remained an open question.


Robert Morier, perhaps as accurately as any other Englishman, predicted the course of the Franco-Prussian War. When news reached England that the two sides were mobilizing, Morier was at a dinner party. The guests knew Morier to be an expert on German issues, so they solicited his opinion on the coming conflict. After
denouncing Britain’s failure to prevent the outbreak of war, Morier proclaimed: “France and Germany will have to fight it out alone, it will be the most horrible war the world has ever seen, and it will end by France being thoroughly and completely beaten.” The prediction was unorthodox, but subsequent events soon proved its wisdom. Morier was also prescient when it came to the subtler aspects of the war. As he wrote to his father after the early Prussian victories, “[W]ith one voice the whole [German] nation is demanding vengeance, and this demand will, I fear, take the form of a demand for French territory, for Alsace–Lorraine, the very worst fault the victorious Germany could commit.”

The issue of Franco-Prussian peace terms—and particularly the subject of Alsace-Lorraine—would be the hinge on which British opinion would turn over the next several months. After Napoleon’s surrender at Sedan, the French formed the Government of National Defense, which was determined to continue the war, and whose foreign minister, Jules Favre, declared proudly that France would not cede “an inch of her territory or a stone of her fortresses.” At this point the war ceased to be, as Gladstone had described, “Government-made”; it was a war of nations, and the conventions of warfare gave way to brutality and carnage.

Most Englishmen, who supported Germany’s claim to national existence and considered the war a Napoleonic invention, naturally assumed the fighting would stop after the French emperor was removed. Considering Napoleon a threat to the peace of Europe, and thinking that French mores had become indulgent and base, the British hoped Sedan would be an edifying moment for France. On the other hand, the “noble” German victors were expected to reach equitable peace terms and begin their reign as a rock of stability in Europe. “Such a downfall is a melancholy thing, but it is meant to teach deep lessons; may we all learn what frivolity, conceit and immorality lead to!” wrote the Crown Princess a few days after Sedan. “It would be grievous for Art’s sake for that beautiful capital [Paris] to suffer. I trust it will not come to that.”

But the war went on, and Paris was laid under siege on September 19, 1870. Three days later, Bismarck announced in a cir-
cular telegram Prussia’s intentions to annex territory in Alsace-Lorraine. While he may have been more concerned with home affairs than foreign, Gladstone, like other British liberals, regarded the principle of nationality and the sanctity of international treaties and norms as matters of paramount importance; indeed his commitment to these issues was as strong as any of his contemporaries. The Prussian circular assailed Gladstone’s dearest creeds. Recent European practice had established the principle that territorial transfers were not to take place without the consent of the inhabitants. Before Savoy and Nice were officially incorporated into France in 1860, plebiscites were held so that the populations themselves could determine their national destinies. After receiving the Prussian circular, Gladstone voiced concern to Granville: “[It] appears to me that Count Bismarck’s paper raises questions of public right, in which all Europe has a common interest, and it is impossible for us to receive in silence.”

Gladstone was stoked into action. He wrote a series of memoranda to his cabinet urging British diplomatic intervention to prevent the annexations, which he surmised would affect 1.25 million people. “The transfer of the allegiance and citizenship … of human beings from one sovereignty to another, without any reference to their own consent,” ruminated the prime minister in the first memo, “has been a great reproach to some former transactions in Europe; has led to many wars and disturbances; is hard to reconcile with considerations of equity; and is repulsive to the sense of modern civilization.” Meanwhile, against Granville’s advice, Gladstone wrote a lengthy treatise on the subject for the Edinburgh Review. In the article, he proclaimed “a new law of nations [that] is gradually taking hold of the mind.” This “law” recognized the right of national self-determination, commanded the peaceful and permanent resolution of conflicts, and declared “the general judgment of civilised mankind” as the “tribunal of paramount authority.” Napoleon’s aggression had violated the law, and for this he was justly censured. If Germany proceeded to transgress by annexing Alsace-Lorraine against the will of the inhabitants, it too should expect reproach.*

* As Gladstone wrote of his “law” of nations: “It has censured the aggression of France; it will
Although the article was published anonymously, the identity of its author was never a mystery to the British public. Within the cabinet Granville led opposition against Gladstone’s plan for intervention, asserting that it would only damage British prestige to stand up for general principles with neither the will nor the power to give them force. Yet Gladstone, repulsed by the idea of transferring “human beings like chattels,” refused to concede defeat.

In Prussia there were also voices opposing the annexations, but they were few and severely outmatched. One such dissenter was Johann Jacoby, a Prussian-Jewish politician, who organized a public outcry. Jacoby and his counterparts were hastily imprisoned in September, shocking British liberals who tended to idealize their German “cousins.” Queen Victoria was forced to reexamine her own views when her friend, Queen Augusta of Prussia, wrote her a letter voicing misgivings about the prospective peace terms. “The boundary line of language should in particular be maintained, and no really French territory be claimed,” expressed Augusta, “… we must ever pray for an honourable and blessed peace!” Such sentiments, however, were drowned out by the mass of German opinion across the political spectrum demanding reward for the heroism of the fallen. Even the Crown Prince Frederick, whose liberal credentials were well established in England, could write in his personal diary, “The annexation of Alsace and perhaps of a part of Lorraine, is surely well earned by the sacrifices Germany has made.”

Disturbed by the volume of German opinion favoring annexation, and defeated in his hopes of intervention by his own cabinet, Gladstone sought to inject his voice into the discussion through private channels. The renowned German liberal intellectual Friedrich Max Müller, then professor of philology at Oxford, was both a correspondent of Gladstone’s and a friend of Bismarck’s personal secretary, Heinrich Abeken. Assuming that Müller was of the same mind, Gladstone sent him a letter: “I want to ask you what we are to think of the Alsace and Lorraine question,” wrote the prime minister. “It would surprise me to find that you thought these people [could] properly be annexed to Germany, if their heart is in censure, if need arise, the greed of Germany.”

---

37 Yet Gladstone, repulsed by the idea of transferring “human beings like chattels,” refused to concede defeat.

38 In Prussia there were also voices opposing the annexations, but they were few and severely outmatched. One such dissenter was Johann Jacoby, a Prussian-Jewish politician, who organized a public outcry. Jacoby and his counterparts were hastily imprisoned in September, shocking British liberals who tended to idealize their German “cousins.”

39 Queen Victoria was forced to reexamine her own views when her friend, Queen Augusta of Prussia, wrote her a letter voicing misgivings about the prospective peace terms. “The boundary line of language should in particular be maintained, and no really French territory be claimed,” expressed Augusta, “… we must ever pray for an honourable and blessed peace!”

40 Such sentiments, however, were drowned out by the mass of German opinion across the political spectrum demanding reward for the heroism of the fallen. Even the Crown Prince Frederick, whose liberal credentials were well established in England, could write in his personal diary, “The annexation of Alsace and perhaps of a part of Lorraine, is surely well earned by the sacrifices Germany has made.”

41 Disturbed by the volume of German opinion favoring annexation, and defeated in his hopes of intervention by his own cabinet, Gladstone sought to inject his voice into the discussion through private channels. The renowned German liberal intellectual Friedrich Max Müller, then professor of philology at Oxford, was both a correspondent of Gladstone’s and a friend of Bismarck’s personal secretary, Heinrich Abeken. Assuming that Müller was of the same mind, Gladstone sent him a letter: “I want to ask you what we are to think of the Alsace and Lorraine question,” wrote the prime minister. “It would surprise me to find that you thought these people [could] properly be annexed to Germany, if their heart is in
France.” But Gladstone was surprised when Müller responded with an ardent defense of the case for annexation: “The thousands and thousands of German hearts that lie buried in Alsace-Lorraine have made that soil German once more.” Müller’s letter revealed that German liberals could not see eye-to-eye with those of Britain on the Alsace–Lorraine question, and it perhaps reflected deeper differences in the characteristics of liberalism in the two countries. “Max Müller’s letter ... shows the hardness of the rock we have to melt or break,” wrote Granville to his disappointed chief.

As the French maintained their resistance and the Prussians were determined to march on Paris, many Englishmen began to have doubts about the forces that the Prussian victory had unleashed. “[O]ur sympathies were entirely with the Germans in the beginning of the war,” wrote George Eliot on September 26, “but I cannot help admitting to myself now, that if they had been in a higher moral condition—I mean the whole [German] nation and its government—the war might not have reached this hideous stage.” Thomas Carlyle was so frustrated by the haste with which British sympathies were turning against Prussia that he published a letter in The Times on November 18, defending the German cause. More representative of popular opinion, however, was the editor’s note printed alongside the letter, denouncing Carlyle for “treating provinces as chattels, and their inhabitants as vermin that may incidentally swarm about them.”

By the time of Carlyle’s letter, it had become evident that Prussia’s victory would have far-reaching consequences—and not all of them for the better. As France lay prostrate, the other European powers gathered like vultures, seeking to profit from the vacuum. On September 20, Italian forces occupied Rome, eviscerating the temporal power of the pope while his protector, France, was immobilized. More troubling to England was the Russian circular of October 31, declaring the czar’s intention to renounce unilaterally the Black Sea clauses of the 1856 Treaty of Paris. Consummated after the Crimean War, the clauses prohibited Russia from commanding a fleet in the Black Sea. Gladstone and other British liberals had opposed the clauses from the first, but the idea that a European
power could unilaterally free itself from treaty obligations flew in the face of the prime minister’s core principles. “It is quite evident that the effect of such doctrine ... is to bring the entire authority and efficacy of all Treaties under the discretionary control of each one of the Powers who may have signed them,” wrote Gladstone following the Russian announcement, “That is to say the result obtained is the entire denunciation of Treaties in their essence.”

With Prussia’s resolve to annex Alsace-Lorraine, Italy’s forceful takeover of Rome, and Russia’s intention to shred the Black Sea clauses, it seemed Prussia’s victory at Sedan had rendered fragile the whole system of international law. But Gladstone was determined not to let it be torn asunder. The prime minister dispatched Odo Russell, England’s former representative at the Vatican, to the German headquarters in France to discuss the Black Sea issue.* Meeting with Bismarck, Russell exceeded his instructions and proclaimed that Britain would go to war, “with or without allies,” if the Russians refused to disavow their unilateral renunciation. Bismarck had no time to contemplate a potential conflict in the east while war against France still raged; he thus consented to call a conference of the powers, which was held in London in early 1871. At the end of negotiations the Black Sea clauses were revoked, but the parties declared it an “essential principle of the law of nations that no Power can liberate itself from the engagements of a treaty, nor modify stipulations thereof, unless with the consent of the contracting Powers by means of an amicable agreement.”

Gladstone had triumphed and the sanctity of international treaties was saved. But in Gladstone’s other great cause—applying the principle of national self-determination in Alsace-Lorraine—success was beyond his reach. In the beginning of January 1871 the German forces prepared to bombard Paris. By this point the British public had swung decisively in favor of France, and the Germans were taking notice. “That we Germans are losing hold of sympathy in England I have observed for a long time with grief,” wrote the Crown

---

* It was clear from the outset that Bismarck had agreed to support Russia’s claim in exchange for the czar’s good will toward Prussian aggrandizement. Eyck Erich, *Bismarck and the German Empire* (New York: W.W. Norton, 1964), p. 190; Langer, *European Alliances*, p. 11.
Prince Frederick to Queen Victoria, “and this will go on increasingly till Paris falls.”

Even Morier had doubts. “Germany is to be held accursed. 1. Because she has not yet made peace. 2. Because she insists on a territorial concession. 3. Because her manner of conducting the war is unworthy of a civilised nation,” wrote Morier, describing the current feeling of the British public. And Morier, perhaps in spite of himself, found reasons to agree with these sentiments. On one occasion his disappointment reached such levels that he fumed to his close friend, Ernst von Stockmar, that “the lust for gloire, kindled as it is within her [Germany], will burn with a much more terrible fierceness than it ever did in the grande nation.”

Stockmar was shocked by his friend’s words and asked for an explanation.

A few weeks later, after the German Empire was declared at Versailles, Morier responded in a more equable frame of mind:

What I meant to imply was that such unparalleled successes as those which have attended the German arms, and the consequent absolute power which the German nation has acquired over Europe, will tend especially to modify the German national character, and that not only necessarily for the better … Is it love of exaggeration to fear that under such circumstances the German Empire based on universal suffrage, i.e., on the suffrages of the 800,000 men who have been fighting in France … may have some of the faults of militarism attaching to it?

Morier had advocated for the form of German unity for decades, but at the moment of creation he was worried about the substance of the Germany that had come into existence. Would a united Germany, drunk on the triumphs of blood and iron, bear the emblem of Bismarck, its illiberal architect? Could Morier really feel confident that Bismarck’s influence would be fleeting, and that after his inevitable downfall Germany would blossom into a paragon of liberalism?

Morier was not alone in his concerns. The British public was rife with fears that England would be Germany’s next victim. George Tomkyns Chesney’s novella, The Battle of Dorking, published in serialized form beginning in early 1871, did much to fuel
dread. Quickly an international sensation, *The Battle of Dorking* is told from the perspective of an old Englishman, who narrates the story of his past to his grandchildren. “The danger did not come on us unawares,” the old man reflects, “... We English have only ourselves to blame for the humiliation which has been brought on the land.”55 He then proceeds to recount how a few years after the Franco-Prussian War, Germany conquered England and destroyed the British Empire. Similarly popular was Henry Pullen’s *The Fight at Dame Europa’s School*, which tells the story of three young students—Johnnie, Louis, and William—who symbolize England, France, and Germany, respectively. Though William was “the biggest and strongest” and coveted Louis’ flowerbeds, he was a “very studious and peaceable boy, and made the rest of the school believe he had never provoked a quarrel in his life.”56 When the rest of the school least expects it, William tramples over Louis’ flowerbeds and annexes them. Dame Europa soon learns of the story and excoriates Johnnie for his spineless neutrality, advising him to “Take care that William, the peaceable, unaggressive Boy, does not contrive (as I fully believe he will contrive) to get a footing on the river, where he can keep a boat, and then one fine morning take your pretty island by surprise.”57 By the end of February 1871, *The Fight at Dame Europa’s School* had sold two hundred thousand copies.58

Among British intellectuals and statesmen, concerns were concentrated on what the birth of Bismarckian Germany meant for the European state system, rather than expectations of future Anglo-German conflict. Far from the old idea that a united Germany would foster continental peace, Gladstone worried that German unification might signal “the beginning of a new series of European complications.”59 The jurist-historian Frederic Harrison, who had initially supported Germany against Napoleon, wrote in *The Fortnightly Review* that Europe faced “essentially a moral struggle; one of principles.”60 Despite hopes for a liberal Germany, “Bismarckism,” as Harrison called it, would provide an ideological challenge to British ideas of international morality. “Europe is thrown into the cauldron to be re-cast, and a new Holy Alliance is forming on the principle of ‘Blood and Iron’ which England must meet abso-
lutely alone,” Harrison cautioned. Prussia was the most militaristic state in the world, and the German people had so thoroughly surrendered their will to Bismarck that there was now more civic participation in czarist Russia. As Harrison asked, “We are told to trust to Germany at the close of her victory assuming a liberal form. What are the grounds for such hope?”

Harrison found such hopes groundless, but other educated Englishmen remained guardedly optimistic. Certainly when Bismarck presented his peace terms to France in February 1871—which included the annexation of Alsace-Lorraine—the British Foreign Office had no thoughts of counterbalancing nascent German power. Britain needed a larger body of evidence before it could conclude that Germany must be bridled.

THE PERIOD OF DISILLUSIONMENT, 1871–1874

Bismarck’s boundless imagination was both his blessing and his curse. On the one hand, it provided him with broad freedom to seek diplomatic combinations, and enabled him to prepare for contingencies that other statesmen could not foresee. But on the other hand, it filled him with constant anxiety; he was keenly aware of the fragility of his position, of how a rapid twist of fate could give rise to his cauchemar des coalitions. In his memoirs Bismarck reveals that fears of potential neutral-power intervention in the Franco–Prussian War “troubled me night and day,” and doubtless his diplomatic counterparts were unsettled by his excitability. When Odo Russell visited German headquarters in France in late 1870, Bismarck received the British emissary warmly, assuring him that “After the present war Germany would care for nothing but peace.”

* The French signed a preliminary peace on February 26, and the formal Treaty of Frankfurt was signed on May 10. The terms included the annexation of Alsace and a portion of Lorraine to the German Empire; a French indemnity of five-milliard francs; and German occupation of parts of France until the reparations were paid in full. Taylor, Struggle, p. 217; A.A.W. Ramsay, Idealism and Foreign Policy: A Study of the Relations of Great Britain with Germany and France, 1860–1878 (London: John Murray, 1925), p. 347; Paul Knaplund, Gladstone’s Foreign Policy (Hamden, CT: Archon Books, 1970), p. 62.

† Bismarck also told Russell that “England and Austria are the natural Allies of Germany,” Odo Russell to Granville (December 18, 1870). Paul Knaplund, ed., Letters from the Berlin Embassy,
Despite the Prussian minister’s good graces, Russell found reasons to be suspicious. “Now that I know him,” reported Russell, “I shall no longer be surprised to see him change the Map of Europe far more than the Emperor Napoleon was expected to do and we must be prepared for many disagreeable surprises.”

In the wake of the war, Bismarck’s anxieties were focused on two closely related issues: one external, the other internal. First and foremost, the new German chancellor feared that the French and the Austrians, who had both succumbed to the Prussian sword, would recover their strength and join forces against Germany in a war of revenge. Second, Bismarck worried about the unity of the fledgling German Empire. A number of constituencies within the empire—notably the Bavarians, the Alsatians, the Lorrainers, and the Poles—were thought to be of questionable loyalty, and owing to their Catholic faith, they represented a potential fifth column in the event of a Franco-Austrian war against Germany. As a result, Bismarck initially pursued a two-fold policy: the prevention of a clericalist party, most likely in the form of a restored monarchy, from taking power in France; and conciliation with the Vatican, in an effort to win Pope Pius IX’s support and moderate the separatist tendencies within the Catholic Center Party, which had formed in Germany in December 1870.

The second prong of Bismarck’s policy failed immediately. When the chancellor reached out his hand, the Vatican not only spurned his overtures, but the papal secretary of state, Cardinal Antonelli, proclaimed the Roman Curia’s solidarity with the Center Party. Bismarck’s imagination was unleashed. He sensed a Franco-clericalist conspiracy against German unity. The Franco-Prussian War itself, he reckoned, had sprouted from an agreement between the Vatican and France, whose Empress Eugénie was under Jesuitical control. Was it mere coincidence that France declared war on Prussia the day after the Vatican decreed papal infallibility? Bismarck’s anxious tendencies could not help but catch the attention of the British. “It is one of the peculiarities of that

great man that he is farsighted to a fault,” averred The Times. “There is something nervous and fidgety in his otherwise dauntless disposition, which renders him unable to find rest in the present, and sets him scheming and striving about an indefinite future.”71 With such an unpredictable helmsman, the German ship of state might sail a disruptive course in international affairs.

Following Antonelli’s rebuff, Bismarck decided to tamp down the Catholic Center Party by using more direct methods. When the Reichstag first convened in March 1871, the National Liberal Party possessed a commanding position.72 As the most enthusiastic supporters of German unity, the National Liberals viewed Bismarck as a hero and willingly followed his lead. A few months after the inaugural meeting, Bismarck commenced what would become known as the *Kulturkampf* by dissolving the Catholic Department of the Prussian Ministry of Worship. Meanwhile, under the leadership of Adolphe Thiers, the French were already showing signs of rejuvenation. By cleverly forging a handshake agreement in the National Assembly—the so-called Pact of Bordeaux—Thiers was able to revitalize French military and economic strength with impressive celerity. Alarmed by France’s recovery, Bismarck moved to further defend the German Empire from the potential fifth column.* While the full extent of Bismarck’s intended war against the Catholic Church was not yet known, the British Foreign Office already noticed signs of renewed Franco-German discord. “The New Year will open gloomily for France. The Germans appear to be alarmed, or at all events irritated, by Thiers’s military boasts and military preparations,” wrote Lord Lyons, the British ambassador to France, in late 1871. “The Germans mean to have their money and keep the territory they have taken, and they say that they had better have it out with France now that she is weak, than wait till she has got strong again.”73

In January 1872, Bismarck selected Adalbert Falk as the new minister of public worship,\(^7^4\) anointing him the juridical mastermind of the *Kulturkampf*.\(^7^4\) Falk’s first action at his post was to draft the School Inspection Law. Inspiring fierce debate throughout Germany and excited interest in Britain, the law secularized German schools, removing Catholic *and* Protestant influence from public education, and thus angering members of the Center Party as well as the Junker Conservatives. British liberals, who tended to view Church control over schools as regressive, received the law with great enthusiasm and were pleased to see Bismarck break with the Conservatives in favor of the National Liberals. British liberals began to hope that Bismarck was shunning his reactionary ways, but British conservatives—still far to the left of the German chancellor—remained skeptical.\(^†\) Odo Russell was well informed on Church matters,\(^7^5\) and he followed Bismarck’s anti-clerical campaign with acute interest. “[A]t this moment they can think and talk of nothing but the coming war with the Church,” reported Russell from Berlin. “Old Thile confided to me that he feared Bismarck was overdoing it and exciting popular animosity against the Clergy through the press to such a pitch that he would find it difficult to manage the anticlerical masses he was now arming and urging on.”\(^7^6\) But at that point, in late winter 1872, the infant *Kulturkampf* seemed very enlightened indeed to the British political classes. Conservatives like Disraeli\(^‡\) and Liberals like Gladstone\(^§\) could agree that the Vatican, which had recently propagated the *Syllabus of Errors* and decreed papal infallibility, represented an impediment to progress.

As Bismarck’s anti-clerical campaign gained momentum,

\(^*\) Bismarck gave Falk straightforward instructions: “To re-establish the rights of the State in relation to the Church, and *with as little fuss as possible.*” Quoted in Eyck, *Bismarck*, pp. 206–07.

\(^†\) The *Pall Mall Gazette* (London)—one of the more conservative papers in Britain at the time—surmised that the Liberals were getting ahead of themselves in their hopes. See “Occasional Notes,” *The Pall Mall Gazette* (London), February 29, 1872.


the French resurgence persisted at a fast pace, and Thiers boasted of the renewed strength of the French army. Although such incendiary language deeply upset German opinion and certainly spread fears throughout Europe of another Franco-German conflict, Russell was convinced that Bismarck’s next war would be an unconventional one. “The Generals tell the Emperor it would be better to fight France before she is ready than after,” wrote Russell, “but Bismarck, who scorns the Generals, advises the Emperor to fight France morally through Rome and the Catholic alliances against United Germany.”

This was exactly Bismarck’s intention. On May 14, 1872, the chancellor declared war against the Catholic Church in the Reichstag.

The Kulturkampf, as the historian Erich Eyck explains, “was looked on by a great part of Europe as one of the most exciting events of the age … there can be no doubt that in those years many of the most enlightened and highly educated men believed that the future of mankind was at stake.”

The veracity of Eyck’s statement is confirmed by the spectacular pronouncements of the British press. To illustrate, The Standard saw in Germany and the Vatican the respective embodiments of “Positivism and Ideality, the conflicting doctrines which have, for years past … been gradually drawing the populations of Europe into two hostile camps; and we are thus unable to divest ourselves of the belief that on the issue of the struggle now imminent hangs the future mastery over Europe of the principles involved in one or the other of these doctrines.”

It was the age of Charles Darwin’s On the Origin of the Species, and the Kulturkampf seemed the physical manifestation of the battle of science and progress versus superstition. In this conflict, the British political classes firmly supported science and progress; they were therefore ecstatic to see Bismarck, the proud Junker, split with the Conservatives, side with the National Liberals, and pioneer a crusade against clerical obscurantism. Nowhere outside of Germany was the Kulturkampf viewed more favorably than in Protestant England. In March 1872, The Times judged Bismarck’s

* Bismarck famously declared: “Have no fear—to Canossa we shall not go, either in body or in spirit.” Quoted in Arlinghaus, “Kulturkampf,” p. 348.
efforts against the Catholic Church a “splendid success” and criticized France for lagging behind Germany on this enlightened path: “[France] ought at last to feel how little it becomes her to be outdone by Germany in a just and liberal course.”

Brimming with confidence, Bismarck unveiled his next major initiative in June 1872: a law that dissolved the Society of Jesuits in Germany and gave the government the right to banish any Jesuit from the empire. This policy was clearly shaded in a darker hue than those hitherto enacted, and Eduard Lasker—leader of the National Liberal Party’s left wing—protested the law, proclaiming that it violated his liberal conscience. The British public seized upon the Jesuit law even more feverishly than it had the School Inspection Law, but this time responses were mostly negative. When one British newspaper commended the Jesuit law, The Standard lashed back fiercely. “Surely no more fitting motto for the Liberals of the present day could be found than these words of Tacitus, written nearly two thousand years ago—*Omnia liberaliter pro dominione*,” harangued The Standard:

> From the representative German Liberal, a petty tyrant in heavy chains, one can hardly expect very broad views on the subject of personal freedom; but when a London paper ... openly justifies religious persecution, we may well ask ourselves if all the progress, all the civilisation of which we boast, is not a dream from which we are gradually awakening.

The *Lancaster Gazette* echoed such sentiments, calling the Jesuit law “an antiquated policy” and criticizing Bismarck for imagining a Jesuit-led conspiracy of Austria, France, and the Vatican against Germany. The British Catholic Union organized a mass meeting, presided over by the Duke of Norfolk, to protest the German legislation. The Earl of Denbigh’s resolution condemning the law as “a wrong done to all natural right” passed comfortably. Denbigh had expected the result: “The meeting ... was bound to carry this motion, not only as Catholics, but as Englishmen and as the champions of liberty and of rights of the subject.” It was well for the British to find clerical obscurantism distasteful, but they must utterly object to policies granting the state such broad capabilities to
oppress civilians.

While the anti-Jesuit law tempered British hopes for an enlightened renaissance in Germany—and called into question the *liberalism* of the National Liberal Party—the Foreign Office was more concerned about Germany’s diplomatic endeavors of September 1872. For in that month Emperor Wilhelm of Germany, Czar Alexander II of Russia, and Emperor Francis Joseph of Austria convened in Berlin for a meeting of indeterminate purpose.* The British government could not help but observe the imperial gathering with suspicion. On the most basic level it evoked distasteful memories of the Holy Alliance, whose reactionary politics and hard-line methods had been viewed with disgust in England. More immediately, a restored alliance of the three emperors, anchored by an all-powerful Germany, would secure Russia’s western flank and free the czar to pursue a more vigorous policy in Central Asia—the most important theater of Anglo-Russian competition. Rumors abounded in Britain that great decisions were being made at the meeting; there were speculations about the rebirth of the Holy Alliance, and Conservative-leaning newspapers blamed Gladstone and the Liberals for letting British influence diminish on the continent.† The British government was so disturbed that it sent an imperial squadron to greet Thiers at Le Havre on September 18.‡ Odo Russell had the opportunity to meet with the foreign ministers of each of the parties at the conference, and they all took pains to assure him that no major political issues had been discussed. What was notable, however, was the statement of the Russian chancellor, Gorchakov, that he “had no interest in the weakness of France; on the contrary, he wished for a strong and

* The British public was also concerned. In the lead-up to the meeting, *The Morning Post* (London) speculated that a new Holy Alliance might be proposed, and that an agreement over the Eastern Question might be reached. See “France,” *The Morning Post* (London), August 9, 1872, pg. 5. *The Pall Mall Gazette* wondered if the meeting might spell doom for France. See *The Pall Mall Gazette* (London), August 5, 1872.

† “[W]e confess ourselves shocked at this clear evidence of the tendency in Europe to a new Holy Alliance…and we shall regret as we have never done before the wane of British influence on the Continent.” “Occasional Notes,” *The Pall Mall Gazette* (London), October 12, 1872.

‡ The gesture surprised the French president, but he gladly accepted it. Langer, *European Alliances*, p. 23.
prudent France,” and Russell’s own shrewd impression that Czar Alexander’s participation had been “unexpected and self-invited.”

The mysterious meeting of the three emperors further subdued British hopes that Bismarckian Germany was taking a liberal turn. To make matters worse, the expansion of the *Kulturkampf* was apparently backfiring. Traditionally the German Catholics had been the most liberal and independent-minded of Europe. Most German bishops had opposed the dogma of papal infallibility at the Vatican Council, and the British press could praise the German Catholics for being less rigid than “the stolid devotees of Italy or Spain.” But the sinister spirit of the *Kulturkampf*, revealed most blatantly in the anti-Jesuit law, pushed them into solidarity with the Vatican. Odo Russell, who normally trusted the chancellor’s wisdom, was disappointed. “I fancy that Bismarck utterly misunderstands and underrates the power of the Church. Thinking himself far more infallible than the Pope he cannot tolerate two Infallibles in Europe,” declared Russell:

> The German Bishops who were politically powerless in Germany and theologically in opposition to the Pope in Rome—have now become powerful political party Leaders in Germany and enthusiastic defenders of the new infallible Faith in Rome, united, disciplined and thirsting for martyrdom, thanks to Bismarck’s uncalled for and illiberal declaration of War on the freedom they had hitherto so peacefully enjoyed.

Yet while the British were questioning the prudence of Germany’s anti-clericalist measures, Bismarck was preparing to intensify them.

The *Kulturkampf* reached its highest stage with the infamous Falk Laws, first introduced in November 1872 and signed into law in May 1873. The British press criticized the legislation from the first. Morier, then British chargé d’affaires at Munich, knew that his

---

*Owing to the liberalism of both Catholicism and Protestantism in Germany, *The Lancaster Gazette* voiced surprise that such a vicious religious conflict could take place in the country. See “The Jesuits,” *The Lancaster Gazette, and General Advertiser for Lancashire, Westmorland, Yorkshire*, etc. (Lancaster, England), July 20, 1872, pg. 2.

† By transferring the education, training, and administration of the German Catholic clergy into the hands of the state, the Falk Laws effectively granted the German government limitless intervention in all aspects of Catholic life. German Catholics, of course, could never accept such measures; the bishops staged protests, urged passive resistance, and formally declared their resolve to disobey the Falk Laws. Arlinghaus, “Kulturkampf,” p. 350.
liberal-minded countrymen would resent the legislation, regardless of their opinions on the Vatican. “[T]his programme must appear monstrous to English eyes as an infringement of the plainest principles of individual liberty,” wrote Morier in January 1873.87 Odo Russell’s frustrations multiplied, and he decried the German government’s “ignorance.”88 Meanwhile, with the signing of a new Anglo–French commercial treaty, and Germany moving closer to Russia in consequence, British worries about Bismarck’s intentions swelled."

In March 1873, Odo Russell deduced that the chancellor had two prime objectives: “The supremacy of Germany in Europe and of the German race in the world,” and “The neutralization of the influence and power of the Latin race in France and elsewhere.”89 Bismarck’s resolve in this mission, Russell concluded, was unbreakable: “To obtain these objects he will go to any lengths while he lives, so that we must be prepared for surprises in the future.”90 And perhaps most disturbing, Europe’s day of reckoning might not be far distant. “The Germans … look upon the war of revenge as unavoidable and are making immense preparations for it,” Russell continued. “Germany is in reality a great camp ready to break up for any war at a week’s notice with a million of men.”91 A year after the British political classes loudly cheered Bismarck and the School Inspection Law, the Foreign Office deemed the chancellor illiberal, untrustworthy, and possibly even dangerous.

Despite fears of increasing Anglo-French amity, the German government was remarkably out of touch with the apprehension its policies had elicited in England. Emperor Wilhelm even thanked the British government for being “an ally in the war he was unhappily obliged to carry on against the Roman Catholic Bishops.”92 Odo Russell, who had received this message at Wilhelm’s seventy-sixth birthday party, felt obligated to rectify the emperor’s misconception: “I explained the difference of the standpoint of Church and state in England and Germany to His Majesty and said that

---

87“We are out of favour with the Germans for preferring the old French alliance to a new German one, as our commercial policy is said to prove,” wrote Odo Russell on March 14, 1873. Odo Russell to Lyons (March 14, 1873). Thomas W. Newton, Lord Lyons: A Record of British Diplomacy, Vol. II (New YOrk: Longmans, Green, and Co., 1913), pp. 41–42.
the love of freedom and toleration were so great with us now that we were not likely to imitate the policy so popular with the liberals of Germany at the present moment of placing the clergy and all denominations under military discipline.”

Bismarck, however, recognized that the Kulturkampf was faltering, and when Marshal MacMahon, an ultramontane monarchist, replaced Thiers as French president on May 24, he was hurled into a tailspin. Amplifying the chancellor's alarm was the fact that, miraculously, France was merely months away from having repaid the full five-milliard war indemnity, which would necessitate German evacuation of French soil. With the onset of summer, Bismarck acutely felt the fragility of his position and was determined to repair it.

Lacking faith in Gladstonian England, the chancellor turned once again to Prussia's erstwhile conservative partners, Russia and Austria. He realized ready success. Before the end of June, Germany had consummated a military partnership with Russia and entered a loose political alliance with Austria and Russia, bringing the Three Emperors' League into existence. As Bismarck labored to shore up Germany's position vis-à-vis France, the chancellor's worst fears of a Franco-clericalist conspiracy seemed to be coming to fruition. Later that summer, the rival royal factions of France agreed to a restoration of the monarchy under the Bourbon Count de Chambord, whose ultramontane sympathies were well known even in England. MacMahon was complicit in the arrangement, and Pope Pius IX himself had helped negotiate the compromise. But much to Bismarck's relief, Chambord nixed the plan at the last moment by refusing to accept the French tricolore over the Bourbon white flag. Around the same time, the Bishop of Nancy issued a letter to his clergy instructing them to pray for the restitution of Alsace-Lorraine to France. These blatant demonstrations of Catholic influence in French politics haunted Bismarck; and when

---


† The Bishop of Nancy's diocese still included part of German Lorraine. The affair caught the attention of the British press. The Graphic (London) reported that a German tribunal at Saverne had sentenced the Bishop of Nancy in absentia to three months in prison. “Foreign—France,” The Graphic (London), May 2, 1874.
France’s final indemnity payment forced Germany to evacuate French territory in September, the chancellor was spurred into action. Bismarck wrote a letter to the French foreign minister insisting that his government denounce the Bishop of Nancy’s instructions and assure that similar episodes would not happen in the future. Despite half-hearted conciliatory gestures on the part of France, the Bishop of Nancy remained steadfast, and the German press churned out belligerent pieces against France and the Catholic Church.

By fall 1873, even so consistent a Germanophile as Queen Victoria began suspecting that the means Bismarck had used to effect unification had poisoned Germany in its nursery. The Crown Prince Frederick himself had concerns, though he did not want his mother-in-law to feel overly discouraged. “The Entwickelung of Germany has not taken place in the way I fondly hoped it would, and there are many measures which I cannot admire or approve of,” wrote the Crown Prince to Queen Victoria, “but I firmly believe that what has been done, has been done for the good of Germany, and of Europe.” In the Reichstag elections of January 1874, the Center Party made substantial gains, becoming the second largest party in parliament. Not one to concede defeat, Bismarck amplified his rhetoric. In a conversation with Gontaut-Biron, French ambassador to Germany, the chancellor assumed a fearsome tone. “In the conflict in which I’m engaged with the Catholic Church,” said Bismarck, “I am energetically determined to conquer … Take care lest the [French] masses become fanaticized in the name of the persecuted Catholic religion, for then the clerical party will seize control and will espouse all the quarrels of the Roman Curia, and you will inevitably be drawn into war against us.”

Bismarck’s threatening language was taken seriously in France, and the new French foreign minister, the Duc de Decazes, wasted no time communicating his government’s distress to Britain.*

* Lord Lyons’ report on the situation vividly illustrates France’s gloom and Britain’s concern: “The fall of France has never, I think, been brought so forcibly home to me, as when I listened yesterday to the humble deprecation which Decazes was obliged to make with regard to Bismarck’s threats, in the same room in which I had so often heard the high language with which the Imperial Minister used to speak of the affairs of Europe. One can only hope that
Doubtless disturbed by Bismarck’s foreboding gestures, the British Foreign Office solicited the help of Queen Victoria, who agreed to appeal to Emperor Wilhelm for Franco-German peace.¹⁰³ “Notwithstanding an active and restless Catholic minority, the English nation, as a whole, is essentially Protestant,” the Queen wrote to Wilhelm, “and its sympathies would be entirely with Germany in any difference with France, unless there was an appearance of a disposition on the part of Germany to avail herself of her greatly superior force to crush and annihilate a beaten foe, and thus to engender the belief that a strong and united Germany was not, after all, the expected mainstay of European peace.” She concluded with an exhortation: “Being sensible that the fate of Europe rests in your hands, after such unparalleled successes, I venture to express my hope that you have the power—and no doubt—also the will to be magnanimous.”¹⁰⁴

Odo Russell found Bismarck’s maneuvers so confounding that he was convinced the chancellor was engaging in deception. “[N]othing can save them [the French] if Bismarck is determined to fight them again; but then, is it France or is it Austria he is preparing to annihilate?” pondered Russell. Believing that Bismarck’s ultimate object was “to mediatize the minor States of Germany and to annex the German Provinces of Austria,” Russell deduced that the Kulturkampf was a tool of the chancellor’s strategy: “His anti-Roman policy will serve him to pick a quarrel with any Power he pleases by declaring that he has discovered an anti-German conspiracy among the clergy of the country he wishes to fight.” This unconventional analysis probably sprouted from the absurdity that Russell saw in Bismarck’s perceptible policy, which held that France was liable to being attacked lest “[she] gag her press, imprison her bishops, quarrel with Rome, refrain from making an army or from seeking alliances with other Powers all out of deference to Germany.”¹⁰⁵

---

¹⁰³ The quoted text is from Queen Victoria to Emperor Wilhelm I, dated January 14, 1874. The text is a primary source and is included in the Historical Texts Project of the German Empire, available through the Digital Research Library of Germany.

¹⁰⁴ Odo Russell’s words are from his letter to Lord Granville, dated January 22, 1874. The text is a primary source and is included in the Newton, Lord Lyons, vol. 2, p. 50.
In the opening months of 1874, Bismarck’s bewildering policies left Britain disconcerted. But the British government could not agree on what Bismarck’s aims actually were, and those officials who had theories could hardly feel confident in their correctness. One thing alone seemed clear: the question of European peace rested in the hands of the German chancellor. As Odo Russell mused darkly: “[T]he only course Governments can follow is to let him do as he pleases and submit to the consequences until he dies.”

THE “WAR-IN-SIGHT” CRISIS AND THE RETURN OF THE BALANCE OF POWER

Bismarck never liked Gladstone. Considering the British prime minister sanctimonious and timid, Bismarck disparagingly referred to him as “Professor Gladstone,” and allegedly said that during the Grand Old Man’s ministry, he had “lost five years of his political life by the foolish belief that England was still a great power.” It thus seemed propitious that in the current state of tension, Disraeli, who cared deeply for foreign affairs, replaced Gladstone as prime minister in February 1874. The very phrase ‘foreign affairs’ makes an Englishman convinced that I am about to treat of subjects with which he has no concern,” said Disraeli on the campaign trail, “Unhappily the relations of England with the rest of the world, which are ‘foreign affairs,’ are the matters which most influence the lot.” For Disraeli, foreign policy was the crux of the matter—the foundation on which all domestic issues depended. He affirmed that the perceived diminution of British power was but an illusion produced by the ineffectual Gladstone Ministry: “I express here my confident conviction that there never was a moment in our history when the power of England was so great and her resources so vast and inexhaustible.” Disraeli was committed to restoring British prestige on the continent, and naturally he first looked to Berlin.

* It was only the second time since 1832 that the Conservative Party won a general election. Jonathan Parry argues (pp. 276–332) that public perceptions of the Gladstone Ministry’s weakness, born during the Franco–Prussian War, played a significant part in bringing about the electoral results. See: Jonathan Parry, The Politics of Patriotism: English Liberalism, National Identity and Europe, 1830-1886 (Cambridge: Cambridge University Press, 2006), pp. 276-77.
Shortly after assuming office, Disraeli’s foreign secretary, Lord Derby, reached out to the German ambassador, Count Münster, and voiced high hopes for Anglo-German partnership.\footnote{On February 28, 1874, Disraeli told Münster that he “had never believed even in Napoleon III’s lifetime that France was, or ever could be, a sincere ally of England. The only people who could go hand in hand, as must ever appear more plainly, were Germany and England.” Quoted in Winifred Tafts, “The War Scare of 1875 (I),” The Slavonic and East European Review, vol. 9, no. 26 (December 1930), p. 341.}

Bismarck’s policy, however, manifested no discernible changes. In May 1874 the National Liberals extended the \textit{Kulturkampf} even further.\footnote{The Reichstag passed a law that permitted the German government to arrest, banish, or deprive citizenship from any priest who had been removed under the 1873 Falk Laws. Shortly thereafter, all Prussian sees were vacant, and 1,400 parishes were without vicars. Arlinghaus, “Kulturkampf,” p. 366; Gordon A. Craig, Germany, 1866-1945 (New York: Oxford University Press, 1978), p. 63.} On July 13, 1874, Bismarck’s campaign against the Vatican nearly cost him his life when a Catholic copper’s apprentice, Eduard Kullman, fired a bullet into his hand. Unsurprisingly, Bismarck used the incident to support his accusation that the Center Party was an enemy of the German Empire, saying to a Center Party deputy in the Reichstag: “You may thrust Kullman aside, but he nevertheless belongs to you.”\footnote{These ironclads were much too heavy to use for just the protection of commercial vessels. Langer, \textit{European Alliances}, p. 42.} The British press widely condemned Bismarck’s cynical use of the incident. “In this speech of Prince Bismarck’s,” declared the \textit{Pall Mall Gazette}, “there is a degree of recklessness, an ostentatious refusal or an inability to be bound by any of the ordinary rules of human prudence.”\footnote{The Reichstag passed a law that permitted the German government to arrest, banish, or deprive citizenship from any priest who had been removed under the 1873 Falk Laws. Shortly thereafter, all Prussian sees were vacant, and 1,400 parishes were without vicars. Arlinghaus, “Kulturkampf,” p. 366; Gordon A. Craig, Germany, 1866-1945 (New York: Oxford University Press, 1978), p. 63.} Just when there seemed to be an opening for Anglo-German partnership, Bismarck alienated the British with his fitful outbursts.

As Bismarck’s policy seemed as aggressive as ever, the French vigorously lobbied the Disraeli Ministry to see the danger Germany posed to European security. Gavard, the French chargé d’affaires in London, implored Derby to consider the growing power of the German navy, which had added three ironclads since 1871.\footnote{These ironclads were much too heavy to use for just the protection of commercial vessels. Langer, \textit{European Alliances}, p. 42.} While this information caught Derby’s attention, it was not enough to convince the British Foreign Office to adopt the French perspective.

In December 1874, a Belgian boilermaker named Duchesne
sent a letter to the Archbishop of Paris offering to assassinate Bismarck for sixty thousand francs. Obtained and published by the German press, the letter caused a stir in both Berlin and London, and Bismarck responded by dispatching a note to Belgium urging the revision of Belgian law to prevent such violent threats in the future.* At about the same time Bismarck sent Joseph von Radowitz, chief of Germany’s department for Eastern affairs, on a special mission to St. Petersburg.† The assignment of such a high-ranking German diplomat to the Russian capital for no perceptible reason elicited gloomy speculations throughout Europe; and one of the most popular was that Radowitz had been sent to offer the czar support for his eastern ambitions in exchange for a free hand against France.‡ Less than a year after Disraeli entered office, Britain’s estrangement from Bismarckian Germany was nearly complete. The chancellor now worried the British not only with his words, but also with his actions; Bismarck’s note to Belgium and Radowitz’s mysterious mission to St. Petersburg seemed to threaten Britain’s vital interests in Belgian neutrality and the preservation of the Ottoman Empire. Meanwhile, recognizing Britain’s misgivings and fearing Bismarck’s ominous designs on his own country, Decazes resolved to rally London against German aggression.†

Franco-German tensions evolved into a full-blown crisis on March 13, when the French passed the Cadre Law, adding a fourth battalion to each regiment of the French army.‡ Both the German military establishment and the press were thrown into

---

* The British public followed the Duchesne affair closely. Belgium, of course, remained the central zone of British interests on the continent. See “Plot Against Bismarck’s Life,” The Star (Saint Peter Fort, England), December 29, 1874, and “Prince Bismarck,” The Times, December 23, 1874, pg. 5, col. A. Fearful rumors about the contents of Bismarck’s note to Belgium also circulated the British press. See “Germany and Belgium,” The Times, April 22, 1875, pg. 5, col. G.

† In his memoirs, Bismarck states that he sent Radowitz to St. Petersburg because Gorchakov, the Russian chancellor, refused to allow his diplomats to communicate with Bismarck in Berlin, and instead managed Russo-German relations solely through the German diplomatic corps in St. Petersburg. Bismarck, Bismarck, vol. 2, p. 190.

‡ Decazes and the French government found the German government’s embargo on the exportation of horses (March 2, 1875) particularly troubling. Apparently, the German government enacted the embargo after receiving reports that the France government had placed orders for 10,000 horses from Germany. E.T.S. Dugdale, ed., German Diplomatic Documents, 1871-1914, Vol. 1: Bismarck’s Relations with England, 1871-1890 (London: Methuen & Co., 1928), p. 2.
frenzy. Moltke calculated that the law would add 144,000 men to the French forces, and the newspaper *Nationalzeitung* described it as an “ad hoc” measure that could only have an immediate, aggressive purpose.\(^{115}\) Although German scaremongering strengthened Decazes’ case, the British government was not convinced that a Franco-German war was imminent. “I do not know and cannot conjecture the cause of Decazes’ anxiety,” wrote Derby, “Nothing has passed or is passing in any part of Europe to justify alarm as to an early disturbance of the peace.”\(^{116}\) At this stage of the crisis, Britain was only concerned about Bismarck’s intentions vis-à-vis Belgium and the Eastern question; the idea that Germany might invade France was considered too outlandish to take seriously.\(^{117}\)

Then, in early April 1875, a series of incendiary German newspaper articles finally prickled British nerves. Reaching its climax with the infamous Berlin *Post* article of April 8—provocatively titled “Is war in sight?”—the press campaign was universally believed to be Bismarck’s doing.\(^{118}\) That the German chancellor was commissioning articles affirming the imminence of war could not but stir Britain and indeed all of Europe. With the exception of Odo Russell, even the foreign diplomats in Berlin were convinced that war was looming.\(^{119}\) And despite Russell’s assurances that “Bismarck is at his old tricks again” and “this crisis will blow over like so many others,” Derby, unaccustomed to the chancellor’s methods, seems to have lost faith in Bismarck’s intentions.\(^{120}\) “The disturbing leading article in the *Post* had become known here that morning,” reported Count Münster to Bismarck from London, “and I found the Minister [Derby], usually so calm, in a somewhat excited state of mind. He said that it almost seemed as though the fears in Paris of an attack by Germany were not altogether unfounded.”\(^{121}\)

After dinner at the British embassy in Berlin on April 21, Radowitz—back from his mission to Russia—engaged in spirited conversation with the French ambassador, Gontaut-Biron. Though at present, Radowitz said, Germans were confident that war was not imminent, there was certainly cause for anxiety.\(^{122}\) Had Radowitz ended the discussion then, nothing would have come of it, but he proceeded. “[I]f revenge is the inmost thought of France—
and it cannot be otherwise—,” said Radowitz, “why wait to attack her until she has recovered her strength and contracted her alliances? You must agree that from a political, from a philosophical, even from a Christian point of view, these deductions are well grounded and these preoccupations are fitted to guide the policy of Germany.”123 What Radowitz was espousing was the theory of preventive war, which had been discussed in military circles and received with trepidation in France, but never thought to be the accepted doctrine of the German government. Yet now a high-level German diplomat had mentioned it, and Gontaut believed it was uttered as official policy. He quickly transcribed the conversation and sent it to Decazes.

Now possessing what he considered definitive evidence that Germany was eyeing war, Decazes commissioned M. de Blowitz, Paris correspondent for The Times, to “expose the entire situation.”124 He also forwarded Gontaut’s report to all the other European capitals.125 Perhaps it was upon learning of Radowitz’s statements that Odo Russell also began doubting Bismarck’s intentions. “The prospect of another war fills me with horror and disgust, and if Bismarck lives a few years longer I do not see how it can be prevented,” wrote Russell in late April.126 His concerns were further heightened on May 1 when Baron Nothomb, the Belgian ambassador to Germany, informed him of recent conversations with Moltke and Bismarck. Moltke had said that war was inevitable within one year unless France revoked the Cadre Law, and Bismarck had spoken even more disturbingly: “Tell your King to get his army ready for defence, because Belgium may be invaded by France sooner than we expect.” As for Nothomb, his personal fear was that the German army might occupy Belgium.127 Russell had been slow to believe that Germany might actually be planning for war, but by May 1875 the preponderance of evidence was impossible to ignore.

With apprehensions mounting, the British government moved to weigh options for curbing Germany’s schemes. The first inclination was to hope that another European power would intervene. “Is there no hope of Russian interference to maintain peace?” asked Derby, “It cannot be the interest of Russia to have France
destroyed and Germany omnipotent.” Interestingly, it seems to have been Queen Victoria who first seriously suggested taking direct action to ensure peace. On May 5 she told Disraeli that “every means should be used to prevent such a monstrous iniquity as a war,” and proposed that Britain “ought, in concert with the other Powers, to hold the strongest language to both Powers, declaring they must not fight, for that Europe would not stand another war!” The next day Odo Russell happily reported that Count Shuvalov, Russian ambassador to Britain, had spent the night in Berlin and had informed him that the czar planned to visit the German capital on May 10, upon which he would “insist on the maintenance of peace in Europe, even at the cost of rupture with Germany.” Disraeli recognized an opportunity to reassert Britain’s influence in European affairs and expunge the Gladstone Ministry’s legacy of diffidence. He communicated his thoughts to Derby on May 6: “My own impression is that we [should] construct some concerted movement to preserve the peace of Europe, like Pam did when he baffled France and expelled the Egyptians from Syria.” That same day Blowitz’s article appeared in The Times under the title “A French Scare.” The stage for intervention was set.

On May 8, Derby telegraphed Odo Russell instructing him to act alongside Alexander II for the preservation of peace in Berlin. Meanwhile Count Shuvalov returned to London, where he promised British diplomats that Russian expansion into Central Asia would cease. While this guarantee did not materially affect the course of events, it was remarkable in and of itself: the present threat of Germany was considered so severe that Russia and Britain were willing to shelve their Great Game competition in Asia to check it. May 10 was the day of decision. Odo Russell and Bismarck ate dinner together, and the British ambassador politely delivered his government’s démarche. Bismarck, who was very fond of Russell, responded gracefully and even expressed appreciation for Britain’s concern for peace. But toward the end of the evening, Gorchakov, the Russian chancellor, barged into the dining room and expressed his own government’s policy with distinctly less tact. To make matters worse, he revealed that Britain and Russia had
coordinated the intercession. Bismarck was infuriated by Gorchakov, whose pretensions of playing the peacemaker he called a “circus performance.” Nevertheless, peace was secured, and Bismarck’s animosity fell upon Gorchakov alone.

The British government was pleased with its performance. Disraeli fancied that the intervention was the most vigorous act of British foreign policy since Palmerston’s time, and he congratulated Derby for his work: “Your policy seems to be very popular, and very successful ... We must not be afraid of saying ‘Bo to a goose.’” Derby agreed that the policy was wise, not least because it “involved no risk and cost no trouble, while it has given us the appearance of having helped, more than we really did, to bring about the result.” Meanwhile in Germany, the Crown Prince and his wife—who had both been so hopeful for a great liberal partnership between Britain and Germany—were left searching for answers. The Crown Prince speculated that tensions might have been overblown by Bismarck’s indiscrete utterances, and the Crown Princess complained that the chancellor “remains the sole and omnipotent ruler of our destinies.” Queen Victoria, whose letters to her daughter were typically so gentle, proudly wrote the Crown Princess taking credit for Britain’s policy, and providing a harsh analysis of Anglo-German relations: “No one wishes more, as you know, than I do for England and Germany to go well together; but Bismarck is so overbearing, violent, grasping and unprincipled that no one can stand it, and all agreed that he was becoming like the first Napoleon whom Europe had to join in PUTTING down.”

The British had always harbored misgivings about Bismarck. Except for a short period in the early stages of the Kulturkampf, most Englishmen could probably agree with the German Crown Princess’ estimation that Bismarck was “mediaeval” and that “the true theories of liberty and of modern government are Hebrew to him.” Yet despite Bismarck’s obvious power, the British had largely refused to let the chancellor color their hopes for an ultimately liberal and peace-loving Germany. As an “enemy of progress,” Bismarck and his influence would only be temporary. He would eventually be swept away by the tide of history, and the Ger-
man state would then finally reflect the liberal wishes of its people. By May 1875, however, Disraeli and his colleagues had recognized this reasoning as illusory. Four years to the day from the signing of the Treaty of Frankfurt, Britain joined with czarist Russia to counter an increasingly illiberal Germany. Far from the vision of Morier’s dreams, “in which free and united Germany, in alliance with England, should impose peace on Europe, and inaugurate a golden age of international security,” by 1875 Bismarck resembled Napoleon in the eyes of British statesmen, and Germany seemed more likely to disturb the peace than impose it. The balance of power had been restored, but not in the way the British had expected four years prior.

FORM AND SUBSTANCE: BRITAIN, BISMARCK, AND THE IDEA OF GERMAN UNIFICATION

Not for nothing is history associated with the figure of Nemesis, which defeats man by fulfilling his wishes in a different form or by answering his prayers too completely. – Henry Kissinger

“By defeating Austria and France while unifying Germany under Prussia, Bismarck could not help but profoundly challenge the European balance of power,” writes David Calleo in The German Problem Reconsidered. “A coalition of hostile powers was nearly inevitable, and indeed appeared as early as 1875, when Britain and Russia both made clear they would not tolerate another German victory over France.” The Anglo-Russian intercession of 1875 was no small event in the diplomatic history of Europe; after a long period of aloofness from European affairs, Britain set aside its Central Asian rivalry to counterbalance German power on the continent. But to describe the Anglo-Russian action as the work of a “coalition of hostile powers” whose emergence was “nearly inevitable” after German unification is to overlook the utterances of British statesmen made all the way up to 1875.

Far from viewing German unity as an eventuality to be prevented, the British political classes tended to believe that a unified Germany would promote European peace and stability. The British
ambassador to Prussia summarized this outlook during the Aus-
tro-Prussian War in 1866:

I could not view with any dissatisfaction or fear of danger to
England an increase of power to Prussia. She was the great Prot-
estant state of continental Europe. She represented the intelli-
gence, the progress and wealth of Germany. We have … nothing
to fear from her. She will become a Power of great importance
in maintaining the peace of Central Europe. She will gradually
advance in a constitutional system of government, and she will
play the part of a moderator in Europe. We have much in com-
mon with her—our race, our religion, our mutual interests are
all interwoven with Prussia, and our political interests should be
identical.¹⁴⁶

British support for German unification rested on two pillars: the
geopolitical premise that a robust Germany at the center of Eu-
rope would restrain France and Russia from challenging Britain’s
extra-continental dominance, and the sociocultural premise that a
unified Germany must be liberal and, in turn, Britain’s “natural ally.”
British statesmen wanted not only the form of united Germany but
also a certain substance; their mistake was to assume that the for-
mer would inexorably produce the latter.

Hence it is unsurprising that England received news of Prus-
sia’s victory at Sedan with loud cheers. France, which had caused
tremendous upheaval on the continent for over two centuries, had
fallen to Prussia, which was a symbol of restraint and shared im-
portant religious, cultural, and ideological ties with Britain. More-
over, Queen Victoria’s daughter was married to the Crown Prince
of Prussia, who socialized in liberal circles and wrote passionately
of “a free German Imperial State, that in the true sense of the word
should march at the forefront of civilisation and be in a position to
develop and bring to bear all noble ideals of the modern world.”¹⁴⁷
But Prussia’s behavior in the aftermath of Sedan, including the
bombardment of Paris and especially the annexation of Alsace-
Lorraine, chafed at British ideas of morality and justice, which were
particularly important to Prime Minister Gladstone.

Most educated Englishmen of the era—who were all liberal
by continental standards—were steeped in the “Whig interpretation” of history. For these men, history was teleological, unfolding ever-higher levels of progress and more perfect forms of liberal government. All people were essentially the same; the only difference between nations was their relative positions on the same path to enlightenment. There was no room in this paradigm for great men. Individuals could either swim with the tide of history or be washed away; they certainly could not alter its flow. In his quest for German unification, Bismarck swam with the tide, and thus should be supported; in his reactionary and aggressive policies, Bismarck swam against it, and thus could be ignored. Germany’s troublesome behavior between the Battle of Sedan and the signing of the Treaty of Frankfurt could be written off to Bismarck, whose influence would be ephemeral. By 1871, therefore, Britain’s highest hopes were shaken, but its ultimate faith in united Germany remained.

British doubts, however, were amplified in the years 1871–1874. In its early stages, the Kulturkampf enjoyed popularity in England. Secularizing German public schools was considered the epitome of progress, and Bismarck’s breaking with the Junker Conservatives to combat clerical obscurantism alongside the National Liberals seemed cause for celebration. The confidence of British statesmen and intellectuals was momentarily restored. But British enthusiasm for the Kulturkampf faded when it lost all semblance of proportion; indeed it was impossible for the British liberal mind to support policies that enabled the state to imprison Jesuits and priests at will. Compounded with Bismarck’s revival of the triple alliance of Germany, Russia, and Austria, the increasingly repressive nature of the Kulturkampf called into question the very liberalism of the National Liberal Party, and the English were forced to re-evaluate their conviction that a unified Germany would be Britain’s ideological ally.

When the Kulturkampf entered the realm of foreign policy, and German officials began making threats to both France and Belgium, the fears of British statesmen came fully alive. Perhaps Germany would not only be illiberal at home, but also a menace to continental stability and possibly even Britain’s vital interests.
This was the worst of both worlds—both geopolitical and sociocultural premises were failing; Germany’s disappointing substance seemed to reverse all expected benefits of the form of unification. In 1874 Prime Minister Disraeli entered office preaching the return of Britain as a force in international politics, and in May 1875 he got his chance. By the time of the Anglo-Russian intercession at Berlin on May 10, Bismarck’s tactless press campaign and the belligerent statements of a variety of German officials had convinced virtually the entire British Foreign Office that Germany must be restrained. Even Odo Russell, whom Bismarck was sure harbored no suspicions of Germany’s intentions vis-à-vis France, applauded the decision to intervene. *

Thomas Carlyle, one of Britain’s most ardent Germanophiles, rejected the “Whig interpretation” of history. “[A]ll things that we see standing accomplished in the world,” wrote Carlyle, “are properly the outer material result, the practical realisation and embodiment, of thoughts that dwelt in the Great Men sent into the world.” 148 By 1875, it was clear that the mere form of German unity would not produce the substance so desired by British intellectuals; Bismarck was one of Carlyle’s “Great Men,” and the Germany born in 1871 was unmistakably his offspring. Perhaps the most ironic element of the situation is that as far back as 1854 British statesmen had identified Bismarck with ideas antithetical to English liberalism, and some had even speculated that his future rise to power could only signify that reactionary forces had triumphed in Prussia. 149 Yet these same statesmen believed nationalism and liberalism were inextricably tethered, and thus they stomached Bismarck’s aggressive means in the belief that not even three wars could thwart Germany from liberal development. But, as it happened, in uniting Germany with “blood and iron,” Bismarck successfully severed nationalism

---

* Bismarck was shocked by the British intercession. As he wrote to Münster a few days afterward: “This phenomenon to me is all the more astonishing since Lord Odo Russell has always reported from here in a contrary sense. I consider him much too good and truthful an observer, considering how long he has had before him evidence of our friendly policy, to have written any differently.” Bismarck to Münster (May 14, 1875). Dugdale, German Diplomatic Documents, vol. 1, p. 9; Odo Russell wrote to Derby five days after the intervention: “I was delighted at the course pursued by Her Majesty’s Government and at the instructions you sent to me, which I feel sure will do good, both at home and abroad.” Odo Russell to Derby (May 15, 1875). Newton, Lord Lyons, vol. 2, p. 80.
from liberalism; and it was not until 1875 that the British finally acknowledged his protean influence.

As Franco-German tensions heightened in 1874, Robert Morier reflected wistfully on “the irony of fortune which grants to men what they earnestly pray for, but usually in such a way as to render the gift useless or worse than useless.” So eager to see Germany unified, and so confident of the inseparability of nationalism and liberalism, Morier had grudgingly supported Bismarck’s wars. But those wars had consequences beyond German unification: notably an enduring Franco-German antagonism, enshrined in the annexation of Alsace-Lorraine, and the vindication of military force as the critical instrument of German statecraft. British statesmen ought not have been surprised, therefore, when German officials decided that aggressive threats were the best means to make France accommodate their interests in 1875. The “war-in-sight” crisis marked the culmination of nineteenth-century Britain’s disillusionment with the idea of German unification, and though it certainly did not make a future Anglo-German conflict inevitable, it revealed for the first time the constellation of powers that would clash in the great wars of the twentieth century.

Notes

3 A number of works providing good narratives of European diplomatic history between 1870 and 1875 have been published. Anglo-German political relations in this period are covered particularly well in William L. Langer, European Alliances and Alignments, 1871-1890 (New York: Alfred A. Knopf, 1956); Raymond James Sontag, Germany and England: Background of Conflict, 1848-1894 (New York: W.W. Norton, 1969); and Paul Kennedy, The Rise of the Anglo-German Antagonism, 1860-1914 (London: George Allen & Unwin, 1982).
5 Quoted in Ibid.
Antagonism, p. 8.

8 Taylor, Struggle, p. 213.

9 Murray, Robert Morier, x, 98. A co-founder of the Cosmopolitan Club with Henry Reeve and A.H. Layard, and an original member of the Cobden Club, Morier was an active participant in English liberal intellectual circles. He was also a renowned expert on German affairs.


11 Mountstuart E. Grant Duff, Studies in European Politics (Edinburgh: Edmonston and Douglas, 1866), p. 234, p. 245. Grant Duff held firm to such views despite his harsh criticisms of Bismarck. As he wrote in 1866: “[I]t is not too much to say that his [Bismarck’s] action upon the affairs of Europe has hitherto been simply evil.”


13 Mosse, European Powers, p. 365.


15 Ramsay, Idealism and Foreign Policy, p. 277.


17 Hans Rothfels, “1848—One Hundred Years After,” The Journal of Modern History, Vol. 20, No. 4 (December 1948), pp. 291-319. Rothfels describes (pp. 300-301) German liberals’ intense and even aggressive antagonism toward Russia during the uprisings of 1848.

18 Ibid., pp. 181–83.

19 Gladstone to Queen Victoria (July 15, 1870). George E. Buckle, ed., The Letters of Queen Victoria: A Selection from Her Majesty’s Correspondence and Journals Between the Years 1862 and 1878, Vol. II: 1870-1878 (London: John Murray, 1926), p. 34.

20 The Crown Princess Frederick to Queen Victoria (July 18, 1871). Ibid., p. 43.

21 Queen Victoria to the Crown Princess Frederick (July 20, 1870). Ibid., p. 44.


28 Quoted in Langer, European Alliances, pp. 10–11.

29 Taylor, Struggle, p. 211.


37 Granville to Gladstone (October 7, 1870). Ramm, *Gladstone and Granville*, vol. 1, p. 139.

38 Gladstone to Granville (September 30, 1870). Ibid., p. 135.


40 Queen Augusta of Prussia to Queen Victoria (September 25, 1870). Buckle, *Letters of Queen Victoria*, vol. 2, p. 73.


42 Quoted in Schreuder, "Gladstone as 'Troublemaker,'" p. 123.

43 Quoted in Ibid., p. 124.

44 Granville to Gladstone (October 9, 1870). Ramm, *Gladstone and Granville*, vol. 1, p. 141.


46 Thomas Carlyle, "Mr. Carlyle on the War," Letter to the Editor, *The Times*, November 18, 1870, pg. 8, col. D.

47 "We Publish to-day a letter from Mr. Carlyle," Editors' Note, *The Times*, November 18, 1870, pg. 7, col. B.


53 Ernst von Stockmar to Robert Morier (December 17, 1870). Ibid.


57 Ibid., p. 22.


61 Ibid., p. 643.

62 Ibid., p. 641.


64 Ibid., p. 32.


67 Bismarck was even prepared to offer the pope asylum in Germany. See Bismarck, *Bismarck*, vol. 2, p. 135.


69 Langer, *European Alliances*, p. 36.

70 Papal infallibility was proclaimed in the constitution *Pastor Aeternus* on July 18, 1870; France declared war on Prussia on July 19 of that same year. Murray, *Robert Morier*, p. 213.

71 “London, Friday, September 6, 1872,” *The Times*, Editorial, September 6, 1872, pg. 7, col. A.

72 The National Liberals controlled 155 of the 399 seats. Craig, *Germany*, p. 64.


74 The appointment of Falk was deemed important enough in Britain to be announced in *The Times*. See “Germany,” *The Times*, January 31, 1872, pg. 5, col. A.

75 Odo Russell had spent twelve years as British representative at the Vatican before his assignment in Germany.


77.5 “Germany,” *The Standard* (London), June 11, 1872, pg. 5.


79 “Prince Bismarck,” *The Times*, March 16, 1872, pg. 9, col. D.

80 The Jesuit bill’s passage in the Reichstag elicited announcements in a variety of British newspapers. See “Closing of the German Parliament,” *The Morning Post* (London), June 20, 1872, pg. 5.

81 Eyck, *Bismarck*, p. 206. Whereas Lasker was one of the only National Liberal officials to denounce the Jesuit bill, the Catholic Center Party considered it an abomination that confirmed the National Liberal Party’s malevolence. August Reichensperger, a Center Party official, described the anti-Jesuit law as “a declaration of modern liberalism’s bankruptcy in the field of the spirit.” Quoted in Craig, *Germany*, p. 77.


84 “Sympathy with the Jesuits,” *The Times*, July 17, 1872, pg. 11, col. A.

85 Odo Russell to Granville (September 12, 1872). Winifred Taffs, ed.,


89 Ibid., p. 41.

90 Ibid.

91 Ibid.


95 The *Times* claimed that Chambord “cannot speak of the Pope without tears in his eyes.” See “Ultramontane Plots Against Prussia,” *The Times*, March 13, 1872, pg. 12, col. A.

96 The final French payment was made eighteen months ahead of schedule. Craig, p. 106.


100 Crown Princess Frederick to Queen Victoria (October 4, 1873). Ibid., p. 284.

101 “The Elections to the German Parliament,” *The Times*, January 17, 1874, pg. 12, col. A.


104 Queen Victoria to Emperor Wilhelm of Germany (February 10, 1874). Ibid., p. 314.


106 Ibid., p. 53.

107 Ramsay, *Idealism and Foreign Policy*, p. 357.


109 Disraeli’s speech at Free Trade Hall in Manchester (April 3, 1872), Buckle, *Life of Disraeli*, vol. 5, p. 191. It is important to note that the key British diplomatic personnel in Germany—including Odo Russell and Robert Morier—maintained their posts despite the change of government.

110 Ibid., p. 192.


112 Ibid.

113 Tafts, “War Scare of 1875 (I),” p. 337.


116 Derby to Lyons (March 16, 1875); Newton, *Lord Lyons*, vol. 2, p. 71.
117 Dr. Geffcken, a liberal German intellectual, informed Morier that “[Bismarck] is
resolved to annihilate Belgium, which he declares to be the central Government of the
political Catholicism, and the heart of coalitional conspiracies.” Dr. Geffcken to Morier (March
119 As Russell reported: “Half the Diplomatic Body have been here since yesterday
to tell me that war was imminent, and when I seek to calm their nerves and disprove their
anticipations, they think that I am thoroughly bamboozled by Bismarck.” Russell to Derby
120 Russell to Derby (April 10, 1875). Ibid.
121 Münster to Bismarck (April 13, 1875). Dugdale, *German Diplomatic Documents*,
vol. 1, p. 4.
123 Quoted in Fuller, “War-Scare of 1875,” p. 203.
124 The article was not published in *The Times* until May 6, 1875. See H.G.S.
126 Russell continued: “The Emperor’s powers of resistance are over; he does what
Bismarck wishes, and the Crown Prince, peace-loving as he is, has not sufficient independence
of character to resist Bismarck’s all-powerful mind and will.” Odo Russell to Lyons (April 24,
1875); Newton, *Lord Lyons*, vol. 2, p. 74.
127 If Germany was resolved to go to war against France, the German army might
occupy Belgium as a prelude. Odo Russell to Derby (May 1, 1875). Ibid.
128 Derby to Russell (May 3, 1875). Ibid., p. 75.
131 Odo Russell to Derby (May 6, 1875). Newton, *Lord Lyons*, vol. 2, p. 76.
132 “Pam” refers to Lord Palmerston. Disraeli to Derby (May 6, 1875). Buckle, *Life of
Disraeli*, vol. 5, p. 422.
133 “A French Scare,” From a French Correspondent [M. de Blowitz], *The New York
Times*, May 6, 1875, pg. 8, col. A.
134 Derby also informed Russell that the British government had proffered the same
suggestion to both the Austrian and Italian governments. Derby to Odo Russell (May 8, 1875).
137 Disraeli to Derby (May 18, 1875). Ibid., p. 423.
138 Derby to Disraeli (May 20, 1875). Ibid., p. 424.
139 The Crown Prince wrote a letter to Queen Victoria two weeks after the affair: “I
am afraid we shall never accurately find out, who is in reality to be blamed for getting up these
alarming rumours. But I could not help observing that perhaps our Chancellor himself, from
his state of nervous irritation, might sometimes be induced to utter remarks, which are liable to
be misconstrued and are often interpreted as meaning more than they were originally intended
to mean.” Crown Prince Frederick of Germany to Queen Victoria (May 24, 1875). Buckle,
140 The Crown Princess blamed Bismarck for ruining Germany’s good reputation. See
Crown Princess Frederick to Queen Victoria (June 5, 1875). Ponsonby, *Empress Frederick*, p.
138.
141 Queen Victoria to the Crown Princess Frederick (June 8, 1875). Buckle, *Letters of
Queen Victoria*, vol. 2, p. 405.
142 Crown Princess Frederick to Queen Victoria (June 5, 1875). Ponsonby, Empress Frederick, p. 139.
147 Crown Prince Frederick’s diary (October 24, 1870). Ponsonby, Empress Frederick, p. 104.
149 Kennedy, Rise of Anglo-German Antagonism, p. 9.

Bibliography

Primary Sources
Memoirs, Letters, Diplomatic Documents, and Biographies containing primary materials

Newton, Thomas W. Lord Lyons: A Record of British Diplomacy, Vol. II. New York: Longmans,
Green, and Co., 1913.


Novellas and Contemporary Political Studies


Parliamentary Debates


Newspaper and Magazine Articles

“A French Scare.” From a French Correspondent. The Times, May 6, 1875, pg. 8, col. A.

Carlyle, Thomas. “Mr. Carlyle on the War.” Letter to the Editor, The Times, November 18, 1870, pg. 8, col. D.

“Closing of the German Parliament.” The Morning Post (London), June 20, 1872, pg. 5.

“Foreign—France.” The Graphic (London), May 2, 1874.

“France.” The Morning Post (London), August 9, 1872, pg. 5.

“Germany and Belgium.” The Times, April 22, 1875, pg. 5, col. G.

“Germany.” The Standard (London), June 11, 1872, pg. 5.

“Germany.” The Standard (London), July 4, 1872, pg. 5.

“Germany.” The Times, January 14, 1873, pg. 12, col. C.

“Germany.” The Times, January 31, 1872, pg. 5, col. A.


“London, Friday, September 6, 1872.” The Times, Editorial, September 6, 1872, pg. 7, col. A.


“Occasional Notes,” The Pall Mall Gazette (London), August 5, 1872.
“Plot Against Bismarck’s Life.” The Star (Saint Peter Fort, England), December 29, 1874.
“Prince Bismarck and the Ultramontanes.” The Pall Mall Gazette (London), December 7, 1874.
“Prince Bismarck.” The Times, December 23, 1874, pg. 5, col. A.
“Prince Bismarck.” The Times, March 16, 1872, pg. 9, col. D.
“Sympathy with the Jesuits.” The Times, July 17, 1872, pg. 11, col. A.
“The Elections to the German Parliament.” The Times, January 17, 1874, pg. 12, col. A.
“Ultramontane Plots Against Prussia.” The Times, March 13, 1872, pg. 12, col. A.
“We Publish to-day a letter from Mr. Carlyle.” Editors’ Note, The Times, November 18, 1870, pg. 7, col. B.

Secondary Sources

Halperin, S. William. Italy and the Vatican at War: A Study of Their Relations from the Outbreak of the Franco-Prussian War to the Death of Pius IX. Chicago: University of Chicago Press, 1939.
F. “The White Revolutionary: Reflections on Bismarck.” Daedalus 97, 3 Philosophers
since the Cold War, debate in the West over the future of the international order has fixated upon the discourse of moral liberalism. Moral liberals envision a liberal democratic world order, focus on the individual, and invoke the need to secure universal human rights—derived variously from naturalistic, consensus, and functional theories, and construed generally to mean claims to human security (i.e., freedom from fear and want)—as justification for interventionist responses, and to claim moral high ground (i.e., “just war”) in the international arena. *

During the 1990s and 2000s, two positions within this discourse—the cosmopolitan and the hegemonic—bifurcated the vast majority of Western statesmen, policymakers, and informed citizens participating in public political discussion. While both cosmopolitan and hegemonic moral liberals intend to advance the type of humanitarian internationalism described above, they differ in the means they propose and the concretized institutional ends they seek. The schism staged a “choice” for denizens of the West to resolve.

The cosmopolitans aspire to formally constitutionalize international law to the point of forging a new kind of “world citizenship” capable of guaranteeing such a set of human rights. To the cosmopolitans, the demands of these rights must countervail the traditional demands of state sovereignty and formal international

---

* I use the term “moral liberalism” to distinguish between the specific understanding outlined above, and other uses of the term “liberal.” These include “IR liberalism,” which asserts an empirical relationship between international institutions and international political behavior, “classical market liberalism,” which extracts political principles from the ideal conditions for capital accumulation, “liberal constitutionalism,” which divides power and ascribes rights to guarantee the equal liberty and worth of individuals from their government, and “Liberalism,” which refers to a progressive social-democratic partisan ideology in modern Western nations.
law. For them, the legalistic discourse of sovereignty, and the inherently political nature of international organizations (i.e., that international organizations are ultimately derivative of states with multiple and regularly conflicting interests, values and aspirations) must not be permitted to interfere with humanitarian “rescue” operations in the face of vast and offensive violations of international law.¹ To some, forcing the actual violation of extant international laws (i.e., pursuing unauthorized unilateral “humanitarian” actions)—analogous to civil disobedience—is the only promising course for reform.² Regardless, under the cosmopolitan model, the political and legal concept of state sovereignty is reduced to a functional “responsibility to protect” that, should the state default on its obligation, could be appropriated to the global community. This would in turn coordinate the dissolution and subsequent reconstruction of the “failed” state.³ United Nations reform—meaning mostly measures that legally clarify and materially secure additional resources to “guarantee the effective implementation of resolutions of the Security Council”—is at the heart of this vision.⁴ Moreover, cosmopolitans argue that UN action has already borne and legitimized the concept of cosmopolitan global citizenship by recognizing individuals as immediate subjects of international law (i.e., with standing in international courts). To the extent that the “constitutional quality” of the UN Charter and other global compacts are still informal, they further contend it will be through future UN action that the formal constitutionalization of international law is eventually achieved.⁵

By contrast, hegemonic internationalism eschews international law and instead calls for the unilateral action of a benevolent hegemon to catalyze the proliferation of liberal democracy, and guarantee the provision of similarly defined human rights to security and protection. Like the cosmopolitans, liberal hegemons also insist that the discourses of state sovereignty and public international law have become irrelevant. However, they propose replacing the system of states with a project of benevolent imperialism intended to provide the basic physical security owed to individuals across the globe. From this perspective, global governance and “hu-
manitarian interventions” are merely vehicles for the empire’s rule.6 Writing in Foreign Affairs in 1990, Charles Krauthammer heralded the position bluntly:

Why it should matter to Americans that their actions get a Security Council nod from Deng Xiaoping and the butchers of Tiananmen Square is beyond me. But to many Americans it matters. It is largely for this reason that American political leaders must be sure to dress unilateral action in multilateral clothing. The danger, of course, is that they might come to believe their own pretense.7

Institutionally, these hegemonic liberals seek to maintain the international structure of formally independent states, but aspire to organize and align them imperially beneath the peace-securing hegemon. Under this model, the internal actions, ambitions, and claims of individual states (including claims to sovereign equality and non-intervention) would be de facto subject to the approval of the hegemon, whose judgment is supreme due not only to its material power, but also an implicit contract it believes to hold with the world’s nations. The world society that hegemonic liberals posit would still be integrated—modern technology and economics, if nothing else, requires that it be so—but only at the level of market interactions and imperial interventions, and not by anything amounting to political relations between sovereign equals, let alone actual world citizens.

On the other hand, constitutional cosmopolitans endeavor to design a highly integrated, law-governed and politically constituted world society. The distinction between internal and external actions of individual states would be blurred, and in any case shaped by the collective attitudes and ideas of a transnational political body. The essence of such a politically constituted world society is that it would be composed both of states and citizens. The flow of legitimation would thus take two paths converging on a single world organization: on the one hand, from individuals qua “national citizens” via a corresponding nation-state to the transnational negotiation system (i.e., diplomacy, partisan politics, civil society organization) that would be “responsible within the framework of
the international community for issues of global domestic politics”; on the other hand, from individuals qua “cosmopolitan citizens” via an international community composed of member states “responsible to their citizens, and to the peace and human rights policy of the world organization.”

That neither a hegemonic nor cosmopolitan order has been institutionalized is ultimately due to the indecisiveness of the world’s most influential actor. Unable to make a choice that would end the schism, the United States has been left caught in a muddled state of perpetual “choosing” for nearly twenty years. It is no secret that, since 1992, the United States has flirted with both major visions. In one respect, the United States has pursued cosmopolitan policies of selective engagement and cooperative security (e.g., UN-sanctioned activities in Somalia, the Balkans, and Iraq, as well as sanctions levied against Iran, North Korea, and others) that have acceded power to the UN. At the same time, the United States has also advanced a decidedly hegemonic agenda (e.g., Iraq).

Without the West’s most influential nation consistently advocating a path forward, international order has bent toward the fundamental moral liberalism of the two predominant alternatives. This foundation bears assumptions about politics and the human condition—notable for their effect upon discourses affecting human rights, sovereignty, and political life itself—that portend to embed themselves into any emerging global political formation. Before world order coalesces into such a structure, it is imperative that the assumptions underlying moral liberalism—and their consequences for the social world—are fully evaluated, and that alternatives are fully explored. Criticism of the first wave of moral liberalism provides solid ground for such an effort.

LIBERALISM AND ITS CRITICS 1919-1939

In the years that followed World War I, moralistic models of

political order attracted the energies of theorists and the aspirations of statesmen. The Great War, Woodrow Wilson announced in January 1918, “will be recalled as the culminating and final war for human liberty.” We now know that his idealism would lead to tragic disappointment: the explosion of a second total war—this one perhaps even more repugnant than the first—exposed critical oversights in the moral liberal platform that, through the structure of the League of Nations, seemed to have embraced dimensions of both the hegemonic and cosmopolitan varieties.

It was in this context that Hans Joachim Morgenthau (1904–1980) began writing about politics. Morgenthau traced the root of social and political problems in the United States during the interwar period back to the fundamentals of moral liberalism: an aversion to politics and a celebration of the private individual. To him, these were vestiges of the bourgeois classical liberalism forged in early modern Europe. Both classical liberalism and its vestiges, Morgenthau argued, could not be understood outside the context of the former’s origins among early modern Europe’s rising middle classes and their attempt to posit an alternative social and political order that challenged the authority and power of the feudal aristocracy.

The building block of the social theory that the bourgeois liberals constructed was a material and rational individual “self” that possessed tangible interests (above all physical security) and that acted in the calculated private pursuit of those interests. When those interests became stymied by the political power of the aristocracy, the bourgeois liberals, Morgenthau argued, came to identify political action with the aspiration for power, which they in turn associated with “a particular manifestation of a ‘lust’ for domination.” Consequently, classic liberals “identified opposition to aristocratic politics with hostility to any kind of politics.”

In its most basic form, classic liberalism was thus a type of anti-political “market liberalism.” It elaborated a utilitarian ethics that moralized private economic pursuits, as well as a democratic theory that reduced politics into the systematized aggregation of competing interests. Morgenthau contends that the construction divorced individual and social “lifeworlds” from the “political activi-
ties” that were considered part and parcel of aristocratic claims to authority and power. The process of this separation reduced understandings of social and political life into scientific questions of material fact and calculations of interest. Politics, as an autonomous sphere of human activity distinct from the pursuit of self-interest, was subsequently dismissed or devalued in later iterations of liberal thought.12

Inheriting this tradition, and rocked by the global carnage of 1914–1917, early twentieth-century moral liberals turned their attention to war as the latest political obstruction to private life and physical security. Politics, which had been reduced by their intellectual ancestors to aristocratic tyranny and an abominable “lust for power,” now became reduced to abhorrently violent interstate aggression. Through broad treaties, narrow pacts, and even attempts at world federation, moral liberals aspired to replace international politics with an international “harmony of interests.”

Ultimately, Morgenthau argued that a mistaken understanding of what political action constituted, and what the human “lust for power” represented, had led to the prompt return of total war when appeasement failed and potential allies, averse to fighting, refused to act early enough. For him, mid-century moral liberals failed to understand that political action was much more than just “power-politics,” and as such could not be expelled from the social world. I interpret Morgenthau to argue that, from the ontological position that it is man’s purpose to encounter his own being, politics constitutes the quintessence of humanity insofar as political action is the activity through which men disclose and assert themselves as equal, unique agents. That the “tragic” ramifications of political activity are not always as elegant as the construction itself, and that sometimes the process of self-determination is sluggish and chaotic, are therefore not justifications to abolish politics altogether. Rather, as Morgenthau seems to assert, we would do best to contain and navigate the tragedies of political action through the adoption of an ethic of responsibility.

Today, in defiance of these intuitions, contemporary moral liberals of both the cosmopolitan and hegemonic variety aspire
to take the project of depoliticization to its furthest extents yet. Through the distinct modes previously described, both point toward the external disaggregation of political communities and the quashing of their internal political activities, ostensibly for their own benefit. Moreover, the content of the universal rights doctrine they both seek to secure fails to assert mankind’s need for polity and therefore not only misses the point, but in fact excludes many (i.e., the stateless and/or politically unincorporated) from the enjoyment of these “universal” rights altogether. At the same time, the continued development of this doctrine as an acceptable justification to unilaterally or selectively reject traditional non-intervention arguments provides pretense for speciously benevolent hegemons or imperial coalitions to aggrandize beneath a banner of apparent justice. The menace of total war ultimately rises from the mass graves of the last go-around as the moral stakes for war are once again elevated to the level of absolute imperatives.

The great failure of moral liberalism is thus a failure to grasp the concept of the political. In the chapters that follow, I endeavor to present an alternative that does. To do so, I will reconstruct what I argue is a genuine reading of Morgenthau’s international political thought by critically reexamining two generations of interpretations. I argue that Morgenthau’s “political conception” of international order has been lost in these interpretations. I call his a “political” conception because the moral values upon which his theory is based are derived specifically from the ontological Aristotelian notion of man as a “political animal,” to which I have alluded previously. This usage should not be confused with either the Rawlsian or the power-political senses of the construction.*

I begin in Section I by retracing the character of Morgenthau’s ideas back to the writings and ideas of Nietzsche and Weber,

---

* In the Rawlsian sense, a “political conception” refers to a non-metaphysical “neutral” basis upon which a diverse set of peoples can find an “overlapping consensus” to justify a common concept of justice that can be used to gauge the legitimacy of basic shared political and social institutions. In the power-political sense, a “political” conception is one that considers the constant self-preservationary accumulation of power to be the principal feature of the social world. Neither denotes the particular meaning I assign. For examples of each, see: Rawls, “The Idea of an Overlapping Consensus,” *Oxford Journal of Legal Studies* 7 (1987): pp. 1–25; John J. Mearsheimer, *The Tragedy of Great Power Politics* (New York: Norton, 2001), pp. 4–8, pp. 39–47.
from whom they rightly derive, developing what I argue is Morgenthau’s actual conceptual foundation. In Section II, I then construct an international normative theory that extends from this foundation: a political conception of international order centered on the moral value of political action, the state, and the state-system.

I argue that the international system that emerges from this political conception is legally instituted upon the twin pillars of sovereign-equality and politically conceived human rights. This “dualistic” formation, which reconciles universal human rights with sovereign equality through a program of constitutional pluralism, contrasts with the cosmopolitan position that privileges human rights above sovereignty altogether. It is also at odds with the hegemonic position that combines the moral liberal concept of universal rights with a principle of hierarchical sovereign-inequality. A politically conceived international order thereby depicts a human political status analytically located between the global citizenship and imperial vassalage of the two liberalisms.

Through his vision for such an order, Morgenthau reminds us not to neglect the basic political character of humanity in crafting an international order. His ideas articulate a viable and compelling alternative that speaks to man’s political nature, the frequent tragedy of his moral aspirations, and ultimately a cautious optimism for the better angels of his judgment.

1. THE HUMAN CONDITION AND THE CONCEPT OF THE POLITICAL*

In what follows, I trace the origins of Morgenthau’s political thought to Nietzsche and explore how Nietzsche had influenced Morgenthau’s understanding of the political and the human condition in general. I then focus on how Weber’s ideas—specifically,

value pluralism and the ethic of responsibility—refined those conceptions and informed the political ethic that Morgenthau would come to embrace as necessary in a tragic and disenchanted modernity.

**The Tragic as a Condition of Human Existence**

Morgenthau’s sense that the human condition was tragic evolved from a belief in the unbridgeable chasm between human desire and capability, he inherited from Nietzsche. For Morgenthau, it was not our failure to refrain from desire, but rather the fact that certain desires unavoidably fall just out of reach—recalling what Nietzsche called “the irremediable, ineluctable, inescapable”—that describes the tragic condition of human existence. This is a dominant theme throughout Morgenthau’s writings. “Suspended between his spiritual destiny which he cannot fulfill, and his animal nature in which he cannot remain,” Morgenthau wrote in *Scientific Man*, “man is forever condemned to experience the contrast between the longings of his mind and his actual condition as his personal, eminently human tragedy.”

Like Nietzsche, Morgenthau worried that the modern West had become not only content with denying the tragic character of the human condition, but also too eager to pursue utopia without regard for the efficacy or consequences of those pursuits. It had become, as per Nietzsche, a definite kind of sickness: “The political and military crises of the thirties and forties … are but the outward manifestations of an intellectual, moral and political disease which has its roots in the basic philosophic assumptions of the age.” These assumptions produced an idyllic intellectual fallacy that ignored the permanence of tragic forces in the social world, particularly those that stemmed from political action. For Morgenthau, the permanence of these tragic forces derived from two sources.

First, human beings are simply unable to calculate and control the results of their actions. Once one acts, her action becomes an

---

*This subsection owes much to Frei’s translations of Morgenthau’s personal writings held at the Library of Congress.*
independent force that shapes, and is shaped by, unforeseen contingencies. “While our hand carries the good intent to what seems to be its consummation,” Morgenthau writes, “the fruit of evil grows from the seed of noble thought.” Since social man owns various moral interests that are intertwined beyond his capacity to understand, “while satisfying one, we must neglect others and the satisfaction of one may even imply the positive violation of another … Whatever choice we make, we must do evil while we try to do good, for we must abandon one moral end in favor of another.” Secondly, in what Morgenthau calls political action, tragedy arises not merely from inevitable, unforeseen, and unintended contingencies, but in fact, “the very essence of the intention and very life-blood of the action.” This has to do with the nature of political action as a “distinct sphere of human activity.”

The Political Drive and the Political

For Morgenthau, political action stems from the animus dominandi—a “political drive” that he seems to have derived from Nietzsche’s will to power. Morgenthau understood the will to power as the fundamental ontological force intrinsic in human nature that drives men to construct a transcendent identity. “It is of the same kind as the mystical desire for union with the universe,” Morgenthau wrote of the animus dominandi, “it attempts to push the individual beyond his natural limits” toward the “transcendent goal” of self-discovery and disclosure.

It is crucial that the animus dominandi can only be satisfied through intersubjective action: fullness can only be attained through the engagement of other beings. This is to say that it is a political drive. Since it originates in the spirit of man, the political is not merely a relationship between individuals, but it is also a quality within individuals that drives them together. To be human is to be homo politicus. Until man joins with other beings, “he cannot fulfill himself, he cannot become what he is destined to be.” As the human telos, political action constitutes the search and expression of meaning and identity.
It is important to note that, for Morgenthau, the political drive cannot achieve its ontological objectives through violence. The political drive, insofar as it quests for the self-assertion of an individual being’s essence in an intersubjective space, necessitates an element of mutual recognition. This is because the *animus dominandi* (as manifested through political action) seeks not obedience from other beings, which might be sufficient for self-preservation, but in fact a connection with other beings that affirms the political actor’s essence.22 Once relations enter the realm of physical force, they are no longer “political” activities, and thus lose their value.

The observable outcome of political action is inevitably a contest: agents in a shared space compete to have their disclosed identities, expressed as ideas, dominate one another to establish the basis for a collective union. This is to say that even the most inspiring bonds unavoidably become entangled in what Morgenthau calls “a struggle for power.”23 Here we see why tragedy is present in the motive force behind political action: political action—for which success is measured by “the degree to which one is able to maintain, increase or demonstrate one’s power over others”—is almost always incongruous with moral action—the test of which is “the degree to which one is capable of treating others not as means to the actor’s ends, but as ends themselves.”24 As the acquisition of fullness comes to entail the partial sacrifice of moral yearnings, and, conversely, as the pursuit of moral conviction comes to interfere with fundamental human aspirations, the human condition becomes ineluctably colored by tragedy.

The Ethic of Responsibility and the National Purpose*

While Nietzsche framed the human condition Morgenthau would come to accept as the basis for his thought, it was Weber who outlined the ethic that would be necessary to navigate its severity. Weber defined the ethic of responsibility in “Politics as a Vocation.” Speaking in 1918, Weber sought to describe the char-

---

*Morgenthau originally had used the term “national interest.” He later revised it in his 1960 *Purpose of American Politics* and began referring to the “national purpose.” I do the same.
acteristics of political leadership urgently needed in a German government that, he worried, had become stymied by a “dehumanized” and “detached” bureaucracy that had become rationalized beyond not only service to the public, but indeed service to any political cause whatsoever. The description similarly targeted the would-be reformers—mostly pacifists, socialists, and other leftist groups—whose strict adherence to ideology and unwillingness to compromise prevented them from effectively mobilizing in any real capacity sufficient to catalyze reforms. In contrast to the paralyzing and destructive ethics of ultimate ends that afflicted both the German bureaucracy and potential reformers, Weber promoted an ethic of responsibility:

For [a man who believes in the ethic of ultimate ends], if an action of good intent leads to bad results then, in the actor’s eyes, not he but the world, or the stupidity of other men … is responsible for the evil. However, a man who believes in an ethic of responsibility takes account of precisely the average deficiencies of people, and he does not feel in a position to burden others with the results of his own actions so far as he was able to foresee them; he will say: these results are ascribed to my actions.

Morgenthau, ever aware of those “ethical paradoxes,” argues that this ethos is the best we have to resolve the tragedy of political action:

What is done in the political sphere by its very nature concerns others who must suffer from unwise action. What is done with good intentions but unwisely and hence with disastrous results is morally defective for it violates the ethic of responsibility to which all action affecting others, and hence political action par excellence, is subject.

At the level of international politics, Morgenthau reformulates this as an ethic of the “national purpose.” According to it, a polity’s normative aspirations and interests are weighed against each other, as well as against universal values, to determine the degree to which each should be pursued according to a prudent evaluation of the consequences to which they are likely to lead. In so doing, an ethic is derived from political reality, and not the other way around.
An ethic of the national purpose acknowledges, as Weber did, “that very frequently, the ‘world images’ that have been created by ‘ideas’ have, like switchmen, determined the tracks along which action has been pushed by the dynamic of interest,” and that the greatest nations “have contributed to the affairs of men more than the successful defense and promotion of their national interests.” Thus, though it considers a community’s material interests in universal moral principles such as survival and prosperity, an ethic of the national purpose weighs these against the particular values that the community espouses, as well as certain universal values that transcend individual communities. * Morgenthau understood nations as complex normatively constituted political communities, functioning first and foremost as forums for self-construction and identity-formation in which individuals have voluntarily joined together on the basis of shared political and cultural beliefs. National purposes can therefore be as diverse as the various moral and political constitutions that a political community may have. † Statesmen subsequently face an array of considerations that must be balanced before reaching a conclusion that may jeopardize some at the expense of others. Thus, the ethical test is whether or not the responsible decision-making agent has prudently evaluated alternative courses of action and weighed potential consequences upon national purposes in reaching her decision.

As might be evident at this point, each political community will have its own distinct national ethic emanating from its own weighted set of values and interests. The variability of value inputs that might supply an ethic of national purpose with its navigational signposts reflects Morgenthau’s value-pluralistic conception of the political and the modern social world, which he also develops from Weber.

Disenchantment, Value Pluralism, and the Political Revisited

The diversity Morgenthau intuits derives itself from Weber’s

---

* For Morgenthau, the perpetuation of the conditions necessary for the “good life” to flourish constitute such a universal value. What the “good life” is, and what these conditions are, will be discussed in section three.

† The “basic universal principles” which accompany a political conception of international order and confer legitimacy upon states and their national purposes will be explicated below.
theory of modernity. For Weber, modernity in the West sprang from the rationalization of the world that made it calculable. The revelation of the underlying causal processes of observed phenomena manufactured a fact-value divide that “disenchanted” men of the magical mysteries of religion that had formerly sated his longing to find fullness in material reality, leaving a disenchanted world in which Christianity could no longer function as a transcendent value that bound together the various spheres of human activity. All the other spheres—the political, aesthetic, erotic, intellectual, etc.—were permanently released as separate, equal, and autonomous value spheres with corresponding central values as a result of disenchantment. To describe the situation, Weber draws the analogy of a “return to polytheism,” of gods (value spheres) interlocked in an “eternal struggle” battling each other for human souls. In this schema, individuals can only “bear the fate of the times” and “choose which is god for him, and which is the devil.” Morgenthau puts it similarly: “Nations meet under an empty sky from which the Gods have departed.” The operative word in Weber’s construction is “choose” because it illustrates the basic characteristic of the modernity that Morgenthau adopted: modern beings, and communities, self-select their values from a near infinite catalog.

Within this typology, Morgenthau renders the political sphere to be primary because it is the focal point of human activity. Its unique attributes—access is universal (i.e., humanity consists of political beings) and the content is limitless (i.e., the pursuit of power, in the ontological sense that Morgenthau applies, can take any form by which men seeks to “assert themselves against the world”)—present the potential for democratic flourishing and totalitarian barbarism. On the one hand, Morgenthau is optimistic that the universality and limitlessness of the political promote a high affinity for the development of democratic institutions and activities. “The doctrine of democracy,” he writes eagerly of the political’s indeterminate content, “starts with the assumption that all citizens are potentially capable of arriving at the right political decision and that, consequently, nobody has a monopoly of political wisdom.” To that end, he defends the political against the other
value-spheres whose natural tendencies are to infiltrate the political and “depoliticize” humanity by importing their absolute values (i.e., specific beliefs about religion, morality, and economics) into what is a principally open, indeterminate, and ideally democratic theater of contestation.

On the other hand, however, Morgenthau recognizes the potential maladies embedded into the universality and limitless-ness of the political sphere. He insists upon maintaining the robustness of the other spheres so that they can balance against the political to prevent it from consuming them and transforming human communities into “pure political communities” which is to say “total states.” To preserve a democratic political community against the dual specters of totalitarianism and depoliticization, Morgenthau aspired to contain and insulate the various value spheres via self-limitation and the division of power. This practice of “balancing” unique independent competitive units constitutes a basis for a broader theory of international order.

II. A POLITICAL CONCEPTION OF INTERNATIONAL ORDER

From these conceptual foundations, I argue for a concept of international order centered on the moral value of political action, the state and the state-system. I begin the section by focusing on a series of lectures Morgenthau gave on Aristotle’s Politics. These lectures help us to better integrate Morgenthau’s arguments and to draw further connections that illuminate and consolidate his ideas.

I then turn to an exercise in what Rawls has called “ideal theory.”36 That is, I will build a political conception of international order using ideal units (i.e., the “ideal state,”“the ideal state system”) that function in an ideal fashion (i.e., reasonable competition and collaboration). Ideal theory entails two assumptions. First, that all actors (citizen and societies) generally comply with the principles of a politically conceived international order based on the moral value of political action we have identified. Ideal theory thus ideal-izes away the possibility of law-breaking (i.e., aggressive war). Sec-
ond, ideal theory assumes reasonably favorable social and economic factual conditions enabling nations to accept the principles in the first assumption. Starting with ideal theory provides us with the optimal example which, through "non-ideal theory," we will reference in order to reform our imperfect reality. Once we understand the ideal principles of international order, we will better see, for example, how the international community should act toward oppressive tyrannies and belligerent states that challenge the peace.

The Political Life and the Moral Value of Political Action

The political concept of international order I argue for begins with the moral value of political action. As we have seen, Morgenthau believed that political action constitutes a self-disclosing ontological expression: men are driven to political action by the very nature of their being and the exercise of such action fulfills their telos as homo politicus. In 1970, Morgenthau gave a series of lectures* at the New School on Aristotle's Politics that elaborated on these beliefs. In them, he also began to express what he considered to be the moral value of the state. "Man cannot achieve his telos outside the state," he said. "The state is essential for the individual's ability to achieve his purpose in life. Without the state, he could not do it."37 Furthering and replicating Aristotle's teleology, Morgenthau argues that because the political purpose of man—i.e., the "highest good"—depends upon the state, "the telos of the state is not just to ensure the bare survival of the citizens."38 Rather, "it is not life as such that the state must preserve, but the good life."39

Morgenthau's time at the New School overlapped with that of Hannah Arendt, a friend and romantic partner who was herself significantly influenced by Aristotle's political thought. Like Morgenthau, Arendt based her political theory upon an analysis of the human condition. Her schema divides this condition into three fundamental activities: Labor, Work, and Action. The first denotes

activities that are necessary to survive, while the second refers to pursuits that seek, or otherwise involve, “artificial” inventions of the human experience. But it is with the third, “Action,” that she is most interested.

In the human condition—which she describes as a “paradoxical plurality of unique beings” sharing the same spatial universe—Action constitutes the self-disclosure of those unique beings through speech, possible only in the presence of others who, “see and hear, and are hence capable of establishing the reality of subjective expression.” As Arendt explains, “In acting and speaking, men show who they are, reveal actively their unique personal identities, and thus their make their appearance in the human world … The disclosure of ‘who’ somebody is, in contradistinction to ‘what’ somebody is … is implicit in everything somebody says or does.”

The point here is that Arendt’s “Action” and Morgenthau’s “political action” are analogous concepts that seem to be derived from, or at least partially influenced by, a common reading of Aristotle’s ontology. Most importantly, Arendt, like Morgenthau, advances Aristotle’s position that political action is man’s purpose, the expression of his freedom, and that leading a political life—what Arendt called the vita activa—constitutes man’s ultimate aim: the political good life of which Morgenthau spoke. “A life without speech and without action,” she wrote, “is literally dead to the world; it ceases to be a human life because it is no longer lived among men.”

But such a life was just what both theorists found in their evaluation of the “affluent society” that emerged in Western nations after World War II. For them, the same kind of moral liberalism that emerged in the 1920s and 1930s was once again eroding the fullness and imperiling the security of human civilization. The culprits were decadent consumerism and the canonization of private pursuits. As the public and the political became penetrated by the private and the economic, Arendt lamented that, “fabrication has come to occupy a rank formerly held by political action.” “The very concept of happiness which in the Declaration of Independence and Federalist refers to public happiness,” Morgenthau wrote
of America in 1960, “now takes on an exclusively private and predominantly material connotation.” Freedom, finally, was becoming sublimated into unworthy chimeras like the unencumbered accumulation of personal property and beliefs about individual inviolability. It took on an apolitical meaning, connoting both anarchy and a chance for unhampered acquisition as well as signifying freedom not only from being ruled, but also from ruling others—the freedom from politics altogether.

For both Morgenthau and Arendt, the interwar period and World War II proved that politics and the political life could neither be denied to individuals, nor ignored by statesmen in their calculations, without terrible consequences. The lesson was not learned by many of their contemporaries. In turning their backs on what they were (i.e., political beings), modern men in the postwar period unknowingly debased themselves by inhibiting the fullest expressions of who they were. To someone that was at one time denied political personhood and subsequently threatened with extermination, the idea that men would voluntarily abdicate their essential humanity was shocking. Reviving, sustaining, and institutionalizing the political life became the guiding principle behind Morgenthau’s political philosophy and the basis for this political conception of international order.

III. IDEAL THEORY: THE MORAL VALUE OF THE STATE

The state is at the center of Morgenthau’s political conception of international order and its moral value derives from its essential relationship to living the political “good life.” As we proceed, this derivation will shape the kind of state that he valorizes in particular.

In his conceptual vocabulary, the state is the compulsorily organized form of society in modernity. The polis, city-state, and kingdom had all, at one time, functioned as just such compulsory organizations. Drawing from Weber, Morgenthau posits that some level of social organization is always compulsory because domestic peace can only be preserved when a monopoly of violence is established and organized. To Morgenthau, the modern state is the
broadest form of social organization that can realistically secure the resources necessary to provide physical security for its citizens without becoming so broad (i.e., becoming a world state) that it could no longer host a political debate framed by the particular moral and ideological convictions shared by a distinct political community.\textsuperscript{51}

But security, as Morgenthau noted is not the source of the state's moral value.\textsuperscript{52} Rather, by enabling the individual to experience political community as a continuum in time and space, as a living idea in whose name men act, bestow, and draw benefits, the state facilitates freedom and the political good life.\textsuperscript{53} As Arendt explains, the state's raison d'être is to establish and keep in existence “a space where freedom as virtuosity can appear, where freedom is a worldly reality, tangible in words which can be heard, and deeds which can be seen, and in events which are talked about, remembered, and turned into stories before they are finally incorporated into the great storybook of human history.”\textsuperscript{54}

The state's moral value in the context of the political life is therefore three-fold: (1) it provides and secures a public space for agent-disclosing political action; (2) it accumulates, organizes and distributes resources in a manner that enables individuals to access and participate in the public space; and (3) it offers an object toward which political action in that space can be directed. It is the studio, palette, and canvas for \textit{homo politicus}. In the first instance, the state extends institutions to essentially prohibit violence from poisoning the public space of appearances. These include police and military forces, but also communications regulations and positive laws that maintain a civil discourse. The sum of these structures secures what Jean Cohen has called “the internal conditions of possibility for self-determination and self-government under law—i.e., for political freedom.”\textsuperscript{55}

Because the freedom of the public life further requires the “conquest of necessity,” the state in the second instance serves to liberate citizens from the chains of subsistence—i.e., what Arendt referred to as Labor and Work.\textsuperscript{56} In modern times, states have developed whole departments devoted to organizing and securing
the private economic lives of its citizens. Central banks, treasuries, and regulation boards all exist in part for the ostensible political purpose of ensuring that citizens have ample resources to be sufficiently available for political participation. Many states have gone further to develop elaborate agencies and transfer programs to provide a base level of subsistence to the destitute. Some advanced democracies have even instituted public access broadcasting, or enacted fair-content requirements, to increase access to the public political forum to those with limited time and resources.

In the third instance, the state functions as a three-dimensional object for political action. As a political community’s real, shared entity existing in both space and time, the state constitutes the immediate substance to be shaped and molded by collective action, and simultaneously the physical record onto which words and deeds can be chronicled and celebrated by posterity. Further, and perhaps most decisively, the state as a personality emits an aura of patriotism that dramatizes and illuminates political action with what Arendt called the “shining brightness we once called glory.” In each of these capacities, the state constitutes an essential component of realizing the political good life Morgenthau identifies.

It is important to note that a democratic constitution is not necessarily entailed by these criteria to establish moral value. Rather, a state needs only to foster political inclusiveness and extend the experience of political community—i.e., the ability to participate in collective self-determining activities and struggles—to all citizens qua political equals in order to possess moral value, membership in the international community, and, as will be discussed, a legitimate claim to sovereignty.

But within the context of ideal theory, it is the democratic state—one founded upon a republican model that reflects of the basic political equality and aspirations of humans qua speech-beings—that Morgenthau identifies as an ideal-type.*

* Critical republican political theorist Cecile Laborde describes this state as one in which, “all citizens enjoy basic but robust civic standing, in the form of political voice, basic personal autonomy, equal opportunities, material capabilities, and intersubjective mutual recognition as equal citizens.” Cecile Laborde, Critical Republicanism (New York: Oxford University Press, 2008), p. 11.
further elaborates on what he means by the democratic state in various discussions of the United States and its founding. Two features endear the United States to Morgenthau as an ideal form. First, the United States is premised on a decoupling of “nation” from “state.” While the former derives from immutable un-willed characteristics of what someone is or appears to be (meaning ethnic, racial, territorial, or otherwise inherited traits of happenstance), the state springs from the willed expression of who someone actually is. “The United States,” he writes, “was founded upon loyalty not to a monarch or a piece of territory, but to a shared purpose,” a continued commitment to a shared constellation of values and beliefs substantiates a connection to the polity. Second, the United States is ideally democratic insofar as it “assumes that all members of society have equal access to the truth, but none of them have a monopoly upon it.” Dissent is emblematic of a democratic model that survives because of its value-pluralistic character and essential neutrality.

Morgenthau contrasts this type of epistemic democracy from both the dictatorial and Jacobin alternatives. In the former, a political elite claims to possess a monopoly on truth (and therefore the “will of the demos”), and so justifies concentrating among themselves additional monopolies of power and law. The elite has then “not only the right, but even the duty to suppress dissent,” for dissent quite literally becomes “tantamount to heresy and treason.” In the Jacobin form, “the will of the majority is the ultimate source of truth in matters political,” and dissent represents a dismissible error in judgment. The suppression of dissent is thus reinterpreted and justified as necessary civic reeducation. For Morgenthau, both alternatives stifle political self-disclosure in an attempt to ho-

---


† “Dissent is legitimate...because of the relativistic ethos that all types of democracy share.” Ibid., pp. 41. See also: Madison, The Federalist 10 and The Federalist 51.

‡ Morgenthau’s democratic ideation seems to echo Lefort, for whom democracy was signified by an “empty seat of power” beneath which fluid and dynamic opposites competed for influence in a theater of contestation. See: Claude Lefort Democracy and Political Theory (Minneapolis: University of Minnesota Press, 1988), pp. 15–17; pp. 256–57.
mogenize diverse beings into a single sterile whole.

By contrast, in the United States, “neither the minority nor the majority is vouchsafed the correctness of its views—both must argue and act on the conviction that they are right, a conviction tempered by the awareness of the possibility that they might be mistaken.”65 To Morgenthau, where both the majority and the minority, “remain within this relativistic ethos of democracy, while at the same time respecting those absolute objective principles that are beyond the ken of that relativism, the vitality of their contest will accrue to the vitality of democracy.”66

Where it exists in such a form, and where it is additionally capable of securing itself and conquering the necessities of its citizens, the state exists in its highest form. For Morgenthau, such states provide the building blocks for a politically conceived ideal state-system.

Ideal Theory: The Moral Value of the State-System

Even if one accepts Morgenthau’s interpretation of the political life and grants that some type of localized, neutral, and democratically organized political community is essential for its realization, it remains plausible to ask, “why a system of territorially autonomous sovereign units”? Are there not, after all, avenues to pursue and preserve self-disclosing political action in other forms of social organization? Could not, for example, a world state facilitate the political life?

In this section, I address why, from a political perspective, a state-system is the ideal form of world order. By exploring why a global state is an insufficient facilitator of the political life, we will be able to more clearly identify the attributes of the state-system that make it the ideal reflection of the human condition, the best institutionalization of the political life, and the most effective curator of them both.
The Problem of a World State

In the mid-twentieth century, the experience of two world wars and the prospect of a third with thermonuclear weapons stirred up unprecedented agitation toward the formation of a world state. In the context of prolonged unspeakable catastrophe, it makes sense that security was of cardinal import to the statesmen and international scholars of the time. Convinced that only a global replication of the domestic security structure could secure international tranquility, proponents of a world state aspired to concentrate a legitimate monopoly of violence in the hands of a central world authority.67

As justified as these proponents’ blind prioritization of security might seem in such a situation, their judgment was void of a central consideration: the political. In effect, the object of the world state was, and still is, to eradicate international politics; to prevent different communities from communicating and contesting their disparate and often clashing interests that derive from the plural sets of values and priorities unique to each distinct political community.68 By overawing the voices of these diverse communities with a reconstituted global Leviathan, the world state denudes political communities of their expressive organs, stifling national conversations and their articulation in foreign policy. No world state can reflect the political organization of the human condition.69 Rather, the state is the “recipient of man’s highest secular loyalties” precisely because it is the embodiment of his most basic political values.70

In a value-pluralistic global terrain, a monolithic apolitical world state necessarily obviates the expression, and thereby devalues and defies the formation, of those plural values.71 The world state values only peace—but whose peace? Peace alone is a hollow virtue. Denied an international theater of independent interaction and halted by the heavy hand of the global sovereign, political communities under a world state become reduced to depoliticized shells with men vulgarized into mere mammals.

The highest forms of human freedom and the expression of
the political life depend upon a fluid, uninhibited vertical dynamic that the world state intrinsically crushes by its very establishment. Only a state-system can simultaneously reflect the diversity of the human condition and facilitate human freedom, while also utilizing characteristics of that system to preserve the conditions that secure diversity and the facilitate that freedom. The sub-sections that follow explore these aspects, respectively.

Sovereignty as an Expression of The Political

Within common law, the concept of sovereignty legally expresses the independence and basic equality of political communities in their legally organized forms. In these capacities, sovereignty transfers into law the political fact that all political communities are comprised of persons endowed with speech and equal in their capacity to contribute to the staging and execution of public life. In the first instance, sovereignty imbues states with a type of legal independence to institutionalize their distinct political character. This manifests at a basic level through the sovereign state’s supreme authority within a certain territory. More decisively, this means that a state is free to arrange its institutions and manage its internal and external affairs according to its discretion. This includes, Morgenthau writes,

The right to give itself any constitution it pleases, to enact whatever laws it wishes regardless of their effects upon its own citizens, and to choose any system of administration. It is free to have whatever kind of military establishment it deems necessary for the purposes of its foreign policy, which, in turn, it is free to determine as it sees fit.

In this positive dimension, sovereignty thus reflects the ability of a political community to freely and voluntarily define and project its values as interests in policy. Sovereignty in this respect dignifies the self-determining political beings that comprise a political community, and also recognizes the diversity intrinsic to the human condition. In the negative dimension, these principles are reinforced by the doctrine of non-intervention.
In the second instance, sovereignty creates a category of legal equality. The “sovereign equality” of states is the basic principle of international common law,* and it is operatively manifested through the appearance of reciprocation structures: for example, in treaty-making, it surfaces through the doctrine of reservation; in diplomatic relations, through dual consent procedures; and in international adjudication and arbitration, through dual-consent or mutually pledged compulsory jurisdiction. That similar structures are embedded throughout international common law signifies that equality is a fundament of the state-system. In practice, the construction gives Panama an equal voice with the United States; symbolically, it legally imbues political communities with the same moral equality in marco-form that speech-beings share in the human condition, valorizing both the diversity of the human condition and the dignity of the political speakers.

Sovereign-Equality and the Division of Power

For Morgenthau, the appeal of equilibrium seems to derive from Aristotle’s adage that virtue springs from self-limitation. In the international case, this refers to the virtue achieved when no single state can overwhelm the whole by dissolving, or subsuming, all the other various political communities into itself. More specifically, the appeal of an organic mechanism within a system composed of independent states possessing diverse energies and interests, is the possibility to achieve stasis within that system without impeding either the plurality of political convictions or the autonomy of the states issuing them. Through the self-regulation of power “balancing” (i.e., the fluid alignment and realignment of sovereign units to sustain and equilibrium in which power remains divided) the state-system of sovereign independent units is therefore able to perpetuate itself, maintain the pluralistic character of the human condition, and secure the political integrity and self-determination of the states composing it.

*It is certainly the foundational principle of the United Nations: “The Organization is based on the principles of the sovereign equality of all its members.” (Article 2.1, U.N. Charter).
Functionally, and most importantly, a system of sovereign states is dynamic enough to resist “great power predations,” while stable enough to withstand fluctuations in the moral and political attitudes of human communities.* The idea is related to Morgenthau’s belief that, sociologically, the spheres of human activity in civil society need to be robust enough to counter the potential expansion of the others, and that, in a democracy, broad pluralism can be utilized to stage a contest that sustains democratic life by keeping power “an empty place.” 78 The broader point Morgenthau makes in both instances is that equilibrium can be orchestrated between a pluralverse of autonomous units to secure the conditions (i.e., a division of power that blocks predation) necessary for the political good life to flourish.

But the division, or “balance,” of power is not a natural outgrowth of the international structure. Rather, it is a technique dependent upon the calculations and judgment of statesmen according to an ethic of national purpose (i.e., an ethic of responsibility) that fully considers the tactical and ethical consequences of their action, or inaction, upon not only their particular interests and values, but also universal values like the preservation of the state-system and the conditions necessary to cultivate the political good life.79 These product of these considerations may at any point urge the formation of alliances, augmentation of armaments, redistribution of materiel, as well as diplomatic machinations, symbolic displays and—when the integrity and continuity of the state-system is threatened altogether by predatory imperialism or inexcusable tyranny that challenges the political lives of all peoples—even full-fledged intervention and combat.†

Non-Ideal Theory: Imperialism, Despotism and War

To this point we have been concerned with ideal theory. In

---


extending a political conception of international order, we have
developed an ideal conception of the conditions necessary to fa-
cilitate what we have been calling the “political good life,” includ-
ing its institutional and legal manifestation through the sovereign
state system. A complete international normative vision, however,
develops through questions arising from the non-ideal features that
exist in reality. Beginning from the premise that there are at least
some ideal political communities who themselves abide by, and
hope to see all others eventually accept, the principles underlying
the conditions necessary to facilitate the political good life globally,
the task of non-ideal theory is to ask how these essentially demo-
cratic political communities should engage the other communities
that do not abide by those principles, either at all or in their highest
form. Our treatment of non-ideal theory in this section proceeds
in two parts. First, it considers the problem of “imperial states” that
pursue predatory foreign policies incongruent with the legal (i.e.,
territorial sovereignty and non-intervention) and moral (i.e., plu-
ralism and human freedom) principles of a politically conceived
international order. Second, it considers the trouble presented by
“despotic states” that are internally ordered, and which frequently
act toward their own citizens, in defiance of those same principles.
In the course of this second consideration, a third dilemma—that
presented by non-ideal, but non-despotic states (i.e., those that are
non-democratic yet legitimate)—is also considered.

Imperial States and Defensive War

An imperial state, in abrogation of legal and political prin-
ciples, seeks to impress its will upon other states at the expense of
those states’ capacity to will independently. This type of action—
whether by actual territorial expansion pursued with armed force
or by significant interference in the essential internal politics of
other nations—necessarily introduces violence into international
politics by silencing the autonomous political agency of the other
political communities (and the persons therein) that are its objects. *

* Morgenthau discusses the various modes and expressions of imperialism in Chapter 5 of
In whatever degree, the imperial state thus demeans the political potential of individuals and their capacity to conduct themselves as citizens. Such action further reduces the pluralism in the global arena and, in the gravest circumstances, risks upsetting the stability of the state-system altogether by threatening to “occupy” (in the Lefortian sense) the global seat of power. Its mobilizations may therefore legitimately invite swift military retaliation, or indeed preemptive military action, to sustain or “otherwise” reequilibrate relations.

Non-intervention, a property of the moral value of the state, represents and transfers the political autonomy of the individual as a self-determining speech-being to the legal level of the community. Violent action—which constitutes non-political action aimed at subjugation, and which in fact undermines political activity altogether—forfeits a state’s claim to non-interference and permits, and may even invite, third-party intervention into its conduct. For example, a creditor state that moves to occupy a neighboring debtor state on the claimed basis of unpaid debts cannot offer its balance sheet as justification for its action. The act of territorial incursion, by crossing the categorical threshold into violence, forfeits any sovereign protections for the negotiation of that debt and invites third-party actors to intervene in the affair.

A political conception of international order consequently realizes that war is a legitimate political instrument in certain circumstances, and that there exists a clear, limited yet permanent need for war, no matter how states organize themselves (i.e., democratically, tyrannically, etc.). Early twentieth century moral liberals, in eager adherence to the ideas Kant developed in his essay “Perpetual Peace,” had perilously equated democratic self-governance with peace and enlightenment. They went so far as to outlaw all wars except for those that pursued democratic “liberation,” with the idea that a world of democracies would decide that force is an obsolete instrument for the practice of international relations. Morgenthau, in asserting the permanence of a limited right to war,


* See for instance Iraq and Kuwait in 1991.
was supremely critical of their conclusions. First, he rejected the notion that the proliferation of democracy would be sufficient to vanquish war once and for all. “The question of war and peace is decided in consideration of permanent factors [geography, national character, tradition, the distribution of power] regardless of the form of government under which a nation happens to live.” War, he insisted, was a permanent if tragic feature of international politics, and nations, of whatever constitution, could be peaceful under one set of circumstances, and aggressive in another. Moreover, it is worth noting that accepting a doctrine of “democratic peace” creates a difficult-to-falsify claim that an imperial state might make to justify its belligerent ambitions. Similarly, by establishing a category of democratic war premised on morality, moral liberals amplify the stakes of combat, excessively dramatizing war to the level that the means and ends of total warfare become justifiable objects. In a nuclear age, this is a dangerous norm to embed.

Second, Morgenthau argued that the undesirability of establishing perpetual peace made attempts to abolish war by voluntary pacts of absolute abstention foolish overtures that imperiled the division of power as well as the conditions necessary for the political good life to flourish. “Liberal governments,” he wrote, “have fought their wars not upon a free choice between war and peace, nor at the moment most propitious to them, but upon the initiative of non-liberal governments.” For Morgenthau, liberal U.S. policymakers’ general aversion to war-fighting masked the fact that a war against imperial fascism was a war enjoined by the national purpose of the United States, which is to say the universal principle to preserve the division of power and halt imperial states. “To deny that Fascist imperialism constituted a threat to the American purpose at home and abroad,” Morgenthau wrote, “was to deny the evidence of one’s senses.” Deluded by an ideology that had been reified in law through meaningless agreements like the Kellogg-Briand Pact (1928), U.S. leaders who “did not want to fight in 1931 or 1935 or 1938 on [favorable] terms,” had to fight in 1941 “on the terms of the enemy.”

Similarly, attempts to definitely codify binding legal cat-
egories of legitimate war-fighting are problematic insofar as they inevitably run up against the reality of political events that only human judgment can navigate. To Morgenthau, these attempts to legally define and standardize an absolute category of *bellum injustum* sacrificed human judgment to a vain legalism. At the precipice of war and grave consequence, only man's imperfect wisdom can evaluate alternatives, select the most prudent, and execute accordingly. While international criminal laws may be useful conduits to aggregate world judgment and establish norms for what constitutes illegitimate and imperial exercises of war power, they cannot—as interwar moral liberals sought, and as their modern-day counterparts still seek—a priori prohibit war-fighting altogether.

Rather, in accord with a political conception of international order that recognizes states as the principal agents in international affairs, it falls upon the community of nations to adjudicate the legitimacy of controversial international actions that violate principles which transcend individual jurisdictions, deliberate possible courses of remediation, and generate consensus upon the chosen response, be it punitive, interventionary, or otherwise.

Some may find this construction uneasy and prone to issue arbitrary and inconsistent judgments. But the unease some may find with this construction comes, Morgenthau writes, from a failure to accept the ineluctable tragedy of political action:

> In the combination of political wisdom, moral courage and moral judgment, man reconciles his political nature with his moral destiny. That this reconciliation is nothing more than a modus vivendi, uneasy, precarious and even paradoxical, can disappoint only those who prefer to gloss over and to distort the tragic contradictions of human existence with the soothing logic of a specious concord.

Moral dilemmas are inherent in political action, and especially difficult to resolve when it comes to identifying and remediating imperial states. Faced with these dilemmas, a politically conceived international order chooses to resolve them through human judgment equipped with an ethic of responsibility rather than a "soothing," but ultimately disengaged legalism. It trusts the poli-
cal judgment of statesmen, and citizens, to weigh unquantifiable competing values and make the responsible choices.

Despotic States, Human Rights, and Permissive Interventions

While for the past fifty years, the world has been generally successful at containing the territorial and political ambitions of imperial states, it has been less effective at dealing with the despotic states that continue to terrorize, oppress, and even murder persons, including their own citizens, within their borders. The central complication plaguing efforts to address these states is an unclear standard for despotism. Imperialism, while perhaps not as clear as “aggression” as defined in international law, is by contrast a simple identification. A common but circular response is that despotism exists wherever human rights are violated, which is to say when foreign parties are motivated to intervene. A political conception of international order strives to at least partially remedy this confusion by developing a clearer conception of what constitutes a human right. In fact, there is not only a cogent political conception of human rights embedded within Morgenthau’s thought, but also one that suggests motivations for what we might deem as humanitarian intervention.

Morgenthau presented his ideas on human rights and humanitarianism most clearly in a lecture at the Carnegie Council’s Center for Religion and International Affairs in 1979. Though he affirmed to his audience that “there are basic moral principles applicable to all human beings,” he cautioned that these principles could not be expressed in terms of “natural” rights because such constructions necessitate reference to theological sources that “have very little to do with a philosophy.” Rather, for Morgenthau, rights can only be intelligible if they are conceived politically with reference to the collective actions and commitments of some political community:

I object to the concept of rights. The concept of rights already presupposes a society that gives the rights. When you talk about

natural rights, you assume the existence of a Divinity that is the King of the universe, in which there are rights.91

Man’s right to have rights therefore involves membership in such a political community (the “society that gives the rights”) because he is a “political animal.” We have human rights not because we are, but rather because of what we are: political beings who require a community to disclose themselves and to encounter the meaning of freedom. Though the extent and content of the rights men give themselves will vary, as Morgenthau notes, “both in time and place,” the decisive aspect is that men have the opportunity to give themselves rights at all, that is, to belong as members to a political community.

But what kind of membership? Joshua Cohen, in positing one version of “politically conceived” human rights, argues that membership entails self-determination. He interprets this to mean access to a political process that responds to the diverse interests and opinions of those subject to the laws and policies of the government.* He further interprets self-determination to create human rights to dissent, expression, and even the right to receive justifications for public policies. Cohen claims that this threshold for membership does not amount to a “human right to democracy” because it still allows for polities (perhaps those that Rawls typifies as “burdened societies” in *The Law of Peoples*) to develop internal orderings that do not reflect an association of political equals. Yet, his broad construction, with its lofty threshold, greatly infringes upon the political principles embedded in the legal concept of sovereign equality.92

A more suitable political conception of human rights, and one which seems to better correspond to Morgenthau’s position, derives from the principle of membership as inclusion.† This construction is perhaps best illustrated through the example of state-

---


† Recall that this was the minimum standard, alluded to earlier, of what conditions would be required for a state to possess moral value and thus make legitimate claims to sovereignty.
less persons. As Morgenthau noted, without the state, individuals cannot achieve their purpose as political beings.\textsuperscript{93} This is because statelessness entails the loss of the relevance of speech and that of all human relationships. Men are only men insofar as they command the power of speech and live in a community. Excluded from political conversation, and de facto removed from political communities, stateless persons are lonely figures trapped outside the pale of human interaction. Denied access to the public realm, stateless persons lose their essential humanity and become easy targets for propagandists, tyrants, and demagogues. Only the loss of a polity itself, in this sense, expels men from humanity.

Here we observe a contrast with the earlier discussed example of the man trapped in a less-than-democratic state. That man, though his voice is muffled by tyranny, still possesses his basic ability to speak and protest. Ignored and beaten, such a man still joins his compatriots in ancient song, \textit{Hine ma-tov u'ma-nayim, shevet akh-im gam ya-had}—“how good and sweet it is when brothers dwell together in unity!”\textsuperscript{94} Meanwhile, the stateless man has no brothers with whom he may dwell, no song to be sung, and no protest to stage. They are, as Morgenthau was, “\textit{ausgeschlossen und ausgestossen}” (excluded and expelled).\textsuperscript{95} To speak of human rights is therefore to speak of the right to access an arena in which, as Michael Walzer described it, “freedom can be fought for, and sometimes won.”\textsuperscript{96}

A despotic state, therefore, is one that exerts exclusionary pressures on a community to the point of committing it to a political death.\textsuperscript{*} The Holocaust, for instance, was a “moral outrage”\textsuperscript{97} to Morgenthau which, unlike other moral judgments of political action, could not be relativized and tolerated as a product of “diverse attitudes” toward the value of human life. The organized depoliticization of the Eastern European Jewish community that culminated in their “mass extermination,” Morgenthau wrote in 1948, must be abhorred “by virtue of an absolute moral principle

\textsuperscript{*} To reformulate an earlier point, we might say that the danger in what Morgenthau typified as “nation-states” can thus be restated as the risk that they will lapse into “despotism” by excluding certain populations from public life on the basis of the ethno-nationality.
the violation of which no consideration ... can justify." Their total dehumanization and near elimination, Morgenthau wrote, "was based upon something different from religion." That is, the fact that, "a Jew, by his very existence, denied the universalistic aims and philosophies" of the Third Reich, and was therefore unfit to participate in public life. Political and eventually physical annihilation "became inescapable" because the Jews, like the Roma, gays, and other non-Aryan groups, represented a pluralism that the Reich could not justify.

Understood in this way, sovereign equality cannot be claimed by states to shield themselves from interventionist action should the community of nations agree that a violation of the principle of membership has occurred. Genocide treaties have begun to accumulate what amounts to a similar principle, though a political conception advises future authors to carefully specify the inclusive political grounds (opposed to vague statements about human dignity and the value of life) upon which the protections in their treaties are based. It would be better to codify such a principle, appropriately substantiated with proper references to political inclusion, into hard international law. Any such codification or treaty should ultimately make clear that any violation of human rights (i.e., exclusion, political death) is subject to international intervention should a properly constituted body representing the international community confirm such a violation.

A different calculus applies to proposed interventions targeting non-democratic, but non-despotic states (i.e., non-ideal yet legitimate states). Contra those who might claim that a "human right to democracy" justifies such interventions, a political conception holds that the universal validity of the democratic political ethos can be realized if, but only if, democracy is independently ascertained and pursued through the self-determined political action of the people. If there exists the possibility of influencing the laws and institutions that govern them (i.e., access to a theater of contestation as political equals), it is for the members of the self-asserting political community to ensure for themselves that the claims made in that theater are ultimately respected by the government."All human
beings want to be free, and, hence, want to have those opportunities for self-expression and self-development,” Morgenthau wrote. A political conception of international order thereby holds democracy to be an idea to which all men intuitively aspire. Self-development, as a principle of democracy, therefore requires democratic states to limit their proliferative activities to what amounts to a type of passive evangelization. Unlike Wilson and contemporary hegemonic liberals who wanted “to transform the world through the will of the United States,” Morgenthau held that it is the obligation of democratic states to “attract the rest of mankind through their example,” and not to “impose” these principles upon other nations. To do otherwise and intervene wherever authoritarianism existed would be to deny members of a political community the opportunity to determine for themselves their government, laws, and, as Jean Cohen points out, the “learning processes entailed by their struggles (including the ability to compromise) to create a more inclusive political process, more just laws, and new interpretations of rights.”

Morgenthau, in his day, held that military force could not be levied in Vietnam to halt or install ideas that are “political in nature.” While it was reasonable to oppose the Soviets in Europe, where throughout the Cold War there was a tangible military threat that the Russians would expand imperially through Berlin and along the Iron Curtain, the ideological nature of the “threat” in Southeast Asia did not merit even limited combat. The values chosen and expressed by the Vietnamese polity were, in other words, for the Vietnamese to decide for themselves and not for the United States to oppose or demean. To do otherwise would depoliticize the Vietnamese political community and deprive its citizens of their most basic political agency. As the major power in that region, Morgenthau believed it was natural that Chinese communist ideology would influence the thinking of self-determining Vietnamese in the construction of their own government and the formulation of their political values. The United States needed to recognize this political and cultural influence “as a fact of life,” and focus instead on rhetorically and diplomatically encouraging democratic elections in Vietnam that would allow the country to flour-
ish as a Titoist state, free from the imperial regional ambitions of either the Soviet Union or China.\textsuperscript{107} “Those universal principles,” Morgenthau wrote, “were not to be exported by fire and sword if necessary, but they were to be presented to the rest of the world through...successful example.”\textsuperscript{108}

A politically conceived international order thus only consents to a limited category of “humanitarian intervention.” Additionally, against the assertions of some of the more enthusiastic cosmopolitan liberal theorists like Kor-Chak Tan and Carla Bagnoli, a politically conceived international order holds that potential “humanitarian” actions incited by despotism are by no means morally obligated. As Morgenthau said at the CRIA lecture, “the normative force of the [universal moral] code is qualified by potential considerations. I mean what we call circumstantial ethics.”\textsuperscript{109} As Stephen Wertheim rightly observes, though the language of absolute obligations “enchants us,” drawing as it does on a “moral purity that aspires to transcend the political,” such ethics of ultimate ends focus upon whether we are moral, not how to act to actually benefit others.\textsuperscript{110} The decision to act, or in this case intervene, must not be decoupled from considerations of how that action or intervention will be executed. If these considerations of the “how” reveal consequences and scenarios that exacerbate the evil or undermine the objectives of the action altogether, then the action should not proceed. From the perspective of a political conception, this is the ethically responsible decision. The blind establishment of a duty to confront evil as a reflexive norm of international politics (i.e., an incontrovertible “responsibility to protect”) would, in its defiance of politics, divert attention from helping victims to challenging evil-doers themselves. Even when it is substantiated by the appropriate precepts and perspectives, such a backward, decadent humanitarianism that in a sense glorifies the righteousness of liberators above the welfare of the damned cannot be countenanced. Judgment, uneasily, remains a statesman’s best and only recourse.

IV. BETWEEN GLOBAL CITIZENSHIP AND EMPIRE

Contra the liberal cosmopolitan and liberal hegemonic positions, the above political conception posits a “dualist” concept for international law and order. This concept reconciles human rights as inclusion (which legally represent man’s need for the polity) with sovereign-equality (which legally represents the equal status of political beings and the diversity of possible value-orientations in a post-differentiated modernity). The dual basis for this international order creates a new human legal-political status between the global citizenship of the cosmopolitan liberals, and the imperial vassalage of the hegemonic liberals.

Legal Concepts for International Order

In the cosmopolitan model that Habermas describes, individuals are doubly citizens of a polity and a global community. By contrast, under the dualist concept of international law espoused by a politically conceived international order, individuals are exclusively citizens of a sovereign state. However, all of those states, without relinquishing their sovereign status, are themselves subject to the selective penetration of global governance institutions and international laws.

The reconciliation of the seemingly antithetical proposition is accomplished by embracing what Neil Walker and Jean Cohen term “constitutional pluralism.”* Constitutional pluralism begins from the premise that there have been “shifts in the substantive rules of sovereignty” as a result of “the emergence of a functionally differentiated autonomous global legal order” that makes its own claims to autonomy, supremacy, and constitutional quality “alongside the continuing claims of states.”111 It is thus feasible to decouple the internal supremacy of a state’s laws from the idea that states possess exclusive legal jurisdiction over their persons and ter-

This is not to propose a radical redefinition of state sovereignty. Sovereignty, as has been argued, is a powerful normative category that extends beyond basic moral principles to protect the self-determining actions of members of a political community. It preserves their capacity to judge and amend the character of their government and its institutions as they see fit. Sovereignty cannot be reduced, for example, to a “responsibility to protect.” Rather, the move to a “pluralist” or “dualist” approach recognizes that the sovereignty argument is not unfalsifiable and indeed limited by universal political values (i.e., human rights as inclusion) consistent with the normative political principles that justify the sovereignty argument in the first place. As alluded to in the preceding section’s discussion of how the international community should engage imperial and despotic states, a constitutional pluralist position merely establishes that the international political community may legitimately refer to international law to reject a state’s sovereignty argument and justify interventionist or punitive action.

As Jean Cohen notes, global governance institutions, especially those with coercive power (i.e., those currently possessed by the UN Security Council), need to be universally inclusive and operate using rules and procedures that incorporate the inputs of all states. This would prevent a partial transference of jurisdiction from abolishing sovereignty while ensuring that decisions that reject the sovereignty argument still legitimately represent the consensus of the international community. Without this step, the legality, legitimacy, and general concert of the politically conceived international order falls apart. A general prescription of this thesis is therefore an exhortation to immediately begin the hard process of reforming these institutions.*

To be sure, the application of an inclusive principle contrasts with other noted concepts for international order (namely Rawls’ in *The Law of Peoples*) by avoiding the “creation of radical rightlessness

---

* Since, as with domestic political institutions, these global governance institutions will be “susceptible to hegemonic capture,” democratic reforms that keep power divided also seem to be appropriate. See: Ibid., p. 597.
and lawlessness zones” (i.e., the designation of what Rawls called “outlaw states”) that fall outside the coverage of international law because they fail to accept the basic principles that undergird the international system. Contra Rawls, criminal governments and persons are subject to the sanctions of international law under a politically conceived international order “precisely because” they are members of international society. This adds not only the benefit of internal theoretical consistency, but it also avoids what Morgenthau considered the risk of excessively amplifying the brutality of conflict and punishment through the implicit “dehumanization” of agents and entities labeled “outside” of humanity.

The human political status created by a politically conceived international order is therefore one that sidesteps global citizenship (by salvaging a concept of the political through its legal representation in sovereignty) and vassalage (by categorizing the imperial hierarchy proposed by hegemonic liberals as a violation of basic normative principles). It affirms and secures the political dignity of individual citizens, including protections from the predations of their own governments.

Against the false choice between cosmopolitan and hegemonic liberalism, this political conception shows an appreciation for man’s political nature and his need of polity by concurrently celebrating and institutionalizing the eminent goodness of the public political life. Since the end of the Cold War, influential democratic states have regularly neglected the long-term ramifications of their actions in their haste to confront evil. As the normative structure of the social world has consequently bent toward the liberal discourse of human security, the extension and cultivation of the public political life—the “good life”—has been imperiled. Without a robust alternative to compete with dialogically, it is unsurprising that the liberal discourse has so far prevailed. Ideas must be mobilized that can supply an alternative to challenge existing liberal modes of thought and salvage the political life from decadence, despotism, and hegemonic usurpation.
Notes


5 Habermas, *Divided West*, pp. 172–73.


11 Ibid., p. 45.

12 Ibid., pp. 15–16.


15 Ibid., p. 6.

16 Ibid., p. 188.

17 Ibid., pp. 189–90.

18 Ibid., p. 195.

19 Ibid., p. 194.

20 Frei, p. 9.

21 Ibid.


24 Ibid., p. 196.


31 Ibid., p. 154.
33 Morgenthau, Politics Among Nations, p. 249.
34 Morgenthau, Science: Servant or Master?, p. 31.
35 Morgenthau, Scientific Man vs. Power Politics, p. 144.
37 Ibid., p. 28. Morgenthau is referring to Book 1, Chapter 2, 1253920–30 in Aristotle’s The Politics.
38 Ibid., p. 30.
39 Ibid.
41 “Men, not Man, live on the earth and inhabit the world.” Ibid., p. 7.
44 Ibid., p. 176.
45 Morgenthau, Purpose, pp. 197–215. See also: Arendt, Human Condition, p.221;
46 Arendt, Human Condition, p. 301.
47 Morgenthau, Purpose, p. 204.
49 Morgenthau, Politics Among Nations, p. 497.
50 Ibid. See also: Weber, “Politics as a Vocation,” in From Max Weber, p. 79.
51 Morgenthau, Politics Among Nations, p. 509.
52 “The telos of the state is not just to ensure the bare survival of the citizens.”
56 Arendt, The Human Condition, pp. 30–31. See also: Cohen and Arato, pp. 180–81. Cohen and Arato note that in the Greek polis to which Arendt generally refers, the conquest of necessity primarily involved organizing the household in a manner that designated economic activity as oikos, thereby permitting only the male head of household with sufficient time to exercise public freedom while relegating women to the home. In the modern social-democratic republic, this deprivation has been largely, though not entirely, resolved. Their critique is nonetheless worth noting.
58 Morgenthau, Politics Among Nations, p. 497.
60 Morgenthau, Purpose of American Politics, p. 56.
61 Morgenthau, Politics Among Nations, p. 507.
62 Morgenthau, “The Right to Dissent,” in Truth and Power (New York: Praeger,
64 Ibid., p. 41.
65 Ibid., p. 44. Morgenthau, like Lefort, was nonetheless entirely aware that democracy was a fragile creation whose own close links to relativism risked, on the one hand, generating alternating periods of blind majority-rule, and on the other, its own decline into something like totalitarianism. As the cited passage concludes, “Otherwise they will strain the delicate ties that keep a democratic society together, and they will risk destroying it while trying to keep it alive.”
66 Ibid.
67 Morgenthau, Politics Among Nations, pp. 491–97.
69 Morgenthau, Politics Among Nations, p. 501.
70 Ibid., p. 502.
72 Morgenthau, Politics Among Nations, p. 318. Note again that a democratic model is not required for a state to claim moral value. It is the idea to which Morgenthau believes all men intuitively aspire. Rather, the criterion the ability for political equals to be included in a community where they can fight for their own freedom, which must be won, not given.
73 Ibid.
75 Morgenthau, Politics Among Nations, p. 319.
77 Aristotle, Nicomachean Ethics, II.6, 1006b36.
80 Morgenthau, Scientific Man, p. 65.
81 Ibid., p. 66.
82 Ibid., p. 67.
83 Ibid., p. 69.
84 Morgenthau, Purpose of American Politics, p. 118.
85 Morgenthau, Scientific Man vs. Power Politics, p. 69.
87 Morgenthau, Politics Among Nations, p. 363. Morgenthau points to the Kellogg-Briand Pact (1928), the League of Nations (1919), and two Hague Peace Conferences (1899, 1907) as examples of such efforts. See also: Morgenthau, “The Twilight of International Morality,” Ethics 58 (1948): 2, p. 85.
88 Morgenthau, Scientific Man, p. 203.
90 Ibid., p. 1, p. 25.
91 Ibid., p. 15. Emphasis added.
94 Psalm 133:1
97 Morgenthau, in *Political Theory and International Affairs*, p. 100.
100 Ibid., p. 255.
101 Ibid., p. 247.
102 Morgenthau, *Politics Among Nations*, p. 266.
106 Ibid.
107 Ibid.
108 Morgenthau, *Purpose of American Politics*, p. 34.
112 Ibid., p. 277.
114 Ibid., p. 594.
115 Ibid., p. 594.

Bibliography


The Concept of the Political. 2nd ed. Chicago: Chicago University Press, 1996.


DISCOUNTING THE STANDARD MODEL: WHY THE RAMSEY MODEL DOES NOT ADEQUATELY EXPLAIN CLIMATE CHANGE POLICY

Yangbo Du

The costs of action or inaction on climate change pose difficulties in terms of analysis in the framework of intergenerational welfare economics. Evaluation of mitigation policy is characterized by either immediate costs with delayed benefits or, when considering a policy of inaction, delayed costs. Sunstein (2007) contends that the delayed benefits of climate change mitigation policy result in a lack of willingness to undertake mitigation measures whereas more immediate benefits of counterterrorism measures results in greater public support of such policy. Differing assumptions regarding discount rates for future costs and benefits contribute in large measure to the high degree of uncertainty surrounding specific estimates of the net economic impact of climate change. For example, the social cost of carbon, or the equivalent Pigouvian tax on carbon emissions to correct for market failure due to externalities from the effects of high greenhouse gas concentrations, varies at the mean from 232 dollars per tonne carbon with a 0 percent discount rate to 18 dollars per tonne carbon with a 3 percent discount rate as reported in Tol (2009, p. 41). While appropriate methods and rates of discounting can be inferred via observable macroeconomic indicators, such methodology may exhibit imperfections even with refinements as outlined by Dasgupta (2008). His findings will be addressed in greater detail in the following sections. The intergenerational welfare implications within the Ramsey model—henceforth known as the standard model due

Yangbo Du graduated from the Massachusetts Institute of Technology in 2011 with a concentration in political economy. He currently serves on four planning and environment committees of the Transportation Research Board at the National Academies, where he seeks opportunities in energy, environment, sustainability, and international development.

*While his conclusion was drawn from a study of preferences among Americans for responses to terrorism and climate change, behavioral factors such as availability heuristic, probability neglect, outrage, and myopia (all of which lead to greater willingness to respond to immediate threats and lesser willingness to respond to distant threats) behind such preferences may result in similar preferences among a geographically wider population sample.
to its widespread use in macroeconomic analysis—are tested against empirical evidence from advanced economies. Importantly, this test demonstrates that an alternative explanatory framework may be necessary for better accounting of impacts from climate change and optimal mitigation policies. Alternatives to the standard model have undergone refinement in recent years from advances in psychological economics, from which ‘bounded rationality’ and time inconsistency in discounting inferred from controlled laboratory experiments can be applied readily to public policy evaluation.* While the focus is on the relationships between market indicators and aggregated decision-making in the form of national savings (or investment if equivalence between the two is assumed), realistically, it might be necessary to incorporate non-economic indicators into benefit-cost analysis even if internalizing all externalities would, in principle, ensure the adequacy of an analysis limited to market observables.

The most visible impact of differences in choice of discount rates lies in the optimal price of carbon emissions. Higher discount rates lead to lower present values of the benefits of lower greenhouse gas concentrations in the atmosphere and, in turn, lower opportunity costs of not emitting carbon; the reverse is true for lower discount rates. Any specific climate policy recommendation—from those promoting a ‘business-as-usual’ trajectory to those promoting aggressive mitigation in the immediate future—can be justified via some particular set of values for relevant parameters as demonstrated in Guo et al. (2006), Tol (2009), and Anthoff, Tol, and Yohe (2009a, 2009b), which makes separating climate policy evaluation from value judgements stemming from individual or societal perspectives impossible. This paper, which provides a backward-looking framework for evaluating future strategies to deal with anthropogenic climate change, is not intended to provide any definitive

* In Dasgupta and Maskin (2005), a study of time inconsistency in discounting involving preferences of starlings and pigeons demonstrated that given uncertain outcomes of alternative choices, longer time horizons tend to favor patience and shorter time horizons tend to favor impatience due to evolutionary pressures. It is worth exploring whether having longer time horizons cause decisions favoring strong action against climate change to dominate those favoring inaction or weak action.
judgment of the means of ensuring intergenerational equity in policy analysis or the optimality of any particular method of discounting; rather, its empirical component is geared toward opening up avenues for future research in the welfare aspects of mitigating climate change.* In a similar vein of argument to that of Tóth (2000), there is no presumption that benefit-cost analysis with discounting is necessarily a proper means of evaluating climate change mitigation policy, but rather an extension of theoretical arguments proposed by Dasgupta (2000, 2009), who has raised concerns about the efficacy of intergenerational welfare economics in addressing issues to which it is directed: empirical data to test the validity of the standard model and to examine its underpinnings in practice.† The practical aspects of the findings from our exercise rely on the presumption that the standard model is a valid descriptor of intergenerational welfare economics. There is ample evidence from both theory and practice to substantiate claims that the standard model is of limited applicability in questions of intergenerational welfare, and the plausible nature of such claims is highlighted in this paper.‡ Arguably, one can interpret the empirical argument of this paper, while focused on discounting, as justification for finding alternative methods of evaluation.

ANALYTICAL FRAMEWORK: THE STANDARD MODEL

The Ramsey equations are used in this paper for determining optimal savings decisions according to the standard model in continuous time. The parameter definitions in this section, originally put forth by Ramsey (1928), are loosely adapted from those by

---

* Among many sources arriving at similar conclusions, Karp and Tsur (2007), who derived ‘efficient’ discount rates over time via a game-theoretic approach, maintained that quantitative analysis of policy-making could not be expected to resolve policy disagreements given inherent value judgements reflected in the structure of any economic study.

† Dasgupta (2008) concludes: “Intergenerational welfare economics raises more questions than it is able to answer satisfactorily” (p. 167). This assertion served as the main impetus behind the writing of this paper.

‡ While challenges to the standard model abound in studies on the economics of climate change and intergenerational welfare from behavioral or other non-neoclassical perspectives, for reasons of scope, here they are not scrutinized to the extent done so for the standard model.
Dasgupta (2008) and Levhari and Srinivasan (2009) to provide a clear basis for our analysis while maintaining relevance to the scope and objective of this paper. That the economy consists only of rational, maximizing actors with constant elasticity of inter-temporal substitution (CEIS) utility ensures the consistency of this model in principle; deviations from this and related assumptions are discussed later in this paper. Save for some changes in definitions of parameters, the standard model underlying the author’s analyses is fundamentally identical to the model employed in Dasgupta (2008).

Formally, the equation for determining the social discount rate is expressed in its deterministic form in (1),

\[ \rho = \delta + \eta g \]

in which \( \rho \) is the social discount rate, \( \delta \) is pure rate of time preference, \( \eta \) is marginal elasticity of inter-temporal substitution or marginal aversion to inequality, and \( g \) is growth in future output or income.\(^*\) The pure rate of time preference can be treated as a quantification of consumer impatience, and higher growth leads to higher discount rates to the extent that the elasticity of inter-temporal substitution is positive. This is typically true since consumption smoothing over one’s lifetime causes current consumption to rise in response to a permanent increase in expected future disposable income. All rates are expressed in per-capita terms and compounded continuously. In its stochastic form, the equation for social discounting becomes (2),

\[ \bar{\rho} = \delta + \eta \bar{g} - \eta^2 \sigma^2_g / 2 \]

\(^*\)Ambiguity persists in finding a proper definition of \( g \). Some sources including Dasgupta (2008) define \( g \) as growth in consumption. Whether growth of consumption or growth of income is more suitable for use in the Ramsey equations depends on whether preferences are revealed \textit{ex post} or decisions are formed based on future expectations; consumption lags income if the latter case applies. One might inquire whether households assign greater weight to expected future consumption or expected future income when attempting to smooth consumption, since if the social discount rate represents returns to consumption, per capita consumption growth is more appropriate of a definition of \( g \). Alternatively, there is a plausible case for defining \( g \) as per capita income growth when considering that higher expected future income leads to higher discount rates, which in turn heightens propensity toward consumption. It is implied in this paper that consumption changes in response to income, henceforth providing a stronger rationale for defining \( g \) as growth in income (normalized for working population and working hours).
in which $\bar{g}$ is future growth as an expected value, $\bar{\rho}$ is the risk-free social discount rate (analogous to the definition of risk-free return on investment in finance literature, and equivalent when perfect information is assumed), and $\sigma_g$ is the standard deviation of growth. All random variables mentioned in this paper are assumed to be log-normally distributed. The effect of elasticity of inter-temporal substitution on the discount rate now depends on mean and variance of growth rates. Because households and firms are assumed to be risk-averse and exhibit precautionary saving behavior, volatility in economic output dampens consumption growth, even as higher income growth promotes it.

Savings rates follow directly from the social discount rate and its determining variables. When growth is deterministic, the savings-output ratio $s$ can be calculated via (3),

$$(3) \quad s = (r - \delta)/(\eta r)$$

in which $r$ is interest rate defined as the opportunity cost of capital in equity or debt markets. Higher levels of consumer impatience, as represented by $\delta$, depress savings, and higher opportunity costs of capital raise savings by rendering investment more attractive relative to consumption, albeit in diminishing increments up to an upper limit of $1/\eta$. It makes no difference whether one uses the market interest rate or the social discount rate as long as capital markets are perfect, but due to imperfections in capital markets, it is reasonable to expect considerable discrepancies between market rates and social discount rates. If growth and market interest rates are stochastic, the savings-output ratio can be determined via (4),

$$(4) \quad s = (\bar{r} - \delta)/(\eta \bar{r}) + (\eta - 1)\sigma_r^2/(2\bar{r})$$

in which $\bar{r}$ is the risk-free return on investment and $\sigma_r$ is the standard deviation of return on investment. This expression accounts for precautionary saving in response to uncertainty over future states of the economy, explicit in the second term. All analyses presented in this paper are based on the stochastic Ramsey equations.

In this paper, at best, a proxy for social discount rates can be defined since their determinants cannot be as easily quantified as
rates of return in capital markets. Whereas capital compensation relative to built capital stock provides easily an observable indication of market interest rates, returns to consumption would be an analogous indicator of social discount rates. Monetizing the returns to consumption is a topic beyond the scope of this paper. Instead, one can consider social returns as returns to all factors of production as opposed to returns to built capital only.

**METHODOLOGY**

The key exogenous parameters in the standard model—the pure rate of time preference and the elasticity of inter-temporal substitution as used in previous economic analyses of climate change by Cline (1992), Stern (2006), and Nordhaus (2007)—are estimated by using long-term trends in savings, growth, interest rates, and returns to capital stock. The applicability of the standard model used for discounting climate change is then examined by comparing actual savings rates and those implied by discount rates derived from observed market behavior. Growth, interest, and return rates reported in this study should be compounded continuously as defined in Section II, and thus differ (albeit slightly in most cases and more so with higher rates) from rates reported on an annual compounding basis. This adjustment enables simplifying calculations while maintaining accuracy.

Data on economic indicators used in this study were accessible online via the EU KLEMS Growth and Productivity Accounts, the International Financial Statistics (IFS) database through UN data, and the EarthTrends—Economics, Business, and the Environment database. These databases were projects of the European Commission, the International Monetary Fund, and the World

* Given the necessity of making value judgments when assigning pecuniary value to non-pecuniary objects (happiness and pleasure for instance), it may not be possible even in theory to monetize utility accurately and impartially.

† Section II of this paper contains all symbolic definitions related to the Ramsey equation for discounting. For example, deriving real interest rates using continuously compoundable figures requires only that one subtract inflation from nominal interest; application of log rules demonstrates the equivalence of this calculation to dividing the annual interest (or discount) factor by the annual inflation factor.
Resources Institute (WRI), respectively.* The EU KLEMS Basic and Additional files contained data in panel time series arranged according to country and industry; details have been documented by O’Mahony and Timmer (2009). Datasets from IFS and Earth-Trends were presented in panel format by country and year of observation, and in time series format by country or bloc, respectively. To facilitate running the regressions described in this section, all data taken directly from the original datasets or derived from raw measurements were formatted as time series by country and calendar year, and reported as annual aggregates, annual averages, or year-on-year indices in the cases of savings, returns, and growth respectively. Only countries listed in the 2009 releases of the EU KLEMS basic and extended datasets were included in the analysis, thus limiting the geographic scope to only the twenty-five European Union member states as of 2007—no data were published for Romania and Bulgaria, which joined the EU that year—plus Australia, Japan, South Korea, and the United States.† Savings, growth, and real interest rate parameters were assumed to be exogenous and were thus estimated via univariate auto-regressive integrated moving average (ARIMA) forecasting models with exact maximum likelihood estimators.‡ The only exception lies with the disaggregated figures for natural, built, and intangible capital, which exist only for the year 2000. Discount rates on consumption were directly calculated from these figures by dividing gross national income per capita over total capital per capita, both with and without the inclusion of natural capital stocks. That future macroeconomic performance is consistent and reflective of past performance was a crucial assumption underlying the choice of ARIMA regres-

---

* WRI gathered data from the World Bank, the United Nations Statistics Division, and the Organisation for Economic Co-operation and Development when compiling its searchable database. Descriptions of methodology are accessible via web link from the data output page.
† Time series data for Canada, while contained in the EU KLEMS database, were not published in the most recently updated version. Canada was omitted from the analysis to maintain consistency within the sample.
‡ Successful estimation of the ARIMA models for interest rates and savings rates require that the time series converge, which was the case with the data set used in this paper. The convergence parameters, which indicate convergence if absolute value is less than one, are listed under ‘Phi’ in the tables along with savings and interest rates produced via the models.
sion. While the long time horizon of data analysis mitigates effects of business cycles and other economic fluctuations on regression output, regression results do not account for developments not observed in the past, with adverse economic consequences of catastrophic climate change being among the most relevant to this study. No seasonal adjustments were deemed necessary due to the annual and year-on-year reportage of data.

Savings rates as used in this study are gross national savings rates as published by the World Bank and the World Resources Institute.* These time series consist of gross national savings, defined as total saving by households, firms, and governments, reported as a proportion of gross national income (GNI) from 1970 to 2006 or during a shorter period within this date range depending on availability of figures as specified in section IV of the appendix under “Savings-output ratio.” To determine the long-run behavior of savings rates, an ARIMA regression of autoregressive order of one and zero differencing and moving average orders was run for each time series of savings rates, hence an AR1 model.

Model 1. 
\[ \hat{s}(t) - \Phi_s s(t - 1) = \mu_s \]

ARIMA(1,0,0) for savings-output ratio \( s \); \( \mu_s \) as mean, \( \Phi_s \) as auto-regressive coefficient

Labor productivity as expressed in gross value added per hour worked was chosen for representing expected future levels of consumption. This measurement confers the advantage over gross national income measurements of not being augmented or depressed by changes in labor force size or changes in total hours worked. No loss of non-working or leisure time is necessary if an agent expects income to increase at the rate of labor productivity growth. Furthermore, labor productivity is constrained by technology only whereas income per capita and gross income fall under constraints of not only technology but also time and labor force size. The basis of growth forecasts therefore consisted of logs of labor productivity indices reported in EU KLEMS, which were already normalized to 1995 levels of 100, since 1970 or the earliest available year up

* Country-level data is identical in compilations by either entity.
to and including 2007 or the latest available year. The number of data points present in each country’s series is listed in section IV of the appendix under “Labor productivity.” ARIMA regressions for labor productivity were then conducted using a differencing order of one and autoregressive and moving average orders of zero, hence a random walk.

\[ \mu_g = Y_{LP}(t) - Y_{LP}(t - 1) \]

ARIMA(0,1,0) for mean labor productivity growth \( \mu_g \); \( Y_{LP} \) as index of labor productivity

Risk-free real interest rates were inferred from nominal interest rates on long-term (ten- to thirty-year) government bonds and price level indices. Periods of high inflation coupled with high nominal interest rates thus do not skew the data in any way.* Nominal interest rate time series were accessed from the United Nations database of financial indicators, and price level time series were published in the EU KLEMS database for the twenty-nine countries contained therein. The dataset spanned years 1971 through 2006, and count of data points for each country’s series is displayed in section IV of the appendix under “Real interest rate.” Taking natural logs of annual interest and inflation factors adjusts the published figures for continuous compounding and simplified calculation as described earlier. Long-term trends were forecast using ARIMA regression with autoregressive order of one and differencing and moving average orders of zero, hence an autoregressive AR1 model.

\[ \hat{\Phi}_r r(t) - \Phi_r r(t - 1) = \mu_r \]

ARIMA(1,0,0) for interest rate \( r \); \( \mu_r \) as mean, \( \Phi_r \) as auto-regressive coefficient

Three different expressions were produced for returns to

---

* It is plausible that households respond to nominal interest rates due to money illusion, in which inflation is not properly taken into account as is done in the case of real interest rates. While it is assumed in this study that real interest rates are the only relevant measure in order to be consistent with the standard model, further study of this phenomenon will be necessary in order to refine models of discounting across very long time horizons. Furthermore, periods of very low or negative real interest rates can be attributed to high inflation in many cases, so the results will still take into account inflationary expectations.
capital: returns to built capital only, returns to all forms of capital, and returns to all forms of capital other than natural capital. These figures serve as a guide for expected returns to equities if applied to the aggregate economy rather than to individual sectors or firms. In the first case, one obtains a ratio by dividing total capital compensation by total built capital stock. Data on thirteen of the twenty-nine EU KLEMS countries were available, with count of data points specified in section IV of the appendix under “Returns to built capital.” This measure indicates only the capital compensation realized in a particular year regardless of the exact year during which the initial investment was made. Overestimation of the economy-wide returns to equities is therefore likely, so the rate of returns to built capital was set as an upper bound for actual such rates. The latter two means of calculating returns to capital attempts to take all capital into account. Assuming that GNI represents the interest earned on the gross national capital stock, the rate of return to capital can be derived by dividing GNI per capita over the sum of natural, built, and intangible capital (human capital and other ‘invisible’ forms of capital) per capita. It is assumed that all returns to labor stem from human and all other capital inputs expressed as unit labor productivity. Alternatively, if it is assumed that built and intangible capital already accounts for GNI derived from natural capital, the rate to capital may be expressed as the ratio of GNI to the total built and intangible capital stock. This assumption, however, is questionable if natural capital is valued by itself and not for the utility of converting it to produced capital; regardless, such figures are still produced under this assumption for purposes of comparison. Data for returns to capital were insufficient for fitting ARIMA models with exact maximum likelihood estimators, so the average rates of return to built capital in the future were assumed to be similar to those from the 1970s to the 2000s, as the dataset provided no substantiation for expecting significantly higher or lower long-term average rates of return. Exact figures for GNI and capital for the year 2000 were used in deriving pure rates of time preference and optimal savings rates from the alternative measurements of returns of capital. That 2000 was a boom year for most of
the advanced economies in our sample may warrant setting future expected returns to all capital lower, but there are currently no data to justify such an adjustment.*

**ANALYSIS AND DISCUSSION**

Our parameter estimation exercise has yielded indicators of discount rates that are reasonably strong to the extent that the standard model is applicable. Mean labor productivity growth rates produced via the univariate random walk regression of labor productivity are identical to the means presented in Table I.2 under “Labor productivity growth.” Output from the AR1 regressions for real risk-free interest rates and savings rates are displayed in Tables II.1 and II.2, respectively. Listed alongside the coefficients and standard errors in Tables II.1 and II.2, statistical significance of the estimates of long-run behavior of savings, interest, productivity, and returns to capital were quite high in general, albeit considerably less so for interest rates than for the other measures. Of all twenty-nine countries in the EU KLEMS dataset, seventeen of them exhibited a strong long-term constant trend in real interest rates upon applying a ninety percent confidence interval. With a narrower 95 to 99 percent confidence interval, 15 and 13 percent of those countries, respectively, still retain a statistically significant mean forecast interest rate. Savings rate data exhibited higher significance of forecast mean and estimated autoregressive coefficient, as shown in Table II.2. The highest significance occurred within the dataset for labor productivity; no $p$-critical value exceeded 0.0001 (in the case of Greece – all other $p$-critical values were negligible).

Theoretical outcomes for savings, pure rate of time preferences, and marginal aversion to inequality proposed in Dasgupta (2008) are compared to outcomes determined via empirical evidence. Results do not appear to be consistent with mechanisms in

*To quantify the costs and benefits of mitigating climate change more precisely will require additional gathering of data on natural and intangible capital and the value of goods and services from them as well as existence value (in the case of preserving ecosystems for purposes other than economic benefit); development and refinement of methods for quantifying non-economic value will be necessary for such an endeavour.
the standard model. Note in Table II.1 how discounting according to risk-free long-term interest rates leads to extremely high and sometimes impossible levels of saving, particularly for countries where average real interest rates on government bonds are negative or very low due to various combinations of high inflation and loose monetary policy.

To test the consistency of the standard model, derived savings rate from returns to capital were regressed against actual (observed) or forecasted (via Model 3) savings rate using ordinary least squares with robust estimators as follows, with $s$ representing actual or forecasted savings-output ratio.

$$s_{\text{model}} = \beta s$$

This regression serves as a test of the null hypothesis that there is no correlation between the derived savings-output ratio and the observed or forecast measures. It should be interpreted exclusively as a test of how well savings-output ratios derived via the standard model match with observations and forecasts from empirical evidence, not as an indication of any causal relation between the observed or forecasted values and the derived values. This test is confined to the three measurements of returns to capital in order to avoid implausible (in some instances impossible) outcomes as produced by deriving optimal savings-output ratios from discount rates based on real risk-free interest rates. These savings figures are displayed according to country in the column “Risk-free interest rate” in Table III.1, and mean, forecasted, and derived savings-output ratios according to the three measurements of returns to capital are likewise listed in the three subsequent columns. The estimates of coefficients of sample mean savings with their standard errors are summarized here.

Returns to built capital only:

$$\beta_{\text{mean}} = 1.02399, \text{error} = 0.0812020$$

Returns to all human capital (built and intangible):
The Student’s t-test results all suggest $p$-values extremely close to zero, leading to a rejection of the null hypothesis that there is no relationship between observed savings-output ratios and those derived from the standard model. It is plausible that observed savings behavior follows closely with that implied by returns to built capital, as demonstrated graphically in Chart III.1. The forty-five degree line is shown as a rough approximation of the relationship in the built capital only case since its coefficient is not statistically different from zero according to the regression output. Coefficient estimates from regressing against forecasted savings were very similar and exhibited similarly-low $p$-values, and inferences similar to those for the regressions against observed savings apply to this case as well, with Chart III.2 serving as graphical demonstration.

Returns to built capital only:
\[ \beta_{\text{forecasted}} = 0.995170, \text{ error } = 0.0797258 \]

Returns to all human capital:
\[ \beta_{\text{forecasted}} = 2.10703, \text{ error } = 0.191045 \]

Returns to all capital:
\[ \beta_{\text{forecasted}} = 2.30881, \text{ error } = 0.194475 \]

While a strong correlation between derived savings and actual savings was found regardless of metric for returns to capital, only the relation involving savings derived from returns on built capital tended toward a slope of unity; the other two metrics appear to produce savings twice that of actual savings. This finding implies a pure rate of time preference from 4 to 12 percent per year given 1.5 as the elasticity of inter-temporal substitution, considerably higher than the 1.5 percent rate adopted by Nordhaus (2007). While the evidence from capital compensation time series imply higher dis-
count rates for benefit-cost analysis of climate change, it should be noted that inter-temporal decision-making related to capital investments may not necessarily carry an intergenerational component. Major capital investments are comparable to medium and long-term (five- to thirty-year) government bonds in terms of risk, whereas a time horizon of a century or more is necessary to account for the effects of long-lived greenhouse gases in the atmosphere. Mitigation, however, would have to be implemented within a much shorter time frame to account for long delays in climate response to human activity, so the return rate to built capital may serve as a reasonable basis for discounting near-term benefits and costs of near-term mitigation projects. Lower discount rates are applicable to longer time horizons since higher cumulative risk reduces certainty-equivalent discount rates relative to those expected over shorter time horizons; a hyperbolic or step-declining discounting schedule may be more realistic than that implied in the stochastic Ramsey model (Weitzman, 2001; Gollier and Weitzman, 2010). The evidence, viewed in its situational context, lends credence to assertions that discount rates ought to vary according to circumstance—a standard protocol for discounting leads to lack of context sensitivity—and that seeking any single ‘correct’ discount rate is inadequate for addressing the complexity of climate change mitigation and its ramifications (Price, 2005).

Given the difference in time horizons for climate mitigation investments and most other business decisions, including natural or intangible capital in calculating real rates of capital compensation, could potentially result in a more realistic benchmark for discounting costs and benefits of climate policies. Estimates of returns to all capital stock, which are available only for the year 2000 and thus may not be representative of rates of return in the long run, are produced in Table I.1. With the case for Latvia as an exception, excluding natural capital increases the rate of return by no more than 10 percent since natural capital constitutes less than 10 percent of capital stock for each country analysed (11 percent for Latvia); these modified estimates are higher than actual returns if the proportion of total returns attributable to natural capital is
more than negligible.

Parameters for pure rate of time preference are displayed in Table III.2 based on figures for marginal aversion to inequality produced in Evans and Sezer (2004, 2005). The values for inequality aversion parameters were determined by the progressiveness of income tax structures in the surveyed countries. Higher values for this parameter are plausible, as demonstrated by Hall (1998) and Dasgupta (2008), based on consumer behavior or, in the context of catastrophic events due to climate change, strong aversion to uncertain risks. Since marginal aversion to inequality and elasticity of inter-temporal substitution are equivalent according to the standard model, it is assumed here that households consider income redistribution among current selves to be no different from income redistribution between current and future selves, in spite of arguments to the contrary. Dasgupta (2008) recommended assuming higher levels of marginal aversion to inequality due to very high uncertainty regarding the impact of climate change in the distant future. Even if pure rates of time preference are in fact as high as those implied by returns to built capital, a sufficiently high level of marginal aversion to inequality will raise precautionary saving dramatically, even above unity.

Such results demonstrate an inconsistency in the standard model: addressing inter-generational welfare issues such as global climate change. Apparently, returns on built capital form the basis of aggregate savings decisions according to this author’s tests, whereas using a metric for social returns as the basis for finding an optimal savings-output ratio results in two-fold overestimation of actual savings. Parameter values often cited in the economic evaluation of climate change mitigation imply optimal savings even greater than those derived via the social returns metric. While the predicted values for savings-output ratio overestimated or underestimated the actual figures by wide margins in some cases as indicated in Table III.1 and Charts III.1 and III.2 despite a statistically significant correlation between observed or forecast savings rates and derived savings rates, the exercise suggests that social discount rates for analysing climate policy using a descriptive approach are bounded be-
low by interest rates on risk-free government bonds and bounded above by rates of return to built capital. It can also be argued that preferences for future consumption among future consumers are not equivalent to current consumers' preferences regarding own future consumption, thus warranting a distinction between *intragen*erational and *inter*generational welfare when conducting benefit-cost analyses of long-term climate policy alternatives. If one intends to maintain equity across generations, lower discount rates, possibly with zero pure rate of time preference, can be justified for the intergenerational welfare case (Azar, 1998, pp. 306–09). Negative pure rates of time preference as implied by some results in Table III.2, which is based on lower rates of return to investment, appear consistent with high levels of uncertainty over future climate change. Such situations may occur under conditions of sufficiently high marginal aversion to inequality and sufficiently low rate of return given a reasonable rate of long-run productivity growth. On one hand, one may propose that, since the Ramsey equation and its related expressions are used to derive *optimal* savings-output ratios, the observed figures not be used in such calculations since agents may not be making optimal decisions as reflected by market observables. On the other hand, the consistency of long-term savings trends as suggested by the raw data series can be interpreted as evidence of optimality of savings decisions according to individual preferences; this claim would hold under assumptions of maximizing behavior among agents in the economy. There is ample empirical evidence however, that economic agents do not always exhibit maximizing behavior but rather ‘satisfying’ behavior. Such behavior describes decision-making that leaves them at some threshold of utility at which they are ‘satisfied’, perceiving no need or ability to increase their utility further due factors including but not limited to information constraints. The possibility of sub-optimality of savings based on private returns is open to further investigation.

It is noteworthy to compare our derivations of pure rates of time preference with those calculated according to life expectancy trends in Brent (1993), reported in Table III.2. These life expectancy discount rates (LEDRs) reflect the long-term annual
rate of change in life expectancy. More rapid improvement in life expectancy corresponds to a higher LEDR. Since life expectancies have been increasing in all sample countries, LEDRs will always be positive as long as the trend of increasing life continues for the foreseeable future. With a few exceptions, the pure rates of time preference derived from market observables were considerably higher than those derived from life expectancy trends. When combined with the marginal rates of aversion to inequality proposed in previous studies, optimal savings rates will likely be far higher than those observed among EU KLEMS data.

The findings in this analysis corroborate assertions by Tóth (2000) that discounting, at least in its most widely utilized form, leads to internal inconsistencies in the evaluation of intergenerational welfare impact of climate change, as well as the conclusion by Dasgupta (2008) that intergenerational welfare economics as a subfield cannot adequately address the questions it poses. Savings rates and pure rates of time preference derived from the standard model diverge from those observed empirically. Lower rates of savings in the returns to built capital is the only case that corresponds to higher implied pure rates of time preference. Lower empirically determined pure rates of time preference and empirically determined marginal aversion to inequality together lead to exaggerated and sometimes implausibly high rates of saving. Hyperbolic discounting provides a limited means of reconciling these outcomes, but cannot completely eliminate problems associated with accounting for intergenerational inequity; a positive discount rate inevitably leads to assigning a negligible value to generations beyond some future point in time. It is nevertheless possible to construct a reasonably equitable discounting schedule via a descriptive approach based on the gathered data. Being unable to determine a single internally consistent discount rate, one can argue for discounting

*While this assumption is generally true, counterexamples can potentially be found when taking into account recent life expectancy trends in areas of Africa with high incidence of HIV/AIDS, provided sufficient and reliable data. One would expect LEDRs in these regions to decline as life expectancy stagnates, though the overall effect of increased death rate and decreased life expectancy on social discounting needs to be examined comprehensively, which is beyond the scope of this paper.
the near future according to the rate of return to built capital, the intermediate future according to the rate of return to all market and non-market capital, and the far future according to the risk-free real interest rate such that lower discount rates are applied for benefits and costs that extend across longer time horizons. In the spirit of Gollier and Weitzman (2010), the asymptotic decline of the effective discount rate toward the lowest rate inferred from economic observables permanently reflects the negative shocks from the effects of climate change and greater precautionary savings in response to this climate change. Given the front-loaded costs and back-loaded benefits of implementing aggressive mitigation policy (versus a business-as-usual trajectory), this segmented discounting schedule and other similar models can render costly mitigation in the immediate future economically efficient under the assumption that duration of benefits are sufficiently long and long run discount rate sufficiently low so that the present value of benefits exceeds the present value of costs in the near future.

CONCLUSIONS AND FURTHER STUDY

This study relied on the premise that the present generation is the generation currently alive, and thus all conclusions from our results should be viewed in the context that generations other than the currently-living one have no standing in policy decisions. Nonetheless, it shall be acknowledged that intergenerational equity concerns will persist as long as benefit-cost analysis of climate policy is confined to the consideration that only the currently-living generation can be considered the present generation. Discounting is not likely to be time-consistent in contexts where the time horizon of policy challenges far exceed those experienced by the planners and actors. Since one can argue that assigning any discount rate to intergenerational costs and benefits results in ‘dictatorship’ of the present, it will be worthwhile to conduct a series of calculations treating subsequent future generations as present ones. Future benefits from mitigating climate change appear greater once future generations are provided sufficient standing. Prior policies
favoring aggressive mitigation will be perceived as more attractive by future generations since, from their frames of reference, benefits appear closer in time and some of the costs may have already been incurred by prior generations and are therefore sunk.\(^*\) Providing future generations standing in evaluation of policies with immediate costs and delayed benefits can alleviate the ‘tyranny of the present’ problem that occurs whenever the impact of policies decided by the present generation spans subsequent generations.

The pool of countries surveyed are all at relatively advanced stages of economic development, so to offer a truly global perspective on the intergenerational welfare economics of climate change, the analysis presented here needs to be applied to data from less developed countries. Pure rate of time preferences in these countries could be hypothesized to be higher than rates for developed countries as long as it is assumed that income and standards of living in poorer countries converge to those of richer countries and that economic growth tends to slow as countries become wealthier.\(^\dagger\) It follows that convergence leads to slower rates of growth in income and improvement of living standards, which suggests that costs and benefits of climate change in developing countries should be discounted hyperbolically. In fact, life expectancy discount rates produced in Brent (1993) offer strong indication that social discounting ought to be steeper in less developed countries, namely those exhibiting rapid or increasing growth in life expectancy and disposable income, and shallower in more developed countries. Demographic transition concomitant with economic transition may also lower discount rates over time. Bishai (2004) indicated that because older generations tend to value future utility more than younger generations do, they should exhibit lower pure rate of time preferences than their younger counterparts. Hyperbolic discounting in such cases, however, demands a time horizon suf-

\(^*\) Even without sunk costs or more immediate benefits, one can expect future generations to find mitigation more attractive if economic growth occurs at a declining rate due to diminishing marginal returns to capital deepening, assuming that the Solow model is an accurate reflection of long-run macroeconomic tendencies.

\(^\dagger\) This tendency is characteristic of the Solow growth model, corroborated by empirical evidence in Mankiw, Romer, and Weil (1992).
iciently long to capture both the diminishing growth rates toward a technology-driven steady state and demographic evolution. To maintain consistency between tendencies toward hyperbolic discounting and observed savings behavior, it should be established that households and firms save and invest over time horizons much shorter than the lifetime horizon proposed in the standard model, let alone an intergenerational time horizon.* Comprehensive examination of the interrelationships among discounting, savings, and time horizons should prove relevant not only to economic analysis of climate policy but also any other subject linked to intergenerational risk and equity.

Past macroeconomic performance underlying the data in this study, however, is a weak indicator of economic prospects in the distant future, and the potential impact of climate change adds to such uncertainty. While economic growth since the 1970s has occurred during a period of steady or falling long-run prices of raw materials, changes in natural resource constraints and availability of substitutes through technological innovation will fundamentally alter growth prospects. If one assumes that total factor productivity is the only source of long-run economic growth, changes in total factor productivity should stand in for consumption growth in the Ramsey equation. Future damages from climate change through increased incidence of extreme weather events will likely mean higher depreciation of capital stock, leading to reduced long-run potential economic output. Discount rates estimated from presently observed savings, growth, and rate of return indicators would thus be overestimated relative to discount rates determined according to future market observables. Benefit-cost analysis of current mitigation policy will need to account for expectations of future macroeconomic performance under a wide assortment of climate scenarios, even though high degrees of uncertainty imply unreal-

* Howlett, Ramesh, and Perl (2009) presented five styles of decision-making, all five of which are crucial when considering climate policy or other policies involving high levels of uncertainty and complexity: optimization, satisficing (already mentioned briefly in this paper; spelled ‘satisficing’ in some sources), search, bargaining, and organizational. It may be more accurate to characterize the actors making savings and investment decisions as satisficing as opposed to optimizing as defined in the standard model.
istically high savings rates as optimal under the standard model (Dasgupta, 2008). Anthoff, Tol, and Yohe (2009a) added that the uncertainty over catastrophic effects of climate change in the future would render any figure for the social cost of carbon arbitrary as discount rates become sufficiently low, for which costs and benefits vary more heavily with the state of the economy in the distant future than they would under higher discount rates. This consequence is one significant drawback, at least from the standpoint of making definitive policy, of providing future generations greater standing. With respect to the series of calculations treating future generations as present, the variation in predicted outcomes for future generations given prior policies may be too large to be of relevance in recommending specific courses of action at the present time.

The extent to which non-market or non-economic considerations affect social discounting compared to economic considerations and their definitive impact on social discount rates merit further investigation. In Desjardins (2006), it has been suggested that the goal of environmental policymaking across long time horizons is to maximize happiness of future generations, which may or may not be quantifiable in a way that benefit-cost analysis can readily handle. Although monetizing natural and intangible capital as the World Bank and the WRI had done to a limited extent enables better reflection on long-term considerations pertinent to policies with extended time horizons, such a refinement fails to account for low elasticities of substitution among natural, built, and intangible capital. Non-substitutability among different forms of capital demands special consideration, especially when certain policy alternatives carry the possibility of major irreversible losses that are difficult to monetize. In order to apply constraints to screen such alternatives, one can disaggregate total benefits and costs of implementing a particular policy into economic, environmental, and social components, each discounted separately using discount rates relevant to its respective categories. Saéz and Requena (2007) proposed a method utilizing separate declining discount rates for human and environmental impact, with a slower pace of discounting.
for environmental benefits and costs. Such ‘differentiated discounting’ according to impact categories in a particular policy package is analogous to hyperbolic discounting as a result of aggregation. Benefits and costs discounted at lower rates carry proportionately greater weight in the longer term; hence the effective discount rate for the entire package declines as time from the present increases. In the context of climate change mitigation, longer time horizons could be expected to raise the profile of distant social and environmental benefits discounted at relatively low rates.

Empirical evidence tested and interpreted in this paper illustrate in practice Dasgupta’s theory-based arguments demonstrating the inherent limitations of formal analyses of the economics of complex problems in environmental sustainability. The methodology employed in this paper relies on market observables and is thus inadequate for taking into account benefits and costs not expressed in market signals, let alone those that can at best be monetized on the sole basis of arbitrarily imposed properties. In particular, the neoclassical economic framework is ill-suited for economic analysis of climate change (and sustainability concerns in general) due to its inadequacy in accounting for the interaction of ecological and sociological conditions with economic activities (Kant and Berry, 2005). Moreover, it is possible within the standard model’s neoclassical framework that the path of savings decisions for satisfying sustainability criteria differs from that for maximizing present expected value (Arrow et al., 2004).* Given eventual limits to real global economic output due to energy and environmental resource constraints, a steady-state economic growth rate of zero in the very long run would be reasonable and consistent with zero discounting that maintains intergenerational equity in the most egalitarian sense (Dasgupta, Mäler, and Barrett, 2000). To resolve intergenerational equity issues in developing optimal (or at least sufficient) strategies to mitigate climate change necessitates addressing sus-

---

* Much of the literature on human behavior and climate change lead to the conclusion that the tools of analysis within the neoclassical rational choice paradigm are insufficient for ensuring sustainability (Gowdy, 2008). Rigorous discourse on shifting away from the neoclassical paradigm is outside the scope of this paper, though such a shift remains in consideration given the exercise demonstrating the limitations of rational choice-based models.
tainability criteria specifically, not merely improving the method of discounting. It remains to be determined whether some alternative to the neoclassical evaluative framework could enable overcoming the said limitations of formal analyses of the economics of climate change and intergenerational welfare.

To view all charts and tables, visit http://www.helvidius.org/du.

Bibliography


PROMOTING THE RULE OF LAW IN CHINA: CRIMINAL DEFENSE LAWYERS AND THE POLITICS OF ARTICLE 306

Rush Doshi

Fifty-year-old Ma Guangjun, a defense lawyer from Inner Mongolia, accepted a controversial case he considered unjust—he decided to defend an impotent patient who had been accused of rape. Ma had found seven witnesses to testify for the defense. While witnesses almost never come to court in China, Ma did not want to rely solely on their written statements. “If I made notes and signed them,” he argued, “then the two witnesses who were not able to write could break their word and say the statement was written by me, the lawyer.” Well aware that law enforcement often prosecuted defense lawyers for forging evidence simply to maintain a high conviction rate, the always cautious Ma decided to make his witnesses testify orally in court.

His prudence, however, did him little good. The day after testifying, all seven witnesses were detained by law enforcement, tortured, and forced to change their statements. The witnesses now claimed that Ma had coerced them into lying, accusing him of the precise charge he had sought to avoid. Soon after their “admission,” Ma was detained and charged under the notorious Article 306 of China’s Criminal Law (CL). For over two hundred days, Ma was beaten and tortured for doing little more than defending his client.¹

The repeated abuse of Article 306 by prosecutors in China has largely intimidated the defense bar, rendering China’s criminal justice system dysfunctional. Article 306 is primarily used by prosecutors at the local level. Thus, understanding the reasons for its abuses requires detailed examination of the local incentive structure. I will argue that the abuse of Article 306 is in part a result of the cultural and institutional difficulties posed by the transition

¹ Rush Doshi graduated summa cum laude and Phi Beta Kappa from the Woodrow Wilson School of Public and International Affairs at Princeton University in 2011. He is currently a Fulbright Scholar in China.
from an inquisitorial system, in which the prosecutor was domi-
nant, to an adversarial one that formally subordinates the prosecu-
tor to judges and places them on equal footing with the defense. 
Within this climate of uncertainty, prosecutors are able to use their 
immense power against their adversaries without any real checks. 
As a result, repeal or revision of Article 306 will accomplish little if 
the local incentive structure that leads to abuse and the central poli-
tics that formed the structure are not addressed. This paper will ex-
amine the enabling factors and current incentives behind prosecu-
torial abuse. It will then examine what reformers might learn from 
the politics and history of China’s criminal reforms in the 1990s. It 
will close with a series of structural reforms that could alter the in-
centive structure at the local level and protect defense lawyers from 
suffering harm for merely representing their clients.

THE PROBLEM

At a glance, Article 306 may seem to be an inoffensive rule of 
law provision that criminalizes the subornation of perjury and the 
falsification of evidence. But when one looks more closely at the 
text, it becomes clear that Article 306 was likely written with the 
purpose of putting criminal defense lawyers in severe professional 
jeopardy. It states the following:

If, in criminal proceedings, a defender or agent ad litem destroys 
or forges evidence, helps any of the parties destroy or forge evi-
dence, or coerces the witness or entices him into changing his 
testimony in defiance of the facts or give false testimony, he shall 
be sentenced to fixed-term imprisonment of not more than three 
years or criminal detention; if the circumstances are serious, he 
shall be sentenced to fixed-term imprisonment of not less than 
three years but not more than seven years.

Where a witness’s testimony or other evidence provided, shown 
or quoted by a defender or agent ad litem is inconsistent with 
the facts but is not forged intentionally, it shall not be regarded 
as forgery of evidence.
Because the opening clause of Article 306 singles out defense lawyers and later clauses on criminalizing evidence fabrication and perjury are so vague, there is widespread potential for it to be wielded unfairly against defense lawyers. Additionally, Article 306 is unnecessary. Article 307 of the CL, which already adequately addresses evidence fabrication and perjury, pertains to all participants in the criminal justice system equally rather than only to defense lawyers. It also has both more appropriate punishments and clearer standards. In contrast, at no point in Article 306 or in the mammoth CL and Criminal Procedure Law (CPL) is it stipulated what it means for a lawyer to “help destroy or forge evidence” or “coerce the witness or entice him into changing his testimony.” Such ambiguities have not been resolved through the court process because, as Terrence Halliday and Sida Liu note, China lacks a “satisfactory system of precedent where courts can settle the meanings of the statutory terms.” This has been a boon for the Procuratorate, a prosecutorial body charged with both prosecuting offenders and with supervising all members of the criminal justice system, including judges, police, and the defense. Article 306’s textual ambiguity and the structural impossibility of clarifying statutes through the courts leaves the Procuracy, tasked with enforcing Article 306, enormous leeway in defining the context in which the law can be used. For them, Article 306 is easily used to harass defense lawyers in criminal trials.

The Magnitude and Costs of Abuse

In the last decade, the Procuratorate has indeed taken advantage of the ease with which Article 306 can be used. Between 1996 and 2006, at least three hundred lawyers all across China have been charged under it. It is the favorite weapon of prosecutors—one study of seventy lawyers who had been criminally charged found that nearly 50 percent were charged under 306. No other legal tool was a close second. It is also a large problem for defense lawyers—Professor Zhao Bingzhi at Renmin University has found that 70 to 80 percent of the complaint cases that bar associations receive
from lawyers reference the threat or use of Article 306. Although a large number of lawyers are charged, investigated, or detained using 306—some for periods of years—only a small number are actually convicted. Out of a sample of fifty-six lawyers charged under Article 306, 70 percent were either found not guilty or were not prosecuted in court. This suggests that “evidence fabrication provisions are being used for the purpose of professional retaliation” and that “charges are dropped once the defense attorney is intimidated” and the prosecutor has secured a conviction.

If the purpose of Article 306 is to intimidate criminal defense lawyers, it has succeeded in doing so at a high cost. According to most observers, Article 306 is one of the major factors contributing to the decreasing popularity of criminal defense work. The Legal Daily, a newspaper published by the Ministry of Justice, has noted with concern that the percentage of defendants with representation has decreased dramatically in spite of a market increase in the number of lawyers in China. Six out of ten defendants currently go without representation. In some counties, the number is as high as nine out of ten. For those who do remain in the field, morale is low and the quality of defense work continues to fall. As a result of prosecutorial intimidation, criminal defense lawyers have “adopted highly defensive measures to minimize their exposure to the police and procuracy,” and therefore avoid gathering evidence or vigorously questioning witnesses. Ironically, and perhaps not inadvertently, the CPL both guarantees defendants a right to counsel and contains, in Article 306, a clause that fundamentally undermines that right.

The Local Political Picture

Although Article 306 creates a significant problem plaguing criminal defense law in China, prosecutors have often abused lawyers through other legal avenues, such as charging them with corruption, malpractice, etc. Even without Article 306 in place, it is likely that prosecutors would still find ways to harass their professional adversaries. An understanding of the political and institu-
tional conditions at the local level that both enable and incentivize abuse by the Procuracy is therefore essential. For the sake of conceptual clarity, enabling factors are the institutional and cultural factors that make abuse possible. The potential for abuse, however, does not make it certain. Prosecutors need a reason to take advantage of the institutional and cultural advantages that enable them to abuse. To understand their behavior, an analysis of the current incentives leading prosecutors to abuse their adversaries is worthwhile.

Enabling Factors

Perhaps the most significant enabling factor leading to prosecutorial abuse is the incompleteness of China’s shift from an inquisitorial to an adversarial criminal system. The inquisitorial system, which was derived from the civil law tradition, pursued justice through an extensive pretrial investigation conducted by prosecutors and/or judges. This, in turn, limited the role of lawyers and trials and expanded the discretion of prosecutors. In contrast, the adversarial tradition drawn from common law tries to pursue justice through competitive trials in which the defense and prosecution are equals under an independent judge who serves as umpire. Although both institutions are capable of protecting rights, the inquisitorial system is notoriously prone to abuse in part because the inequality between prosecution and defense makes the trial process a mere formality. Although the 1996 reforms to the CPL were intended to mitigate this power imbalance by shifting away from the inquisitorial model, they have actually exacerbated its effects. This is in large part because the Procuracy was able to successfully sabotage moves toward an adversarial system and retain its power.

China’s inquisitorial system was one in which prosecutors conducted investigations unrestricted by evidentiary rules, a defense bar, or judges. The switch to an adversarial system in the 1996 CPL brought defense lawyers and prosecutors into direct competition in the trial process and gave judges the power to pick winners, putting the two sides into sharper conflict than ever be-
fore. Reforms that put lawyers and prosecutors into direct competition and formal equality failed to relieve prosecutors of their institutional advantages.¹⁸

Foremost among these myriad advantages is the Procuracy’s supervisory power over the judicial system. Indeed, the Procuracy not only “represents the interests of the State and the people,” but also functions as “the State’s legal supervisory organ” tasked with inspecting “the activities of the other participants as a neutral and objective referee.”¹⁹ This places the prosecutor in charge of complaints regarding procedural violations, extending Procuratorate authority over the police, lawyers, courts, detention centers, defendants, witnesses, and evidence.²⁰ While American judges settle pretrial disputes between the prosecution and defense, here the responsibility falls to the Procuracy. The Procuratorate is thus a player and a referee in the criminal process. Current institutional safeguards that seek to avoid the conflict of interest by separating the supervisory and prosecutorial functions of each Procuratorate into two separate departments are ineffective. Officers that prosecute offenders are often close colleagues of those who are tasked with ensuring the fairness of legal proceedings and lateral transfers between departments remain common.²¹ The Procuratorate is functionally an integrated entity without external or internal checks on its supervisory power.

The absence of such checks means that the Procuracy can investigate and initiate cases against lawyers for fraud, coercion, or violation of Article 306 with impunity. From this perspective, the net effect of the CPL was to throw the weaker defense bar into a dangerous competition with the vastly more powerful and heavily entrenched procurator. Because this competition is “taking place in the context of great legal uncertainty, where rules are vague and their meanings unsettled,” the prosecution—as a judicial supervisor—has the ability and power to take unfair advantage of uncertainty, whether it be about evidentiary procedure, proper questioning of a witness, or Article 306.²²

The second major enabling factor is culture. China’s recent inquisitorial system is rooted in a long Confucian tradition.
Confucianism, historically at odds with Chinese and Western legalism, holds truth and social harmony as more important than legal procedure. The inquisitorial system China adopted from the Soviets conformed to this ideal, as the Procuratorate was charged with finding the truth—often through coerced confessions—and the defense was effectively irrelevant. Prosecutors, judges, and the public thus find it difficult to accept the values implicit in an adversarial system, where the judge is an impartial referee between the prosecutor and defense. Instead, they continue to see procedural safeguards as loopholes and legal etiquette, and view defense lawyers best as obstructionist and criminal accomplices at worst.

In this context, the rapidity with which the Procuracy has been made subordinate to the court and equal to the defense has left its professionals feeling they have lost face. As a result, they have fought a “series of petty but highly symbolic skirmishes, such as over whether prosecutors are required to stand up when judges enter the room and whether judges may sit at elevated podiums.”

To add insult to injury, private lawyers often earn several times more than prosecutors, a fact that prosecutors find humiliating. Given this mixture of contempt and envy, the threshold for prosecutors to use Article 306 when it suits their interests is far lower than it otherwise might have been.

**Current Incentives**

While the previous enabling factors make misconduct more likely, they do not explain why prosecutors decide to abuse their position. For that, we need to examine the incentive structure operating at the local level, which pushes prosecutors to harass their professional adversaries. In the current structure, the costs of losing a case are high for prosecutors while the ability of the defense to win cases is rising as a result of the 1996 reforms. There are three major incentives preventing prosecutors from taking defeat lightly. As it stands, prosecutors need to win, and because it is getting harder for them to win, they are motivated to cheat.
First, the passage of the 1995 State Compensation Law (SCL), intended to make local government more accountable by allowing citizens to receive compensation when their rights are violated by the state or when they suffer undue harm, makes losing costly for prosecutors, both literally and figuratively. The law allows those wrongfully accused or detained to sue the Procuracy and police for compensation. Moreover, these cases tarnish the reputation of the prosecutors involved and can trigger an internal review. To avoid this situation, prosecutors have strong incentives to wield Article 306 to eviscerate a defense. A Shenzhen judge explained how these incentives shaped a recent trial by noting that, if the defense triumphed in court, “the Procuracy would have to compensate for the false detention of the subject ... [so] to prove the guilt of the suspect the prosecutor needed to prove that the lawyer had falsified evidence.” While the number of cases filed under the 1995 SCL is unknown, prosecutors fear the prospect of paying compensation and believe claims will rise if they do not maintain a tough line against them.

Secondly, on an institutional level, the pressure on law enforcement to crack down on crime is greater than ever. China’s vast economic transformation and the attendant social disruptions have produced a surge in crime that has increased six-fold since 1979. Since instability threatens the party’s legitimacy, the CCP has publicly made social order a key objective. It has done so by dedicating itself to several high-profile and effectively never-ending campaigns against crime, known as “Strike Hard” campaigns, during which, as even the government-run China Daily notes, “police usually take tougher measures against crime and judicial authorities hand down swifter and harsher penalties.” Since local law enforcement agencies are assessed based on the number of convictions they secure, especially during these campaigns, they have strong incentives to push through as many convictions as possible. As securing convictions in court grows more difficult, the police and public security bureaus face an uncomfortable dilemma: they “will bear much of the blame for failing to curb crimes while at the same time suffering diminished powers to fight the war.” In this light, Article 306 pre-
vents irksome defense lawyers from interfering with institutional objectives.

Finally, individual prosecutors have strong professional incentives to target their opposition, even when compensation is not at risk. This is largely because law enforcement agencies are themselves assessed by their total number of convictions and reward prosecutors who help agency-level quotas. The current bureaucratic structure is one in which pay and promotion are often directly impacted by a prosecutor’s win-loss record, creating strong incentives to avoid losses no matter the details of a case.\textsuperscript{33}

The pressures caused by the 1995 SCL, China’s “Strike Hard” campaigns, and career have mounted on prosecutors while, at the same time, the chance of losing in court has increased, giving prosecutors a greater incentive to circumvent the system altogether. Two main developments toward the rule-of-law have, ironically, led to prosecutorial abuse. First, China’s increasingly independent judges are willing to declare a defendant not guilty on procedural grounds and have given defense lawyers some success in depriving prosecutors and police of techniques upon which they have long relied.\textsuperscript{34} Secondly, the trial is no longer a mere formality. Before reform, judges were part of the investigatory process. Because they were familiar with all of the prosecutor’s evidence, they made decisions before the trial. This changed after the 1996 reforms. As one judge wrote, “I now have little knowledge about the case … As a result, I have to listen to their arguments more seriously and sometimes I intentionally encourage the defense to advance attacks on the prosecution’s case, so as to help me examine the case from a different perspective.”\textsuperscript{35} Although prosecutors have the clear balance of power on their side, an increasingly independent judiciary is giving the defense a chance.

In an environment where defeat is possible, China’s prosecutors “find themselves feeling frustrated, embarrassed, and humiliated” because they are “ill-prepared to face the unprecedented challenges” caused by the 1996 reforms.\textsuperscript{36} They therefore prefer to cheat rather than to play by the rules. A defense lawyer’s exculpatory evidence must have been “forged,” his key witnesses must have
been “enticed,” and the withdrawal of his client’s confession, though extracted under torture, must similarly have been “coerced.” In this way, prosecutors generate convictions by circumventing the judiciary and arresting their opponents, thereby forestalling the slow move toward procedural justice. Moreover, with only a few high-profile uses of Article 306, prosecutors can “lock-in” the gains by creating the impression that defense work is futile, thereby avoiding the need to combat the defense or adjust to the reforms of the 1996 CPL. And, as explained above, prosecutors can get away with this because the current institutional structure and cultural attitude facilitate precisely this sort of behavior. As one researcher puts it, “It is indeed the overall political structure that should be blamed most for the failure of the CPL revision in 1996.”

THE CENTRAL POLITICS OF REFORM

If the current local political and institutional structure can be blamed for the abuse of Article 306 and the failure of reforms, then an investigation into how that structure came about is essential to understanding how to push for further reform.

In the 1990s, the main impetus for legal reform was the self-interest of the CCP. By the early 1990s, the Tiananmen uprisings and the decreasing salience of communist ideology left the CCP without a basis for its legitimacy. To remain in power, CCP elites turned to economic performance as a new justification for their rule, and began to promote the rule-of-law. First, CCP elites believed that the rule-of-law would create stable institutions that protected contracts and encouraged transactions. Second, they hoped that law could supplant the death of communist ideology and ground the state in the Constitution rather than decaying concepts such as “Mao Zedong Thought.” It was within this context that internal criticism from lawyers and academics gained influence. Reformers pushed for revisions to the old legal codes that had failed to provide adequate procedural protections. Law professors, academics, and researchers were the initial drafters of the 1996 CPL and 1997 CL, and, citing human
rights standards as the impetus for legal reform, included the presumption of innocence, participation of defense in all stages of trial, and the inadmissibility of evidence extracted under torture in these new frameworks. In the initial phase of reform, CCP interests drove the process while internal criticism directed it toward liberal purposes.

Although academics exercised a great degree of influence on the initial reform documents, bureaucratic politics determined most of the subsequent law. The 1996 CPL and 1997 CL are filled with tradeoffs, indicating that the disjointed documents seek on the one hand to commit China in a formal way toward the rule of law, while simultaneously using local discretion and law enforcement to prevent reform from going too far. Lacking a coherent set of goals and philosophies, these laws are instead the product of horse trading and political competition “so fractious that it involved almost every major player in the Chinese bureaucracy.”

The Ministry of Public Security (MPS) and the Supreme People’s Procuratorate (SPP) stood to lose most from reform and were successful in watering down the reforms, maintaining their power, and creating obstacles for defense lawyers. They were also successful in delaying the reform process, launched in 1981, for several years. The Party relies on the MPS and SPP for day-to-day social stability, so when protests flared or the economy slowed, it only strengthened the position of hardliners who could then make a convincing case for retaining the machinery of social control in pristine shape and weakening the hands of reformers.

This is not to say that reformists did not have their own advocates in the process. The Ministry of Justice (MOJ), an initial proponent of China’s shift to an adversarial system, is generally opposed to restrictions on lawyers’ autonomy because the Ministry is most relevant when lawyers are relevant. The MOJ opposed the inclusion of Article 306 when it was introduced by the SPP and PSB, and failing in that, successfully secured the addition of a second clause to Article 306 so that unintentional violations would be exempted from prosecution. The increasingly independent National People’s Congress (NPC) was also an ally. During the
1996 reform process, three thousand NPC delegates vociferously objected to a MPS amendment supported by party elites that would make any discharge of a police firearm in the line of duty a legitimate act of self-defense in all cases, regardless of the circumstances. Ultimately, reformers were successful in ushering in a formal shift to an adversarial system, with protections for defendants, a foundation for an increasingly independent judiciary, and a greater role for the defense bar.

As a result of bureaucratic contestation, the 1996 CPL is profoundly ambivalent on the question of reform. While reformers were successful in winning many formal protections for the defendants and securing a greater place in the trial for defense lawyers, they were unable to ensure these new provisions would be implemented. In contrast, hardliners were successful in ensuring that the SPP and PSB remained vastly more powerful than their adversaries, the defense lawyers. Ultimately, the CPL created provisions that allowed the defense lawyers to stand up in the criminal process while ensuring prosecutors had the power to knock them back down by leaving the local political structure intact. The lesson of the reform process is that textual or formal reform which leaves unaffected the power imbalance between prosecutors and defense lawyers is ineffective. For this reason, the reform of China’s criminal laws in the 1990s remains an unfinished project.

RECOMMENDED REFORMS FOR CHINA

If the local political structure is at fault, then the question becomes how to reform it. Reforms can take three main forms. The first type seeks to remove the potential for abuse by revising the legal and textual problems in the language of Article 306. The second type seeks to make abuse less likely by addressing the enabling factors and the structure which permits prosecutorial abuse. The final type seeks to address the current incentives which lead prosecutors to misuse their authority. While legal reforms are unlikely to prove effective and structural ones are unlikely to garner political acceptance, the final type of reforms might be non-threatening to
prosecutors and reduce the likelihood of abuse.

**Ideal Solutions: Technical/Textual Reforms**

On a technical level, the solutions to the dilemma posed by Article 306 seem relatively straightforward. First, the National People’s Congress (NPC) ought to repeal Article 306. Article 307 of the 1996 CPL already sufficiently addresses perjury and the fabrication of evidence by judicial officers and pertains to all participants in the criminal justice system equally. In contrast, Article 306 applies only to defense lawyers, has vaguer standards, and carries heavier penalties.48

Secondly, the NPC could raise evidentiary requirements for convicting lawyers of having fabricated evidence by establishing a *de minimis* requirement so that minor testimonial discrepancies would not justify prosecution.49 Additionally, the NPC might require more than one witness to demonstrate that a lawyer was guilty of engaging in or assisting fabrication. Revisions that reduce prosecutorial discretion over Article 306 by clarifying vague phrases like “entice” and “assist” might also make abuse less likely.

Unfortunately, textual reforms will not be enough to change the status quo. Without an attempt to change the institutional structure and current incentives, legal reform will go nowhere because prosecutors can always find new tools for harassment, like Article 307 of the CL and Article 38 of the CPL. Worse, those who push for textual reforms may actually harm the reform effort. Randal Peerenboom argues that prosecutors might prefer a debate framed in terms of Article 306 rather than one over whether they should be held personally liable for wrongful prosecution or whether their supervisory role should be given to disciplinary committees.50 Pushing for the abolition of Article 306 may thus squander the political capital of reformers.

**Changing the Power Structure: Addressing Enabling Factors**

Meaningful reforms will require increasing the power of oth-
er actors at the local level relative to the prosecution and creating a system of checks on prosecutorial power. In this vein, the first reform might be to strip prosecutors of primary responsibility over Article 306 and transfer it to China’s bar association, the All China Lawyers Association (ACLA). This would reduce local abuse while simultaneously strengthening the professional identity of lawyers. While the ACLA is controlled by the MOJ, this proposal is unlikely to backfire because the MOJ risks irrelevance when lawyers are not autonomous and has thus often combated efforts to restrict lawyers’ autonomy. Alternatively, prosecutors could maintain primary responsibility for Article 306 while the ACLA or courts are given appellate powers to prevent local prosecutorial abuse.

Secondly, if prosecutors retain power of Article 306, reformers should advocate structural provisions that deter its abuse. Creating a separate financial penalty for the false accusation of lawyers under the 1995 SCL might deter use of Article 306 by making it more costly. Holding prosecutors legally culpable and civilly liable for the harassment of defense lawyers would change the cost-benefit analysis for all prosecutors. Finally, an internal reform that orders the withholding of promotions or bonuses for prosecutors who misuse Article 306 might also prove effective.

On an institutional level, reformers should try to divide the Procutorate’s supervisory and investigatory functions so that the institution would not be both the player and referee in the criminal system. Banning transfers between the supervisory and investigatory departments or creating a dual leadership system with one chief from each wing might help solve this problem. Similarly, reforms that incentivize the supervisory wing to take an active role in overseeing its investigatory counterpart might create a check on abuse. While it might be ideal to abolish the supervisory function of the Procurate outright, this is likely to be unrealistic given that its supervisory role is mandated by the Chinese Constitution and Criminal Law. As one author proposes, “a more feasible proposal at this stage is to introduce earlier judicial involvement and to channel the majority of pre-trial procedural issues” away from the Procurate and to the courts. Prosecutorial control over the
pretrial process, the phase in which evidentiary disputes between prosecutors and the defense often emerge, makes defense lawyers exceptionally vulnerable to Article 306 and gives prosecutors the ability to deny defense lawyers fair access to the client or evidence. Giving judges control over the pretrial process would reduce the potential for abuse. The relatively straightforward nature of pretrial disputes means that less experienced judges and recent law school graduates could be assigned with such responsibilities without dramatically increasing the court’s workload.\textsuperscript{57}

Non-Threatening Reforms: Addressing Current Incentives

But the textual reforms outlined above are not enough. While such solutions would be effective, they would not secure the political support of the Procuracy, which stands to lose power. However, there are reforms that would address the current incentives for abuse without directly undermining the Procuracy’s power. These reforms can be thought of as “non-threatening reforms” because they offer the possibility of change without attracting the political opposition of law enforcement.

First, reformers could push for the revision of the 1995 SCL, which holds the Procuracy liable if an individual defendant is found innocent. This law motivates prosecutors to fight harder for conviction even when the evidence points to the defendant’s innocence. If the law were revised to reduce or eliminate prosecutorial liability when the prosecutor ended legal action against an innocent defendant, the pursuit of justice would be aligned with the interests of prosecutors. Additionally, raising awareness about the SCL might reduce the stigma that prosecutors face internally when served with a compensation claim and also decrease the impetus for retaliation against claimants.\textsuperscript{58}

Second, the Procuracy’s compliance with anti-crime campaigns should be assessed not on the number of convictions secured, as is current practice, but rather on the crime rate. The current basis for assessment is both legal and cultural. Bureaucrats believe that the greater the number of convictions, the safer society
is. Thus, they incentivize prosecutors to collect the usual suspects and charge defense lawyers under Article 306 if they threaten the conviction rate. Shifting the assessment criteria to the crime rate will reduce the pressure on prosecutors to pursue meaningless convictions. Finally, the standards for internal promotions, based in large part on a prosecutor’s conviction rate, could be altered so as to be based on qualitative factors and performance reviews. Such reforms would probably garner support for the Procuracy while simultaneously changing the situation for criminal defense lawyers.

While the mere repeal or revision of Article 306 is unlikely to alter the political imbalance on the ground, the aforementioned reforms work toward changing the political structure and incentives that lead to abuse.

CONCLUSION

The promise of the 1996 CPL and 1997 CL was dashed by the realities of politics. Hardliners at the center kept the Procuratorate powerful by using local discretion to undermine reform, thereby creating a system in which prosecutors and defense lawyers are officially but not actually equal. Article 306 is a manifestation of China’s incomplete transition from an inquisitorial to an adversarial system. Defense lawyers are now responsible for clients and may even have a chance of winning, but prosecutors are left with enough power to arrest their adversaries and avoid defeat.

Prosecutorial abuse is enabled by the fact that the Procuratorate is both the referee and player within the criminal process and can therefore arrest lawyers with impunity. Cultural factors, such as a belief amongst prosecutors that they have lost face by being made formally equal to defense lawyers as well as a belief that such lawyers are obstructive, also lower the bar for abuse. But although abuse is possible, it is not certain. Prosecutorial abuse is incentivized because law enforcement agencies face pressure to generate convictions. With individual careers dependent on win-loss records and lawsuits resulting from defeat, the path to abuse is thus incentivized. At the same time, China’s criminal justice reforms are
making it harder for prosecutors to generate convictions as judges are now independent and legal procedure becomes more accepted. Article 306 offers prosecutors a way out of defeat even if the trade-off is a miscarriage of justice.

Textual solutions, such as clarifying vague terms or including de minimis standards, might make a small difference, but it would not be enough. Real reform must address enabling factors like the Procuratorate’s power advantage. Giving bar associations power over Article 306, dividing the Procuratorate’s investigatory and supervisory powers, or establishing penalties for Procuratorate abuse would protect defense lawyers. Such reforms, however, would be challenged by the Procuratorate. Reforms that mitigate current incentives for abuse rather than curb the power of the Procuratorate might be more welcome. Amending the State Compensation Law and assessing law enforcement agency prosecutors on a metric other than their conviction rate would reduce the impetus for abuse.

Unfortunately, the road to substantive reform is a long one. China is now in the midst of a four-year crackdown that began during the Olympics and became semi-permanent after the financial crisis. Chinese elites are deeply concerned about the rise in mass protests, which have increased from 8,700 in 1993 to nearly 83,600 in 2005—the last year figures were made available by the Ministry of Public Security. These factors have created an atmosphere of profound anxiety among CCP elites while simultaneously calling CCP legitimacy into question, thus making reform less likely but more necessary than ever.

Although China has made steps away from reform in the last four years, that does not mean outsiders should despair. Forty years ago, China was a totalitarian state in the grips of the Cultural Revolution with a nonexistent legal community. That China will dramatically change once again in the next forty years is undeniable; how it will change remains the central question. It is in the interest of the world and the Chinese people to steer that change toward judicial reform and more respect for human rights.
Notes


9 Ran, “Criminals,” p.1024.


12 Congressional-Executive China Commission, p. 6.

13 Ibid., p.6.


15 For more information on other forms of abuse and how often they are used, see Fu, “Lawyers,” p.5–6.


18 This is not a uniquely Chinese problem. According to a Crowley Mission report on Mexico, the incompleteness of reform led to a hybrid system that ‘embodied the worst elements of the inquisitorial and adversarial systems’ because continued prosecutorial dominance rendered the formal role for defense lawyers and judges meaningless. The same analysis applies equally well to China.


20 Ibid.

21 Ibid.


25 Peerenboom, China Modernizes, p. 203.


29 Fu, "Lawyers," p. 41.
32 Peerenboom, China Modernizes, p. 202
33 Ibid.
34 Ibid., p. 40.
35 Ibid., p. 31.
37 Ran, "Liberal Moment?" p. 1027.
42 Hecht, Wrongs and Rights, p. 20.
44 Halliday and Liu, pp. 87–88.
47 Hecht, Wrongs and Rights, p. 20.
48 Ran, "Liberal Moment?" p. 1012.
49 Congressional-Executive Committee, p. 8.
51 CECC, p. 9.
52 Ibid.
54 Ran, "Liberal Moment?" p. 1042.
56 Ibid.
57 Ibid.
58 Ibid.


DEPORTATION, DUE PROCESS, AND DEFERENCE: RECENT DEVELOPMENTS IN IMMIGRATION LAW

Sarah Paige

However heinous his crimes, deportation is to him exile, a dreadful punishment, abandoned by the common consent of all civilized peoples. – Judge Learned Hand

The American immigration adjudication system has witnessed profound change in recent years. Starting in the 1980s, a series of legislative and administrative shifts has produced an immigration adjudication system that increasingly mirrors the criminal justice system. This phenomenon, termed the “criminalization of immigration,” reflects a paradigm shift in the discourse on immigration law and policy. Historically focused on the exclusion and removal of the undocumented, immigration law and policy now increasingly prioritize removal of noncitizens with any criminal history. These developments have both revealed and exacerbated internal tensions in the Supreme Court's treatment of immigration law, leading to a moment of unparalleled dissonance between the Supreme Court's doctrine on immigration and the reality of the American immigration enforcement and adjudication systems.

LEGISLATING THE CRIMINALIZATION OF IMMIGRATION

Legislative politics of immigration in the 1980s and 1990s reflected contradictory moral intuitions in American society about immigration. On one hand, American identity has historically been intertwined with the idea of a land of immigrants and the notion that all persons by virtue of their humanity deserve the op-

---

Sarah Paige is a 2011 graduate of Princeton University, where she majored in political theory with a certificate in values and public life. She graduated with honors, and was awarded the 2010 Arthur Liman summer fellowship to support an internship in public interest law.
portunity to seek a better life. On the other hand, deeply rooted and at times racially charged fears have often been expressed about the economic and security concerns associated with immigration. The category of the criminal alien, referring to not only the undocumented but also legal residents with any criminal history, has emerged as a mechanism for navigating the politics of immigration law.

Starting in the mid 1980s, a flurry of legislation, animated by the idea of the criminal alien, increasingly connected the regulation of immigration with the regulation of crime. Despite the increasing overlap between criminal and immigration law, procedural protections in immigration law were continually weakened to look less and less like protections historically associated with the criminal justice system. Stephen Legomsky suggests that this process reflects a “selective, asymmetric … importation … of the criminal justice model into the domain of immigration law.”

In 1986, Congress passed the Immigration Reform and Control Act (IRCA), a piece of legislation emblematic of an emerging trend toward creating more opportunities for admission while emphasizing the removal of criminal aliens. IRCA introduced a new amnesty program for undocumented noncitizens living in the United States while dramatically increasing funding for border enforcement and criminalizing knowingly hiring undocumented workers. Throughout the 1980s, Congress also passed increasingly punitive legislation governing drug policy and frequently built immigration consequences for criminal offenses into the legislation. The 1986 Anti-Drug Abuse Act, for example, tied drug offenses directly to bases of inadmissibility to the United States.

The criminalization of immigration continued with the 1990 Immigration Act (IMMAct). Like IRCA, IMMAct expanded opportunities for immigration by raising the limit on admissions but created harsher penalties for immigration violations and more severe immigration consequences for criminal violations. The Violent Crime Control and Law Enforcement Act of 1994 continued the trend by increasing the criminal penalty “for unlawful reentry after deportation that followed criminal convictions .... and for as-
sisting noncitizens to enter unlawfully,” among other measures.\textsuperscript{10} The Act also increased federal resources allocated to immigration enforcement, particularly enforcement directed toward noncitizens with criminal records.\textsuperscript{11}

**AEDPA AND IIRIRA**

Though harsher immigration enforcement policies and greater overlap between criminal and immigration law marked the decade following IRCA, the Antiterrorism and Effective Death Penalty Act (AEDPA) and the Illegal Immigration Reform and Immigration Responsibility Act (IIRIRA) changed immigration law more expansively and comprehensively than any other legislation in the previous decade.

AEDPA expanded the grounds of deportability primarily by amending the definition of aggravated felony to include certain racketeering, gambling, forgery, bribery, perjury, and obstruction of justice offenses among others. AEDPA dramatically changed the term’s place in immigration law.\textsuperscript{12} IIRIRA expanded the term’s definition even further.\textsuperscript{13} AEDPA and IIRIRA also broadened the reach of the term aggravated felony by defining it in terms of the length of the sentence imposed in a criminal proceeding as well as redefining the term conviction in immigration law. Through this measure, common plea bargain practices such as imposing a suspended sentence were retroactively redefined as convictions for immigration purposes. These convictions could then serve as the basis for an offense’s recategorization as an aggravated felony. Judicial discretion that had been exercised by providing a suspended sentence was effectively revoked. These changes allowed for crimes generally understood as minor crimes and treated as such in state law to be categorized as aggravated felonies in immigration law.

AEDPA and IIRIRA also expanded the grounds of deportability in other ways. AEDPA amended the INA to “[extend] deportation to aliens convicted of a crime of moral turpitude within five years of their date of entry, for which a sentence of one or more years is statutorily permitted.”\textsuperscript{14} IIRIRA created new federal im-
migration crimes such as “knowingly making a false claim of U.S. citizenship .... and failing to cooperate in the execution of one’s removal order.” The retroactive application of these changes meant that many noncitizens who were legally present and had already served their sentences for minor crimes suddenly became deportable. Additionally, AEDPA and IIRIRA expanded the use of detention in immigration proceedings. The legislation attached mandatory detention to a “wide range of crimes including simple drug possession” and expanded the use of detention primarily through redefinitions of aggravated felony since that category triggers mandatory preventive detention.

The 1996 laws changed the avenues of relief available to noncitizens in removal proceedings as well. Most significantly, Congress repealed Section 212(c) of the INA, a provision that had previously allowed immigration judges significant discretion in determining the fairness of deportation in a particular case. Section 212(c) had required immigration judges to “balance the adverse factors indicating the alien’s undesirability as a permanent resident with the social and human considerations presented on his behalf.” The 1996 laws also changed the standard for relief for another commonly invoked provision, Section 244. Under Section 244, the INA had provided relief for noncitizens who could show ten years of continuous presence and good moral character, while proving that deportation “would result in extreme hardship to the alien or a U.S. citizen or permanent resident who is the alien’s spouse, parent, or child.” The 1996 legislation replaced the extreme hardship clause with a more exacting requirement of “exceptional and extremely unusual hardship,” and also precluded any discretionary relief for aggravated felonies.

Subsequent legislation continued this trend toward criminalization of immigration and greater restrictions on procedural protections. The USA Patriot Act, passed on October 26, 2001, increased executive power of deportation and limited requirements of judicial review in cases “that the attorney general had ‘reason to believe’ [a noncitizen] might commit, further, or facilitate acts of terrorism.” The REAL ID Act of 2005 further limited judicial
review of immigration cases\textsuperscript{25} by, among other measures, limiting the availability of habeas corpus claims to noncitizens, restricting those claims to courts of appeals, and not allowing them to be brought into federal district courts.\textsuperscript{26} In addition, the REAL ID Act “expanded the grounds of inadmissibility and deportability for terrorism-related charges, and introduced even more rigorous criteria for asylum cases.”\textsuperscript{27}

The cumulative effect of this wave of legislation was a “revolution in the law of immigration.”\textsuperscript{28} Immigration and criminal law were increasingly intertwined as new criminal sanctions were attached to immigration offenses and new immigration consequences were attached to criminal offenses.

Against a backdrop of anti-crime, anti-drug rhetoric, Congress relied on the morally loaded classification of the criminal alien to negotiate a deep tension between societal fears about increasing immigration and American identity as a nation of immigrants. This category of criminal alien, however, cut across historical categories organizing the discourse on immigration policy according to legal status. Criminal alien at times referred to undocumented citizens and at times to legally present persons with criminal histories, and blurring this boundary often had consequences that legislators did not seem to have fully anticipated. In addition, noncitizens’ disenfranchisement undermined efforts to voice concerns in the face of what was termed a “legislative steamroller,”\textsuperscript{29} attaching harsher penalties to immigration violations and restricting procedural protections in immigration proceedings.

**Administrative Changes and Current Enforcement Policy**

Following the 1996 legislation, a series of shifts in administrative policy affecting both enforcement and adjudication practices furthered the trend of criminalization of immigration and restrictions on procedural protections. Immediately following the passage of the 1996 laws, the “number of INS law enforcement agents grew by 40 percent … [and] the increase in spending also, predictably, produced an increase in immigration related prosecutions.”\textsuperscript{30} How-
ever, the largest shift in enforcement policy followed the Homeland Security Act of 2002, which restructured the government agencies relating to national security under the umbrella organization: the Department of Homeland Security (DHS).\textsuperscript{31}

The Immigration and Naturalization Service (INS) was absorbed by the Department of Homeland Security only six weeks after DHS’s creation, and INS’s functions were divided into three bureaus: the Citizenship and Immigration Services (CIS), Immigration and Customs Enforcement (ICE), and Customs and Border Protection (CBP).\textsuperscript{32} The DHS’s mission, stated in its strategic plan, reflected shifting priorities in enforcement: “We will lead the unified national effort to secure America. We will prevent and deter terrorist attacks and protect against and respond to threats and hazards to the Nation. We will secure our national borders while welcoming lawful immigrants, visitors, and trade.” While the INS was a service with a clearly regulatory purpose, DHS focused on terrorist attacks and “respond[ing] to threats and hazards” while regulating immigration.\textsuperscript{33} Unprecedented sums were appropriated to these new agencies. In fiscal year 2008, ICE and CBP had an operating budget of over $15 billion,\textsuperscript{34} a massive increase in funds compared to the INS budget of just over $3.6 billion in 1998, particularly since the INS had been responsible for the functions of CIS in addition to ICE and CBP.\textsuperscript{35}

Additionally, immigration officials have increasingly worked directly with police, as illustrated by INS beginning to enter civil immigration information into the FBI’s criminal database in 2001.\textsuperscript{36} This collaboration has expanded with DHS’s increasing emphasis under the Obama administration on Secure Communities, a program that “mobilizes local law enforcement agencies’ resources to enforce federal civil immigration law.”\textsuperscript{37} DHS has also enforced immigration regulations with increasing intensity, illustrated by their ambitious ten-year plan created in 2003 termed “Operation Endgame,” calling for removal of all removable noncitizens by 2012,\textsuperscript{38} at the time an estimated twelve million people.\textsuperscript{39}

As enforcement policy was changing, so was adjudication policy. Attorney General Ashcroft’s 2002 “reform” of the Board of
Immigration Appeals (BIA) dramatically changed the procedural protections afforded to noncitizens in immigration proceedings. The BIA hears appeals of immigration judges’ decisions, and prior to the “reform” (also referred to as the “streamlining” of BIA procedures), the Board would presumptively hear appeals in panels of three. The reform provided for adjudication by a single Board member, unless the case fell into one of six categories or was specifically requested in writing by the appellee. The reform also changed guidelines for Board members’ opinions, promoting and in many cases requiring summary affirmances in which a single Board member would affirm an immigration judge’s decision by using predetermined language defined by regulation instead of writing a case specific opinion. “The Attorney General went so far as to authorize individual Board members to dispose of cases through the affirmation without opinion procedure even if there were errors in the immigration judge’s decision below, and even if the Board member did not agree with the reasoning of the decision below.”

These trends in administrative policy have continued under the Obama administration. Janet Napolitano, President Obama’s Secretary of Homeland Security, has prioritized removal of legal residents deportable due to criminal charges. A DHS press release states, “Secretary Napolitano’s announcements … reflect this administration’s continued focus on smart and effective immigration enforcement over the past twenty months – prioritizing the identification and removal of criminal aliens who pose a threat to public safety.”

Criminalized Immigration

The result of the legislation passed throughout the 1980s and 1990s and recent trends in administrative policy was, as Dan Kanstroom notes, a “paradoxical picture of a nation-state that has expanded both the number of people whom it admits and the number of people whom it expels.” As Bruce Western suggests, the politics of crime were connected in new ways to the politics of race and the economic concerns of the 1960s, 1970s, and 1980s.
nderlying anxieties about race and economic insecurity that drove the anti-drug, anti-crime rhetoric of the 1980s at times drove anti-immigrant sentiment as well. These overlapping concerns help explain the insertion of this rhetoric into the immigration debate, a process that resulted in a paradigmatic shift in the government’s approach to the regulation of immigration. Noncitizens in immigration proceedings now experience an immigration adjudication system that functions dramatically differently than any other time in U.S. history. Reorganization of immigration law and policy around the removal of criminal aliens has affected every level of the immigration enforcement and adjudication systems. Rates of detention and deportation, as well as government spending on all levels of the immigration adjudication and enforcement system, are unparalleled.

RATES OF DEPORTATION

Deportation rates have steadily increased since the 1980s, with 18,013 persons deported in 1980, 30,039 in 1990, 188,467 in 2000, and 393,289 in 2009. Following the 1996 legislation the deportation rate increased more than 60 percent, jumping from 69,680 deportations in 1996 to 114,432 in 1997. Crime related deportations have increased even more dramatically. Before 1986, crime-related deportations, “rarely, if ever, reached 2,000 per year.” Following the passage of IRCA in 1986, however, the number of crime related deportations began to rise. There were 4,385 crime related deportations in 1987, 42,014 in 1999, and 88,000 in 2004. This trend only accelerated during the 1990s with the passage of AEDPA and IIRIRA. From 1996 to 2003, there were 1.2 million deportations, 517,861 of which were for criminal violations.

This trend is still accelerating, as Napolitano reported a record 392,862 deportations in 2010, with about half of that number, 195,772, convicted criminals. Even these figures, however, fail to convey the full scale of this phenomenon, given that federal statistics count only orders of removal and not voluntary depar-
tures. As Jennifer Chacon points out, currently “for every noncitizen who receives a formal order of removal, another four depart ‘voluntarily’ as a result of their encounters with the immigration enforcement bureaucracy.”

DHS reports more that 1 million “returns” as opposed to “removals” in eighteen of the twenty-four years for which statistics have been reported.

**IMMIGRATION DETENTION**

Rates of detention have increased along with increased efforts at enforcement and increased eligibility for mandatory detention. Between 1995 and 2007, the average daily population of U.S. immigration detainees increased from about six thousand to more than 27,000, and in 2008 and 2009 an average of 33,000 individuals were detained on any given night. From 1994 to 2008, the overall number of yearly detentions rose from roughly 81,000 to 380,000, and estimates place the number of detentions for 2009 between 380,000 and 442,000 at an annual cost of $1.7 billion.

The government detains noncitizens in Service Processing Centers, Contract Detention Facilities, Intergovernmental Service Agreement Facilities, and in U.S. Bureau of Prisons. Intergovernmental Service Agreement Facilities, generally state or county jails with beds designated for immigration detainees, house the vast majority—almost 67 percent—of detainees. Conditions in immigration detention facilities largely mirror conditions of prisons or jails. Detainees are treated similarly to those convicted of crimes and have reported being “constrained with handcuffs, belly chains, and leg restraints.” Additionally, the immigration detention system is “woefully unregulated” and lacks legal binding standards, “sending a clear message that noncompliance [with standards] carries no real penalty.” As rates of detention have increased dramatically without a corresponding development of capacity and standards for detention, an overburdened detention system has been widely criticized for alleged violations of human rights and failures to adequately respond to claims of abuse.
COSTS IN TERMS OF FAIRNESS

Concerns about whether the immigration adjudication system is meeting basic standards of fairness have increasingly been voiced from within the legal community. For example, Judge Richard Posner has argued, “the adjudication of these cases at the administrative level has fallen below the minimum standards of legal justice,” and in various opinions has suggested immigration courts’ decisions are “arbitrary, unreasoned, irrational, inconsistent, and uninformed.”

The BIA has also been widely criticized for its failure to meet basic standards of justice. The Seventh Circuit alleged that the BIA was “not aware of the most basic facts.” Additionally, “Board decisions in favor of noncitizens fell from 25 percent to 10 percent” immediately following the “streamlining,” raising questions about whether the streamlining had changed the decision-making process in a way that compromised the fundamental fairness of the proceeding.

The costs in terms of fairness associated with the dramatic changes in immigration law and policy are more difficult to measure than absolute numbers of deportations or rates of detention. Fairness is an elusive concept. While some hold that fairness requires individualized review, others assert that fairness requires treating like cases in a similar fashion. These two components can often pull in different directions, one toward allowing greater judicial discretion and the other toward more standardization through more specific rules.

However, even as immigration adjudication procedures have moved steadily away from individualized review by increasing the number of cases subject to expedited removal, restricting opportunities for judicial review, and “streamlining” adjudication procedures, studies of immigration court outcomes show wide disparities in grant rates depending on factors that few believe should determine outcomes. For example, a thorough study of disparities in asylum adjudication conducted by Jaya Ramji-Nogales, Andrew Shoenholtz, and Philip Schrag determined that (1) whether asy-
lum seekers were represented by legal counsel (2) the asylum seeker’s number of dependants, (3) the gender of the judge, and (4) the prior work experience of the judge each had a statistically significant relationship to asylum grant rates, at a 99 percent confidence level.72 Certainly, asylum cases are not representative of all immigration cases, but the dramatic disparities emphasize the costs in terms of fairness associated with recent policy changes, as more and more noncitizens experience an immigration adjudication system that defies expectations for consistency as a component of fairness.

Cumulatively, these changes reflect a fundamental shift in American immigration law and policy. The criminalization of immigration has created new tensions within constitutional immigration law that still relies on a nineteenth century picture of immigration law and policy and is increasingly disconnected from the current reality of the American immigration system.

“A CONSTITUTIONAL ODDITY”: THE LEGAL HISTORY OF THE SUPREME COURT’S TREATMENT OF IMMIGRATION

Constitutional law governing immigration adjudication is an area of law in which, as Anna Law has noted, “it appears, the normal rules do not apply.”73 The Supreme Court is generally called upon to provide meaningful substantive protection of constitutional rights and to justify these protections against critiques that say these protections are counter-majoritarian and anti-democratic. As Ronald Dworkin has explained, “The right method [constitutional scholars] say, is something in between which strikes the right balance between protecting essential individual rights and deferring to popular will. But they do not indicate what the right balance is, or even what kind of scale we should use to find it.”74

These challenging questions about the proper scope of judicial review, the proper interpretive methods, and the proper interpretations, however, have been largely absent from the development of constitutional immigration law. The Court has justified its silence on these questions with reference to the “plenary power”
doctrine\textsuperscript{75} that “confer[s] upon Congress and the Executive Branch the power to regulate immigration without judicial restraint.”\textsuperscript{76} The Supreme Court’s consistent deference in immigration cases sets immigration law apart from any other area of public law in which different justices have historically approached constitutional interpretation in dramatically different ways despite the existence of clearer textual bases for interpretation.\textsuperscript{77}

**Early Immigration Cases and the Plenary Power**

Early Supreme Court opinions provided the foundation for the plenary power doctrine at the heart of the Court’s deferential approach to judicial review of immigration cases. In these cases, the Court relied heavily on extralegal judgments about immigration, an approach that likely resulted from the lack of a significant body of federal immigration law at the time.\textsuperscript{78} Justice Field writes in *Chae Chan Ping v. United States* (1889) that “the powers to declare war, make treaties, suppress insurrections ... are all sovereign powers, restricted in their exercise only by the Constitution itself,” suggesting that though the Constitution may constrain government action, the Court is not empowered to interpret the nature of that constraint.\textsuperscript{79}

Chae Chan Ping’s case arose from his detention at the border, even though he had been a legal resident of the United States, and was framed in terms of exclusion. Justice Field wrote:

If ... the government of the United States, through its legislative department, considers the presence of foreigners of a different race in this country, who will not assimilate with us, to be dangerous to its peace and security, their exclusion is not to be stayed because at the time there are no actual hostilities with the nation of which the foreigners are subjects. The existence of war would render the necessity of the proceeding only more obvious and pressing. The same necessity, in a less pressing degree, may arise when war does not exist, and the same authority which adjudges the necessity in one case must also determine it in the other. In both cases its determination is conclusive upon the judiciary.\textsuperscript{80}
This comparison between the judicial review of immigration adjudication and the judicial review of congressional powers in war helps contextualize Justice Field’s judgment about the scope of interpretive authority. However, Justice Field does not explain why these types of cases merit different interpretive authority or why the facts of *Chae Chan Ping* (1889) make the case analogous to national sovereignty matters in the salient respect that merits exemption from judicial review. *Chae Chan Ping*’s exclusion was based on nationality, and beyond his identity as a Chinese national, the government made no indication that it considered his presence a threat to national sovereignty. The Court’s inchoate explanation of the relationship between national sovereignty, national security, and interpretive authority would typify future holdings in immigration cases.

In the case *Nishimura Ekiu v. United States* (1892), the Court reformulated the rationale for its holdings on interpretive authority in immigration matters. In *Ekiu*, the Court suggests that sovereignty and self-preservation justify judicial interpretive deference. Ekiu arrived by steamship to meet her husband with $22 in cash. She told the immigration inspector that though she did not have her husband’s phone number, they had planned for him to call her at a prearranged hotel. Still, the inspector excluded her on the grounds that she was likely to become a public charge. She applied for habeas corpus, providing evidence to refute the accusation, but that evidence was excluded by the circuit court due to its finding that the administrative decision regarding her likelihood of becoming a public charge “was conclusive upon the judiciary.”

Addressing the question of whether or not the statute violated due process by making the administrative findings of fact conclusive upon the judiciary, Justice Gray wrote, “As to [foreigners who have never been naturalized, nor acquired any domicile or residence within the United States, nor even been admitted into the country pursuant to law], the decisions of executive … officers, acting within powers expressly conferred by Congress, are due process of law.” By claiming that the executive’s decision is due process, Justice Gray suggests that due process is required but judicial
review may not be necessary to ensure due process. Justice Gray’s rhetorical focus on national security and foreign relations throughout the opinion again helps contextualize this interpretation of due process that leaves ambiguous whether the Court is asserting a lack of interpretive authority or a requirement for interpretive deference.

In *Chae Chan Ping* and *Ekiu*, the Court held that federal power to exclude merited deference comparable to that afforded to wartime actions and that due process afforded noncitizens in exclusion proceedings necessarily met constitutional requirements. *Fong Yue Ting v. United States* (1893) extended those holdings to include expulsion (deportation) as well. In *Fong Yue Ting*, three Chinese laborers were arrested for not possessing required certificates of residence proving that they had been living in the United States when the Act of May 5, 1892, was passed. That Act stipulated that a laborer would be deportable without a certificate of residence or proof that he had been unable to secure the certificate and testimony about his residence by "at least one credible white witness."  

Evaluating the three laborers’ challenge to the constitutionality of the white witness requirement, Justice Gray wrote, “The power to exclude aliens and the power to expel them rest upon one foundation, are derived from one source, are supported by the same reasons, and are in truth but parts of one and the same power.” 85 Justice Gray cited the “power ... inherent in sovereignty” asserted in *Ekiu* and the “independence” and “relation to foreign countries” asserted in *Chae Chan Ping* as the relevant reasons, suggesting that these reasons for restricting judicial review in cases involving exclusion were also reasons for judicial deference in cases involving deportation. 86 Justice Gray continued, “The order of deportation is not a punishment for crime. It is not a banishment, in the sense in which that word is often applied to the expulsion of a citizen from his country by way of punishment.” 87

Justice Gray’s emphasis that deportation is not punishment illustrates the tension created by his argument for judicial deference. Though he argues for extreme judicial deference and possibly that the government should have final interpretive authority in this
case, he nonetheless attempts to provide a constitutional interpretation justifying the government’s action—in this case basing that interpretation of the requirements of due process on the argument that deportation is not punishment. This holding still powerfully shapes constitutional jurisprudence on immigration today.

Taken together, these opinions “can be viewed as the three basic building blocks of the ‘plenary’ congressional power over immigration. The Chinese Exclusion Case recognized an inherent federal power to exclude noncitizens; Ekiu appeared to reject due process limits on the exercise of that power, and Fong Yue Ting extended the principle of both cases from exclusion to deportation.” 88 Additionally, in Chae Chan Ping, the Court connected judgments about immigration to an argument about interpretive authority, in Ekiu that argument was for judicial interpretive deference, and in Fong Yue Ting the reasons given for restricting judicial authority or for interpretive deference were applied in a case involving deportation rather than exclusion. In these cases, the Court relies on a characterization of immigration as raising the same concerns as sovereignty or foreign relations, rather than on statutory interpretation, to justify its deference. The opinions, however, never fully articulate why these concerns should exempt these cases from judicial review or change the Court’s interpretive method. Similarly, the Court never fully articulates why immigration proceedings should not be categorized as punishment. Still, these judgments about the nature of immigration and immigration law have remained remarkably durable features of the Court’s jurisprudence on immigration, even as trends in immigration law and policy have undermined their plausibility, particularly the plausibility of the assertion that deportation is not punishment.

Due Process Protections in Early Immigration Law

Though Fong Yue Ting v. United States (1893) provided a cornerstone for the plenary power doctrine that justified “extraordinary judicial deference,” 89 the three dissents in that case also laid the foundation for the subsequently accepted proposition that non-
citizens should be afforded due process rights. In Justice Brewer’s dissent, he emphasized that “whatever rights a resident alien might have in any other nation, here he is within the express protection of the Constitution.”90 He further emphasizes that “deportation is punishment,” and that “no person who has once come within the protection of the Constitution can be punished without a trial.”91 Additionally, Justice Field, the author of *Chae Chan Ping* and *Ekiu*, wrote:

> The moment any human being from a country at peace with us comes within the jurisdiction of the United States, with their consent – and such consent will always be implied when not expressly withheld … he becomes subject to all their laws, is amenable to their punishment and entitled to their protection.92

All three dissents argued for the importance of affording constitutional protections to noncitizen residents in the United States, implicitly challenging the majority’s ruling that these constitutional protections could be sufficiently guaranteed without judicial review by the fact of their dissent. Justice Brewer described the argument for constitutional protections for noncitizens, particularly due process protections, most clearly when he wrote:

> The Fourteenth Amendment to the Constitution is not confined to the protection of citizens. It says: ‘Nor shall any State deprive any person of life, liberty, or property without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws. These provisions are universal in their application to all persons within the territorial jurisdiction, without regard to any differences of race, of color, or of nationality; and the equal protection of the laws is a pledge of the protection of equal laws … If the use of the word “person” in the Fourteenth Amendment protects all individuals lawfully within the State, the use of the same word “person” in the Fifth must be equally comprehensive, and secures to all persons lawfully within the territory of the United States the protection named therein.’93

These Justices’ dissents suggest that their willingness to assert judicially reviewable constitutional rights of Chinese immigrants was based on an understanding of the centrality of protecting due process to maintaining meaningful protection of rights under law, as
well as concern with unchecked government power and the implications of such extreme judicial deference in terms of rule of law.

_Yamataya v. Fisher (1903)_

The due process protections for noncitizens asserted in these dissents were later affirmed in _Yamataya v. Fisher_ (1903), though the scope of these protections was still strongly influenced by the views articulated in _Chae Chan Ping, Ekiu_, and _Fong Yue Ting_. In _Yamataya v. Fisher_, Justice Harlan wrote, “this court has never held, nor must we now be understood as holding, that administrative officers, when executing the provisions of a statute involving the liberty of persons, may disregard the fundamental principles that inhere in ‘due process of law’ as understood at the time of the adoption of the Constitution.”

Harlan continued, however, that the due process protection required, “not necessarily an opportunity upon a regular, set occasion, and according to the forms of judicial procedure, but one that will … be appropriate to the nature of the case.” Harlan even specified that due process did not require that appellant understand the procedure.

Harlan’s ruling would foreshadow the Court’s treatment of noncitizens’ due process protections in immigration proceedings. The Court has shown strong and consistent support for the noncitizen due process protections in cases such as _Wong Wing v. United States_ (1896), which guaranteed that noncitizens may not be subject to criminal punishment without due process protections, and _Zadvydas v. Davis_ (2001), which asserts that indefinite detention of irremovable admitted aliens violates due process requirements. However, like Harlan’s opinion in _Yamataya v. Fisher_, despite asserting due process protections, the Court has defined the requirements of due process so narrowly that it raises the question of whether the Court’s interpretive deference crossed a line into a functional abdication of interpretive authority. This interpretation highlights the indeterminacy inherent in due process and suggests how this indeterminacy allows for interpretations of due process that can betray the spirit of rights protection.
The indeterminacy of the phrase combined with the Court’s extreme interpretive deference have allowed for a tension to develop between the Court’s deferential treatment of due process claims and an immigration law that increasingly merits stronger procedural protections according to standards the Court has developed in other areas of law. The Supreme Court has explained the inadequacies of these protections by pointing out both that noncitizens may be treated differently from citizens and that immigration proceedings present a unique set of concerns that merit different procedural protections than would, for example, a criminal proceeding.

In *Demore v. Kim* (2003), the Court decided whether allowing mandatory preventive detention for a wide range of noncitizens in immigration proceedings violated due process. Justice Rehnquist wrote for the majority, upholding the constitutionality of the detention and noting, “this Court has firmly and repeatedly endorsed the proposition that Congress may make rules as to aliens that would be unacceptable if applied to citizens.” Though the Court suggested in *Demore v. Kim* that different protections should be afforded based on the identity of the person claiming due process protection, the Court has more often justified its narrow interpretation of the procedural protections owed noncitizens with arguments about why different protections are due.

The holding in *Ekiu* that any procedure dictated by the government would necessarily constitute due process seems to have formed the baseline for the Court’s evaluation of the procedural protections “due” to noncitizens in immigration proceedings. In *Zadvydas v. Davis*, Justice Clark noted, “the Constitution may well preclude granting an administrative body the unreviewable authority to make determinations implicating fundamental rights.” But against what standard the courts may review that authority has been left largely undefined, despite the admittedly fundamental rights of personal freedom at stake.

In determining the requirements of due process in immigration proceedings, the Court has repeatedly deferred to the legislature. The Court demonstrated this deference when addressing the question of whether deportation proceedings must meet the
requirements of the Administrative Procedure Act in order to meet the required constitutional standard. In 1950, Justice Jackson wrote in *Wong Yang Sung v. McGrath*:

> The Administrative Procedure Act ... does cover deportation proceedings conducted by the Immigration Service ... [since] it might be difficult to justify as measuring up to constitutional standards of impartiality a hearing tribunal for deportation proceedings the like of which has been condemned by Congress [in the APA] as unfair even where less vital matters of property rights are at stake.  

Justice Jackson also pointed to the particular importance of procedural protections in cases of "voteless...litigants ... [lacking] the influence of citizens ... strangers to the laws and customs in which they find themselves...and who often do not even understand the tongue in which they are accused." The Court held that the APA should apply.

However, in 1951, Congress included a provision in the Supplemental Appropriation Act that the relevant sections of the APA should not govern deportation proceedings. When asked to determine whether Congress had implicitly reinstated the APA requirement with the passage of the INA in 1952, the Court held in *Marcello v. Bonds* (1955) that "the [INA] expressly supersedes the hearing provisions of that Act." The Court dismissed the contention that the procedural protections provided would not meet the requirements of due process, writing "[this] contention is without substance when considered against the long-standing practice in deportation proceedings, judicially approved in numerous decisions in the federal court, and against the special considerations applicable to deportation which Congress may take into account in exercising its particularly broad discretion in immigration matters." Kanstroom suggests that "the Court, at the height of the Cold War, had simply lost the strong 'rule of law' spirit that had been engendered by the 1946 APA ... [and] relegated [immigration law] to extra-constitutional status." Now, although deportation hearings now conform to many APA requirements, compliance is not constitutionally required.
More recently, the Court has indicated some willingness to assert interpretive authority over immigration cases, pointing out that immigration provisions are subject to constitutional limits. However, the few cases in which the Court has ruled in favor of noncitizen appellants have been primarily cases involving statutory interpretation rather than protections of constitutional rights.

In *INS v. St Cyr* (2001), the Supreme Court held that noncitizens who would have been eligible to seek 212(c) relief had they not taken plea agreements, following the passage of IIRIRA in 1996, could still apply. However, Justice Stevens, writing for the majority, did not specifically invoke noncitizens' due process protections. Rather than focusing on the constitutional guarantee against ex post facto criminal penalties, the Court emphasized case law providing for a “presumption against retroactivity”\(^{108}\) and a consideration of legislative intent. Justice Stevens points out, “We find nothing in IIRIRA unmistakably indicating that Congress considered the question whether to apply its repeal of § 212(c) retroactively to such aliens.”\(^{109}\)

Similarly, the Court held in *Carachuri-Rosendo v. Holder* (2010) that a noncitizen convicted twice in state court for non-felony simple possession should not be categorized as an aggravated felon and be precluded from seeking discretionary relief.\(^{110}\) Justice Stevens wrote for the majority:

> As the text and structure of the relevant statutory provisions demonstrate, the defendant must also have been actually convicted of a crime that is itself punishable as a felony under federal law. The mere possibility that the defendant’s conduct, coupled with facts outside of the record of conviction, could have authorized a felony conviction under federal law is insufficient to satisfy the statutory command that a noncitizen be convicted of an aggravated felony before he loses the opportunity to seek cancellation of removal.\(^{111}\)

Again, rather than relying on interpretation of constitutional requirements, Justice Stevens reasons from “the text and structure of the relevant statutory provisions.”\(^{112}\)

In *Zadvydas v. Davis* (2001), the Court seemed more willing
to recognize noncitizens’ constitutional rights. In an opinion stating that the government may not indefinitely detain admitted noncitizens subject to removal, Justice Breyer emphasized, “the ‘plenary power’ [doctrine]...is subject to important constitutional limitations.”

Despite suggesting its responsibility to uphold noncitizens’ constitutional protections, however, the Court also emphasized that “if Congress has made its intent in the statute clear, ‘we must give effect to that intent.’”

As these opinions illustrate, Supreme Court decisions relating to immigration seem to assert the applicability of the Constitution and at times even assert the Court’s authority to interpret constitutional requirements. However, when those rights may conflict with Congressional intent, the Court has functionally failed to protect noncitizens’ constitutional rights, often not even acknowledging constitutional issues at stake.


In a departure from this long trend of refusal to consider the precepts underlying the plenary power doctrine, the Court gave some indication in *Padilla v. Kentucky* (2010) that it may reconsider the designation of immigration law as civil rather than criminal, an assumption that has persisted since *Fong Yue Ting*. In *Padilla*, the Court considered whether Padilla’s attorney’s mistaken advice that taking a plea in a criminal proceeding would not have immigration consequences constituted a violation of the Sixth Amendment’s guarantee of effective assistance of counsel.

The Court’s decision hinged upon its determination of whether deportation was, “merely a ‘collateral’ consequence of his conviction.” This distinction is meaningful because courts have generally found defendants entitled to constitutional protections of criminal proceedings when a consequence of a conviction is punitive rather than remedial and direct rather than collateral. The Court ultimately ruled that, “constitutionally competent counsel would have advised [Padilla] that his conviction for drug distribution made him subject to automatic deportation.”

Justice Stevens wrote, “deportation as a consequence of a
criminal conviction is, because of its close connection to the criminal process, uniquely difficult to classify as either a direct or a collateral consequence.” As Peter Markowitz explained, “we can understand the Court’s inability to classify deportation as direct or collateral as a proxy for, or at a minimum strongly suggesting, a similar conclusion that deportation is neither purely civil, nor purely criminal.”

The opinion also contained strong language that suggested a willingness to reconsider the civil designation of deportation proceedings. Justice Stevens’ opinion chronicled the progressive overlap between criminal and immigration law as well as the increasingly automatic immigration consequences of criminal convictions. He wrote:

The landscape of federal immigration law has changed dramatically over the last 90 years. While once there was only a narrow class of deportable offenses and judges wielded broad discretionary authority to prevent deportation, immigration reforms over time have expanded the class of deportable offenses and limited the authority of judges to alleviate the harsh consequences of deportation … Although removal proceedings are civil in nature … deportation is nevertheless intimately related to the criminal process … Importantly, recent changes in our immigration law have made removal nearly an automatic result for a broad class of noncitizen offenders. Thus we find it “most difficult” to divorce the penalty from the conviction in the deportation context.

Though the Court’s ruling in Padilla does not directly affect the due process protections afforded noncitizens, it suggests a willingness to reconsider some of the assumptions that have persisted in constitutional immigration law since the early immigration cases, a willingness that is a significant departure from a long history of extreme deference in constitutional immigration law.

DEVELOPMENT OF THE CONSTITUTIONAL DOCTRINE ON IMMIGRATION

Constitutional law governing the immigration adjudication system reflects “two different ideological threads: the one deny-
ing that a society owes aliens any obligation to which it does not consent, the other affirming the existence of certain obligations to aliens owed simply by reason of their humanity.” Since *Yamataya*, the Court has consistently and strongly asserted that noncitizens may claim protection under the Due Process Clause by virtue of their personhood. However, the Court has consistently deferred to Congress in interpreting the scope of noncitizens’ constitutional protections in immigration proceedings, often citing different obligations to noncitizens. Under this doctrine, a tension has developed between the Court’s assertion of constitutional protections and its unwillingness to interpret those constitutional protections in any way that may conflict with Congressional intent in regulating immigration. The result is an approach to interpretation of immigration law best understood as “a constitutional oddity.”

Changes in immigration law and policy since the 1980s have brought to light the implications of such a doctrine, as the assertion that “deportation is not punishment” has become an increasingly unreasonable justification for the Court’s refusal to meaningfully protect noncitizens’ due process claims in immigration proceedings. The Court may argue that since deportation is not punishment, noncitizens’ due process protections are sufficiently protected even despite the Court’s deference. However, the discrepancy between the Court’s interpretation of due process in immigration proceedings and in other cases where increasingly similar interests are at stake undermines this assertion.

As trends toward the criminalization of immigration continue to intensify, the implications of the Supreme Court’s unique deference in constitutional immigration law are increasingly exposed. Most notably, should the Supreme Court acknowledge—as Stevens’ opinion in *Padilla* indicated some willingness to do—that deportation is punishment, the Court would be forced to explain allowing for drastically different procedural protections in admittedly similar circumstances.
EVALUATING ARGUMENTS FOR JUDICIAL DEERENCE IN IMMIGRATION CASES

The proper role of courts in a democratic society has long been the subject of a philosophical debate that reflects deeply entrenched views about what constitutional democracy means. This paper does not attempt to argue for a particular role that courts should play in society or for a particular approach that the Supreme Court should adopt in interpreting the Constitution. Rather, this paper seeks to establish that the fact that a case relates to immigration should neither weaken the argument for judicial review nor fundamentally change the Court’s approach to Constitutional interpretation.

The Court’s arguments for deference in cases related to immigration have relied heavily on certain precepts about immigration law and policy that have remained largely unexamined since their introduction in the nineteenth century. First, constitutional law relating to immigration relies on an argument for deference that at times conflates arguments about interpretive authority with arguments about the correct interpretive approach. That argument characterizes immigration as meriting deference because of its relationship to national sovereignty and foreign relations. Second, constitutional law contends that deportation is not punishment. Recent changes in immigration law and policy, however, have undermined the plausibility of these precepts.

PLENARY POWER AND THE ARGUMENT AGAINST JUDICIAL REVIEW

Early Supreme Court cases framed immigration in terms of national sovereignty and foreign relations to justify the restriction of judicial review. Arguments that the Supreme Court should not have interpretive authority over immigration matters have been widely criticized and implicitly rejected by the Court itself, as it has resolved constitutional questions of immigration matters. Nonetheless, the reasons initially given for restricting judicial re-
view merit evaluation because they are still invoked to explain judicial deference in constitutional immigration law.

The two central arguments for restricting judicial review are immigration law’s connection to foreign relations and to national sovereignty. Th

ough the Court has not clearly articulated what about these cases merits special deference, one account of the foreign relations argument suggests that the Court owes special deference in immigration cases because immigration decisions “operat[e] on the subject of a foreign state,” and the Court should not interfere with government relations with that state through a decision about the immigration status of that state’s subject.

Even taking as a premise the unexplained claim that all cases relating to foreign relations merit special judicial deference, this argument identifies cases that should be shielded from judicial review at once too broadly and too narrowly. Not all cases relating to immigration affect foreign relations. Although the United States’ treatment of refugees arriving at its border may have foreign relations implications, a long-term resident alien deported following a criminal conviction for an offense committed within the territorial United States would not likely raise foreign relations concerns. Additionally, many cases that are subject to judicial review, such as foreign nationals’ criminal cases, do have clear foreign relations implications. Similarly, the Constitution provides for judicial review of “cases arising under treaties, disputes between an American state or its citizens, and even cases affecting ambassadors,” all instances in which foreign relations concerns do arise.

In Hamdi v. Rumsfeld (2004), the Court upheld the due process claims of an American citizen captured in Afghanistan and labeled an enemy combatant. Justice O’Connor wrote, “due process demands that a citizen held in the United States as an enemy combatant be given a meaningful opportunity to contest the factual basis for that detention before a neutral decisionmaker.” In this case, the government claimed that the authority to detain Hamdi under the Authorization for Use of Military Force, which “empower[ed] the President to ‘use all necessary and appropriate force’ against “nations, organizations, or persons” that he determined “planned,
authorized, committed, or aided” in the September 11, 2001 al Qa-
edra terrorist attacks.” Foreign relations were directly implicated
in the case, yet the Court did not seem to approach this case with
special deference.

Though this contradiction suggests that a foreign relations
concern may not be the complete justification for restricting judi-
cial review, one possible response could be that erring on the side of
non-interference with foreign relation cases is preferable to erring
on the side of preserving judicial review despite possible foreign
relations concerns. However, the Court need not categorize all im-
migration cases in the same way. The Court could, as Legomsky
suggests, balance foreign relations concerns in a particular case
against individual rights at issue. This overbroad categorization
of immigration cases as relating to foreign relations is also an in-
creasingly less accurate picture of immigration proceedings as the
number of crime-related immigration cases has risen dramatically
over the past twenty years.

The other argument for restricting judicial review is national
sovereignty. Sovereignty, so the argument goes, “entail[s] the
unlimited power of the nation … to decide whether, under what
conditions, and with what effects it would consent to enter into
a relationship with a stranger.” However, as Gerald Neuman
points out:

One can readily agree that the sovereignty and independence
of the United States would be impaired if other nations could
unilaterally force it to accept or retain their citizens without its
consent. But … in the postwar era, international human rights
norms may impose limits on a nation’s discretion to expel or ex-
clude aliens. Sovereignty is a more relativized concept today and
absolute control over the movement of persons in its territory is
no longer regarded as a necessary ingredient of sovereignty.

Sovereignty is an even less compelling explanation for judicial de-
ference in cases of removal on the grounds of post-entry conduct. In
these cases, the nation has already entered into a relationship with
and assumed certain obligations to a noncitizen. Moreover, even ac-
cepting the unlikely proposition that all immigration cases directly
imply a threat to national sovereignty does not explain why that relationship to sovereignty should preclude judicial review. Judicial review suggests a requirement of adherence to the Constitution and does not necessarily preclude national ability to shape decisions about membership.

Furthermore, as Legomsky points out, the national sovereignty argument, like the foreign relations argument, bases a claim about the proper scope of judicial review on “the existence of an inherent, nonenumerated Congressional power [from which] the plenary power doctrine follows”. The Court’s willingness to accept a nonenumerated Congressional power as the basis for restricting judicial review contradicts the Court’s strong presumption for judicial review, a presumption illustrated by its willingness to review even cases in which Congress has a clear enumerated power.

These two arguments against judicial review center on the dependence of self-determination and national autonomy on a government’s ability to make decisions and take action. However, the arguments largely fail to discuss how judicial review or a particular interpretive approach would undermine the government’s capacity for self-determination. After all, the Supreme Court is an integral part of American government. These arguments seem to suggest an unarticulated argument for a different balance of power between branches of government when certain interests are at stake.

Although constitutional authority for judicial review is at times contested, the authority of the Supreme Court to interpret the requirements of the Constitution has emerged as a definitional feature of American federalism. If immigration merits a departure from this system, the precise boundaries and nature of the reasons for this departure should be clearly articulated. Failing to articulate these reasons seems to disregard the historical and philosophical reasons for affording the Supreme Court judicial review.

ADDITIONAL ARGUMENT FOR JUDICIAL DEFERENCE

In constitutional immigration law, previously discussed arguments for deference tend to invoke the explanations relating to
national sovereignty or foreign relations that were originally used as arguments for restricting interpretive authority. Since immigration is a subset of administrative law, arguments for deference have also been drawn from general arguments for deference invoked in administrative law.

When agency decisions that do not clearly implicate constitutional requirements are at issue, the Chevron doctrine counsels deference to “reasonable agency interpretations of ambiguous statutory provisions, even if the court disagrees with those interpretations.” Since the Executive Office for Immigration Review (EOIR) is an administrative agency under the authority of the Department of Justice, Chevron has been invoked to call for Court deference to agency decisions regarding immigration law.

Chevron’s importance in administrative law derives not only from the Court’s ruling for a fairly strong presumption of deference toward agency interpretations but also from “Justice Stevens’ broad articulation of the reasons judges should defer.” William Eskridge explains:

Most originally [Stevens claimed that] agencies are relatively more legitimate policy-balancers than courts, because the executive branch is more ‘directly accountable to the people.’ Thus, when Congress (the most accountable branch) has not directly addressed the issue, and the agency has filled the statutory gap in a reasonable way, “federal judges—who have no constituency—have a duty to respect legitimate policy choices made by those who do.”

Deference under the Chevron doctrine is also “often defended on the ground that administrative agencies have greater expertise ... than courts.” Even these arguments for deference to agency interpretations are extremely weak in the context of immigration court, particularly given recent changes in immigration law and policy.

The structure of immigration courts and the selection process of immigration judges do not suggest that immigration courts would have greater expertise that should merit deference. Immigration judges were “special inquiry officers” with the INS until their
adjudicatory function was transferred to the EOIR in 1983. They are not members of the judicial branch, nor are they strictly considered administrative law judges, as administrative law judges are required to pass an administrative judge examination to qualify for the position. The Appleseed Network reported in 2009 that 56 percent of immigration judges had a prior job defined as adversarial to immigrants, including “INS/DHS Trial Attorneys, Office of Immigration Litigation attorneys, special Assistant United States Attorneys, border patrol lawyers, and other similar positions,” and another 24 percent of immigration judges had other prior government jobs.

Additionally, a unique feature of immigration court presents a stronger reason against judicial deference in the interpretation of agency decisions. The “supervision of an adjudicative body by a prosecuting official runs counter to the APA philosophy that governs other administrative proceedings.” Yet a scenario in which both the immigration judges and ICE operate under the authority of the DOJ has been allowed in immigration periods. In fact, “immigration judges are appointed by the Attorney General and act under his control and supervision.” Despite the fairness, the APA’s philosophy reflects an understanding of the potential costs of allowing supervision of an adjudicating body by prosecuting officials. Given the potential for abuse in such a situation, judicial deference seems less desirable than in other areas of administrative law. Moreover, the sharp criticism of the immigration courts from the judiciary in recent years has undermined the plausibility of the argument that immigration courts have expertise that merits deference.

Given these strains on arguments for deference to agency decisions, arguments that are consistent across different areas of administrative law, any argument for special deference in constitutional immigration law seems even less plausible.
ARGUMENTS AGAINST JUDICIAL DEFERRENCE IN IMMIGRATION CASES

The consequences of employing such a different approach to the interpretation of constitutional requirements into one area of law raises challenging moral questions. As Kanstroom points out, an overwhelmingly deferential approach to judicial review of enforcement legislation targeting certain racial or ethnic groups illustrates the reasons why judicial review is revered as central to the preservation of American constitutional democracy.  

Though there are many different theories about the proper scope of judicial review and the proper interpretive approach the Court should adopt, one philosophical tradition holds that the Court should have a role in ensuring the protection of minority rights and voice in the political process. This view is reflected in the interpretive approach based on Justice Stone’s famous footnote four in *United States v. Carolene Products Co.*, in which he asked, “whether prejudice against discrete and insular minorities may be a special condition, which tends generally to curtail the operation of those political processes ordinarily to be relied upon to protect minorities, and which may call for a correspondingly more searching judicial inquiry.”  

From such a perspective, noncitizens in immigration proceedings would, if anything, be strong candidates for more, rather than less, searching judicial inquiry, since the lack of the right to vote limits noncitizens’ capacity to correct against discrimination through political processes.

Furthermore, changes in immigration law and policy—particularly the increased use of immigration detention—strengthen the argument that noncitizens in immigration proceedings may be considered a discrete and insular minority. As noncitizens in immigration proceedings are increasingly removed from normal societal interactions through detention, they arguably become more vulnerable to discrimination and less likely to be able to effectively assert their rights. Additionally, Stephen Legomsky suggests that, “aliens’ relative lack of familiarity with the legal system, the language, and the customs” may undermine their ability to participate in the po-
creating precisely the type of circumstance that drives concern for protecting the interests of discrete and insular minorities.

John Hart Ely has advanced the clearest theoretical justifications for what he calls a “representation-reinforcing, participation-oriented approach to judicial review,” similar to the approach outlined in footnote four. Ely argues that the Constitution itself supports this approach and that the approach is “entirely supportive of, the underlying premises of the American system of representative democracy.” This approach also “involves tasks that courts, as experts on process and (more important) as political outsiders, can sensibly claim to be better qualified and situated to perform than political officials.”

While the argument from the text of the Constitution has no special applicability to immigration cases, the argument about the court’s capacity emphasizes why judicial deference is less desirable in immigration cases. As Ely points out, “lawyers are experts in process writ small, the processes by which facts are found and contending parties are allowed to present their claim.” Since procedural rights are central in constitutional immigration law, the Court’s expertise in process may justify judicial review in immigration cases.

One response may be that Ely’s argument was constructed with reference to citizens, and since noncitizens do not have comparable rights to political participation, Ely’s concern with potential restrictions on participatory rights would not apply. However, affording noncitizens due process rights at all reflects recognition that these rights should not be contingent upon political participation.

PUNISHMENT

The Court has historically invoked the claim that “deportation is not punishment” to resolve tensions created by its uniquely deferential approach to constitutional interpretation of immigration cases. Since deportation is not punishment, so the argument goes, noncitizens are not owed the same procedural protections
afforded those in the criminal justice system. That this claim has become less plausible illustrates the dangers associated with employing a uniquely deferential interpretive approach in one area of constitutional law.

Firstly, deportation and immigration detention are experienced as punishment. As Judge Sarokin emphasized in a 1996 opinion:

The legal fiction that deportation following a criminal conviction is not punishment is difficult to reconcile with reality, especially in the context of this case. Mr. Scheidemann entered this country at age twelve; he has lived here for thirty-six years; he has been married to an American citizen for twenty-four years; he has raised three children all of whom are American citizens; his elderly parents are naturalized American citizens; two of his four siblings are naturalized American citizens, and all four of them reside permanently in the United States; he has no ties to Colombia, the country to which he is to be deported; and he has fully served the sentence imposed on him. If deportation under such circumstances is not punishment, it is difficult to envision what is.154

Deported or detained noncitizens suffer significant harm due to the often degrading conditions in immigration detention facilities and the sometimes permanent, forced separation from friends, family, and community. Also, lawful permanent residents often accept harsher consequences in sentencing in criminal proceedings such as longer terms of incarceration or extensions of parole to avoid triggering negative immigration consequences.155 Immigration consequences are often even directly considered by the court during sentencing in criminal proceedings, according to a Boston attorney specializing in immigration law and criminal defense.156

While the severity of the consequences in immigration proceedings leads noncitizens to experience and understand detention and deportation as punishment, severity of harm alone does not constitute punishment.157 As Huge Bedau and Erin Kelly have explained, punishment “in its very conception is now acknowledged to be an inherently retributive practice.”158 Theories of punishment have ranged, as Bedau and Kelly discuss, from “deterrent effects of
punishment … [to] social defense through incarceration [to] retributivism.” Increasingly, American immigration law and policy have assumed these definitional features of punishment. In certain immigration cases, such as those in which a legal permanent resident is placed in removal proceedings following a conviction for an aggravated felony, the punitive intent seems clear.

Kanstroom provides a useful framework for understanding the different intent in different types of immigration cases by distinguishing between extended border control policy and post-entry social control enforcement. While the extended border control model “implements basic features of sovereign power: the control of territory by the state and the legal distinction between citizens and noncitizens,” post-entry social control imposes immigration penalties based on actions of noncitizens once admitted. As the legislation starting in the 1980s significantly accelerated the overlap between immigration and criminal law, the immigration system shifted away from the border control, regulatory model and toward the “post-entry social control” model, a model that strongly implies a punitive rather than remedial purpose for removal.

Certain immigration proceedings have also increasingly taken on a distinguishing philosophical purpose of punishment and, particularly, of incarceration: incapacitation. Incapacitation is present as an element of the current immigration adjudication system in two ways. Most literally, preventive detention, a “staple of the criminal justice system,” is increasingly employed as a mechanism of incapacitation in immigration proceedings over the past twenty years. At a more theoretical level, as Legomsky notes, criminal sentencing should reflect the desired “degree of retribution, deterrence, and incapacitation.” If one argues that immigration consequences are neither retributive nor deterrent, the best characterization of deportation for post entry crimes is as the country’s “ridding [itself] of those noncitizens whose presence is undesirable.” Employing deportation to effectuate this judgment about desirability could be considered a form of incapacitation that operates by removing the offender from society. Finally, the Department of Homeland Security describes noncitizens with criminal records
as “criminal aliens” who are among the “worst of the worst.” The expressive component of this labeling illustrates the retributive intent of immigration policy.

In explaining the conclusion that deportation is not punishment, courts have largely employed circular reasoning, pointing to the designation of immigration law as civil rather than criminal to justify continuing to categorize deportation or immigration detention as not-punishment. Though the argument to consider immigration proceedings as punishment is certainly stronger in some cases than others, in many cases the Court’s argument that deportation is not punishment has become untenable.

DEFERENCE AND CONSTITUTIONALISM

Although the Court gave some indication in Padilla v. Kentucky (2010) of willingness to reconsider the designation of deportation as non-punishment, the Court’s treatment of immigration cases still largely reflects the deference embedded in the plenary power doctrine. Particularly given the changes in immigration law and policy in recent years, the claim that deportation is not a punishment has become a less convincing explanation for the divergence between the Court’s assumptions about immigration and the reality experienced by hundreds of thousands of noncitizens in immigration proceedings each year.

The consequence of this divergence is a doctrine that nominally protects noncitizens’ constitutional claims to due process protections but functionally fails to do so, effectively placing noncitizens who are in immigration proceedings outside the protection of the Constitution. Failure to meaningfully protect constitutional guarantees undermines the United States’ claims of commitment to due process and constitutional democracy. Even though the Court’s own explanations for definitions of due process in other areas of law apply with increasing force to immigration laws, the Court has continued to show striking deference in interpreting the scope of noncitizens’ due process rights in immigration proceedings. For instance, the Court’s explanation for finding a due process right to
government appointed counsel in criminal proceedings in *Gideon v. Wainwright* (1963) also applies to immigration proceedings. Justice Black wrote:

> In our adversary system of criminal justice, any person haled into court, who is too poor to hire a lawyer, cannot be assured a fair trial unless counsel is provided for him … That government hires lawyers to prosecute and defendants who have the money hire lawyers to defend are the strongest indications of the widespread belief that lawyers in criminal courts are necessities, not luxuries.\(^{168}\)

In immigration court, the government hires lawyers to prosecute DHS’s case in an adversarial system, and defendants with the resources consistently hire an attorney. Statistical analyses have demonstrated different outcomes for those afforded legal counsel in immigration proceedings, further suggesting that the concern with fairness that motivated the Court to hold that due process required provision of an attorney in criminal proceedings is also relevant in immigration proceedings.

Furthermore, as Legomsky points out, “the reason for building such stringent procedural safeguards into the criminal justice system is that the consequences of criminal convictions are potentially so severe.”\(^{169}\) Yet, the consequences in immigration court increasingly mirror those in criminal court. Immigration detention has become more common, and most noncitizens in criminal proceedings would accept harsher penalties in criminal court in order to avoid immigration consequences.

The criminal justice and immigration adjudication systems in the United States serve different purposes and may legitimately provide for different procedural protections. However, as the reasons for requiring a certain procedure in one situation increasingly apply to the other, different circumstances seem to be a less compelling explanation for affording such different rights protections.

This tension at the intersection of constitutional and immigration law is ethically problematic because of its implications for constitutionalism as well as the historical and philosophical significance of procedural due process protections.
PROCEDURAL DUE PROCESS

The concept of procedural due process is inextricably linked to rule of law and constitutionalism. Anna Law suggests that, “the idea [of due process] also represents a value that derives directly from the roles and mission of legal institutions.” Putting aside questions of how due process would be interpreted, the idea of procedural protections against inappropriate exercise of government power resonates deeply in American society and ethics, thus reflecting a societal commitment to an ideal of fairness. Though procedural requirements do not guarantee against government taking a certain action, they reflect the understanding that the exercise of power is constrained by law that applies to all persons. For this reason, Larry May argues, “the idea of due process of law is recognized as the cornerstone of domestic legal systems ... Due process rights thus bring deep-seated considerations against the arbitrary exercise of power into some kind of institutional structure, especially one that connects these moral ideas to legal practicalities.”

In American law, procedural due process restricts the government’s ability to deprive persons of “life, liberty, or property” without first adhering to certain procedures designed to ensure that deprivation is not undue. Though the procedural protections due in a particular case may translate into different substantive guarantees in different cases, procedural protections reflect an underlying faith in law as a mechanism for protecting against fundamentally unfair government action. The Court’s interpretation of due process is increasingly explained by the Court’s judgments about the identity of the claimant—a noncitizen in immigration proceedings, for example—rather than by situational concerns associated with those proceedings. Determining that lesser procedural protections would be owed based on the claimant’s identity rather than on situational concerns that may affect the likelihood of ‘undue’ government deprivation of life, liberty, or property contradicts the concern with basic fairness so intertwined with procedural due process.

Failure to meaningfully protect due process rights afforded noncitizens suggests a lesser concern with inappropriate govern-
ment violations of those persons' claims to life, liberty, or property, and even, arguably, a willingness to place these persons outside the law. The unacceptability of functionally stripping a person of rights deemed so important that they should be conferred by personhood is illustrated by Alexander Bickel's observation that, “It has always been easier, it always will be easier, to think of someone as a non-citizen than to decide that he is a non-person, which is the point of the *Dred Scott* case.”¹⁷² Denying legal protections associated with personhood violates deeply held notions about respect for human dignity and what democratic governments owe persons by virtue of their personhood.

**DEPORTATION AND CONSTITUTIONAL DEMOCRACY**

The rhetoric of moral desert in the discourse on immigration policy historically recognized legal status as marking the cleavage between the deserving and the undeserving. Throughout the 1980s and 1990s, however, this paradigm shifted, and documented residents with any criminal history were re-categorized as undeserving. As law increasingly targeted these criminal aliens, a divergence occurred between the procedural protections afforded in immigration proceedings and the similarity of these cases to criminal cases in the salient respects identified as meriting stronger procedural protection.

At the same time, the Supreme Court identified immigration law as different from other types of law. The plenary power doctrine was developed in the late nineteenth century, relying heavily on extra-constitutional arguments about the nature of immigration to justify special treatment of immigration law. Under the plenary power doctrine, early arguments for restricting judicial review were translated into a largely unexplained requirement for judicial deference. In an attempt to reconcile the tension between the claim that the Constitution protected noncitizens' due process rights and the ruling that the Court did not have the authority to interpret the scope of those constitutional requirements, the Court announced that deportation was not punishment. The Court de-
flected arguments of inappropriate deference to Congress and the executive by insisting that constitutional requirements in immigration proceedings were not being violated, no matter how narrowly other branches of government protected those requirements. After all, deportation was not punishment and did not merit comparable procedural protections to those afforded in criminal court. As the landscape of immigration law and policy has changed, however, this explanation has become increasingly implausible. The growing overlap between criminal and immigration law underscores the discrepancy between the Court’s treatment of due process requirements in criminal and immigration cases. In this way, changes in immigration law and policy exacerbate internal tensions in the Court’s interpretation of constitutional requirements in immigration cases.

These developments in immigration law and policy have pragmatic, legal, and moral concerns that are worthy of consideration. On a practical level, the immigration enforcement and adjudication system currently operates at massive cost to the government. Precisely what the goals of those two systems are and whether current policies allow for those systems to fulfill their intended purposes are debatable. In many instances, policies do not effectively fulfill their stated purposes. For example, if concern for public safety is the driving force behind the Obama administration’s focus on apprehending and deporting criminal aliens, the administration should evaluate whether the massive costs associated with the enforcement of immigration laws is the most effective allocation of those funds. That the criminal justice system had already purportedly imposed a just penalty undercuts the idea that all noncitizens processed through the criminal justice system pose a threat to public safety.

Not only do the immigration enforcement and adjudication systems currently operate at massive cost to the government, they also cause massive suffering. Given the severity of the trauma of deportation or immigration detention, reasonable interpretations of the constitutional due process requirements suggest that fairly robust procedural protections must accompany any imposition of
those penalties. If certain procedural protections, such as the right to legal counsel, have been deemed necessary to ensure fundamental fairness in criminal court, failure to provide those protections in immigration proceedings—where comparable interests are at stake—suggests a lack of concern with the fundamental fairness of those proceedings.

A wide range of commentators on the ethics of immigration regulation accept the proposition that the United States has moral obligations to all persons, and particularly strong moral obligations to long-term permanent residents the state has accepted. That obligation, at a minimum, requires that the United States not impose harsh penalties without minimal procedural safeguards to protect against fundamentally unfair imposition of those penalties. Protection of noncitizens in immigration proceedings under the Due Process clause need not yield the same procedural protections in all cases or the same procedural protections as those afforded in criminal courts. However, in at least certain cases the United States has a moral obligation to afford noncitizens stronger procedural protections in immigration proceedings in order to protect the fundamental fairness of those proceedings. Different beliefs about American constitutional democracy may yield different judgments about the acceptability of the Court’s deference. However, even those who believe the Court ought to show greater deference in all cases should accept the claim that the Court ought to give stronger reasons for interpreting one area of law so differently from others, if consistency and coherence are to be valued as conferring institutional legitimacy.

This notion of limits on the exercise of coercive government power is central to the American constitutional democracy, and while reasonable people may disagree about the proper scope of those limits, most accept that procedural protections against government impositions on fundamental interests form a baseline of what is constitutionally required. From this perspective, the Supreme Court, as the branch of government historically charged with interpreting the scope of constitutional limits on the exercise of government power, has a strong duty to at a minimum protect
the procedural rights of the "12.4 million permanent residents and 1.7 million legal temporary migrants" in the United States as of March 2010.

Notes

3. Throughout this thesis, the term "constitutional immigration law" will be used to refer to constitutional law governing immigration adjudication procedures. The term is not intended to encompass all possible constitutional claims that immigrants may advance.
13. Ibid.
20. Ibid., p. 700.
22. Ibid.
35. Ibid.
41. Those six categorizes are listed in 8 C.F.R. §1003.1(e)(6); American Bar Association Commission on Immigration “Report to the House of Delegates.” detentionwatchnetwork.org/files/ABA%20Due%20Process%20Resolution.pdf.
44. Ibid, p. 832.
48. Ibid.
50. Ibid.
51. Ibid.
62. Ibid.
65. Ibid.
68. Kanstroom, Deportation Nation, p. 238.
433.

77. Ibid.


79. Ibid.


82. Ibid.

83. Ibid.

84. Fong Yue Ting v. United States, 149 U.S. 698 (1893).

85. Ibid.

86. Ibid.

87. Ibid.


90. Fong Yue Ting v. United States, 149 U.S. 698 (1893), Brewer Dissent.

91. Ibid.

92. Fong Yue Ting v. United States, 149 U.S. 698 (1893), Field Dissent.

93. Ibid., Brewer Dissent.


95. Ibid.

96. Ibid.

97. Wong Wing v. United States, 163 US 228 (1896).

98. “Indefinite Detention of Immigrant Parolees: An Unconstitutional Condition?”


102. Ibid.


105. Ibid.


107. Ibid.


109. Ibid.


112. Ibid.


114. Ibid.


118. Ibid.

119. Peter L. Markowitz, “Deportation is Different,” Cardozo Legal Studies
192  Paige • Deportation, Due Process, and Deference


125. Fong Yue Ting v. United States, 149 U. S. 698 (1893).
127. Ibid.
130. Ibid.
131. Legomsky, Immigration and the Judiciary, p. 313.
132. Nishimura Ekiu v. United States, 142 U.S. 651 (1892)
136. Ibid., p. 275.
139. Ibid., pp. 1086–87.
143. Ibid., p. 477.
149. Legomsky, Immigration and the Judiciary, p. 300.
151. Ibid.
152. Ibid., p. 88.
153. Ibid., p. 102.
155. Sweeney, “Fact or Fiction,” p. 50.
159. Ibid.
160. Kanstroom, Deportation Nation, p. 5.
161. Ibid.
163. Ibid., p. 519.
164. Ibid.
165. Ibid., p. 520.
166. Sweeney, “Fact or Fiction,” p. 85.
170. Law, The Immigration Battle, p. 189.

Bibliography

Chachón, Jennifer M. “A Diversion of Attention? Immigration Courts and the Adjudication


Moore, Andrew. “Criminal Deportation, Post-Conviction Relief and the Lost Cause of


Tumlin, Karen, Linton Joaquin, and Ranjana Natarajan. "A Broken System: Confidential

U.S. Citizenship and Immigration Service.”Immigration Act of 1990.” http://www.uscis.gov/portal/site/uscis/menuitem.5af9bb95919f35c666f6141765436d1a/?vgnextoid=84ff95c4f635f010VgnVCM1000000ccd190aRCRD&vgnextchannel=b328194d3e88d010VgnVCM10000048f3d6a1RCRD.


POST-MERGER BUNDLING
IN THE CABLE INDUSTRY:
THE CASE OF TIME WARNER AND
TURNER BROADCASTING

Sarah Amanda Levis

The cable industry consists of both upstream and downstream firms: the Cable Television Programming Service Providers (CTPSPs) and the Multichannel Service Operators (MSOs). The CTPSPs create and provide basic cable programming to MSOs in exchange for monthly per-subscriber fees; they also receive funding from advertisers (JRANK Encyclopedia). MSOs purchase franchise rights to operate in specific regions, where they act as monopoly providers of cable, providing subscribers with their choice of programming bundle for a monthly fee. In this paper, I look at the effects upstream bundling may have on subscribers. Gregory Crawford shows that downstream bundling of highly sought-after programming lowers subscriber surplus but raises profits for the cable providers (Crawford 2006). However, a model allowing for differentiated consumer preferences reveals that the price-discriminatory effects of bundling† may in fact raise consumer surplus. Indeed, while Michael Whinston emphasizes that bundling independent goods is an effective and profitable means for deterring entry (Whinston 1990), my model points to a somewhat counterintuitive outcome: product bundling can, at times, facilitate market entry by competitors and improve the welfare of consumers. My approach supports

---

Sarah Amanda Levis is a senior at Yale University, majoring in economics and international studies.

* In a 2005 interview for the Hauser Oral and Video History Collection, former TCI employee Darlene Payne recounts that franchise renewals were not complicated by competition because over-building was costly and ineffective. While many debate whether cable systems should be treated as natural monopolies (see, for example, Bolick 1984), franchise agreements have effectively provided exclusive regional monopolies (Cui et al. 2005).

† The use of bundling to allow for price discrimination between consumers has been widely discussed (Stigler 1968; Adams and Yellen 1976; McAfee et al. 1989). More recently, Gregory Crawford concluded that bundling may provide a way for monopolists to reduce heterogeneity in consumer preferences for various goods and to capture higher profits than they could without bundling (Crawford 2005).
previous research conducted by Whinston that models heterogeneous preferences for one monopoly good that is tied to another monopoly good in a market that faces entry. My contribution is an evaluation of the effect of bundling on entry deterrence in light of heterogeneous preferences for both goods. My model also applies to previous research conducted by Whinston that models heterogeneous preferences for one monopoly good that is tied to another monopoly good in a market that faces entry. My contribution is an evaluation of the effect of bundling on entry deterrence in light of heterogeneous preferences for both goods. My model also applies to a specific case in the cable industry; the Time Warner-Turner merger finalized in September of 1996 (FTC v. Time Warner et al., Complaint, 1997) serves as both an illustration and inspiration for my model.

One recurring debate in the literature on bundling goods focuses on the intersection of leverage theory and tying strategies. The Chicago School did not view tying as a viable strategy for a firm producing both a monopolized and a competitively supplied good. Its main argument was that the monopoly would have no apparent motivation for leveraging its market power over a bundled good when it could extract the same monopoly profit from selling the unbundled components (see Posner 1976 or Bork 1978). Whinston later changes this argument by saying that a similar firm may find tying to be a profitable strategy, as precommitments to tying could allow the firm to monopolize the market in which it currently produces the competitively supplied good (Whinston 1990). In addition to showing that bundling can be used to profitably deter entry, Whinston finds that heterogeneous consumer preferences eliminate the need for precommitment, as bundling may also act as a profitable strategy post-entry. Thus, in many cases, bundling may lead the other entrants to exit the market through strategic foreclosure. In his argument that a precommitment to tying is not necessary, Barry Nalebuff concludes that tying can actually mitigate profit losses in the case of entry (2004); this only holds under certain conditions in my model. For the purposes of my model, the precommitment discussion is not particularly relevant, as I will be assessing the value of a particular court agreement. I assume that Time Warner’s bundling decision always holds. If there were a chance that Time
Warner could change its bundling decision post-entry, then the ban on bundling would be trivial. Assuming that some form of precommitment can be made, results that show that bundling does not always deter entry become interesting.

The model used in this paper is a variation on the Whinston model for the pure bundling* of two independent goods. In section one, I give a brief historical introduction to the merger. In section two, I present the model, the results and limitations of which I describe in section three. In section four, I conclude.

I. CASE BACKGROUND

In 1995, the three largest multichannel system operators, Telecommunications, Inc. (TCI), Time Warner (TW), and Comcast, served approximately 20, 16, and 6 percent of cable subscribers, respectively.† TCI and TW acted as both MSOs and CTPSPs, producing their own programming and providing it to other MSOs (FTC v. Time Warner et al.; Complaint, 1997). Despite high levels of concentration in the cable industry, limitations on anticompetitive behavior, especially among vertically integrated firms, were enforced in the early 1990s. In particular, the 1992 Cable Act prevented MSOs from entering into exclusive contracts with CTPSPs, in which they had a financial interest, and forbade CTPSP price discrimination against the competitors of its own downstream MSO, among other restrictions.

By 1995, rapidly rising cable rates led to a series of attempts to encourage competition in the industry through deregulation; however, the Telecommunications Act of 1996 failed to foster effective competition for the cable firms.‡ Instead, the

---

* Pure bundling means that only the bundle is provided, whereas mixed bundling offers both the bundle and its components independently. When I refer to bundling in this paper, I am referring to "pure bundling" only; I make no distinction between bundling and tying.

† Data compiled from the Encyclopedia for Business ("Cable and other pay television services") and the NCTA website ("Industry Data").

‡ Neither telephone nor satellite companies were able to provide effective competition for cable until much later; the mid-1990s saw cable prices continue to rise, largely uninhibited...
newly deregulated cable industry saw even more acquisitions and higher levels of concentration (Encyclopedia for Business) as prices continued to rise (Consumer Federation of America 2001). It was in this atmosphere that Time Warner announced its decision to merge with Turner Broadcasting.

The Federal Trade Commission (FTC) reacted immediately to the merger, as it posed anticompetitive threats in both horizontal and vertical integration aspects; not only would the TW CTPSP’s incentives be aligned with those of the Turner CTPSP, but the Turner CTPSP would also be vertically joined to the TW MSO. These relationships were further complicated by TCI’s partial ownership (24 percent) of Turner prior to the merger, which translated to an approximately 9 percent non-voting share in new Time Warner.* Additionally, TCI could continue to purchase shares until a poison pill † would finally dilute its ownership after it passed 18 percent (still a silent, or non-voting, share). Nonetheless, the FTC was afraid that the close alignment of the three firms, strengthened by agreements that forced TCI to carry certain Turner programming on its MSOs, would leave TCI with little incentive to compete with Turner broadcasting. This fear, when compounded with the potential for high prices through horizontal integration and the potential for foreclosures induced by vertical integration, led to extensive FTC scrutiny of the merger.‡ Ultimately, a consent agreement was formed that temporarily ended the TCI-Turner carriage agreements, limited TCI’s partial ownership share to 9.2 percent of Time Warner’s fully diluted equity, and placed further stipulations on Time Warner’s business decisions, one of which will

---

* The facts I cite in my discussion of the historical background of the merger come from a case study by the consultants to TCIs attorneys in the FTC proceedings (Besen, Murdoch, O’Brien, Salop, and Woodbury 1999).

† A poison pill is a strategy used by a company to make itself less attractive to an acquirer considering a hostile takeover.

‡ Chipty finds that vertically integrated MSOs often refuse to carry the networks of CTPSPs that compete with their own CTPSPs, e.g., Comcast cable in Philadelphia does not offer the commonly offered network, Fox Sports, but instead offers its own sports program, Comcast SportsNet (Chipty 2001).
be discussed in detail in this paper (FTC v. Time Warner et al., Consent Agreement, 1997).

This paper focuses on part five of the consent agreement, which forbade Time Warner from bundling TW premium cable broadcasts with Turner-affiliated basic programming; specifically, the consent agreement included Time Warner's Home Box Office (HBO) and Turner's Cable News Network (CNN) in the bundling restriction (FTC v. Time Warner et al., Consent Agreement, 1997). At the time of the merger, HBO received the largest amount of subscription revenues of all U.S. programming and was considered by MSOs to be a “marquee” service necessary for acquiring a significant subscriber base (FTC v. Time Warner et al., Complaint, 1997). Although CNN operated as a monopoly in the twenty-four hour news market prior to the Time Warner-Turner merger, the announcement of the merger coincided with a declaration from FOX, a competing CPTSP, that it intended to enter the market by creating the FOX News Channel (Besen et al. 1999). One may interpret the consent agreement's restriction of HBO-CNN bundling as an attempt by the FTC to protect the entry of FOX News. However, my model will explore the possibility that HBO-CNN bundling may, in some cases, have made FOX entry easier than it was under the non-bundling policy. My model also provides some insight into the effect of bundling on consumer surplus.

II. THE MODEL

The following model seeks to reconstruct the relationship between CPTSPs and cable subscribers. Because each MSO operates as a regional monopoly, pays a monthly per-subscriber fee to the CPTSPs, and receives a monthly fee from subscribers, the relationship between firms can be greatly simplified with a few assumptions. First, I look at only those MSOs with no financial investment in either Time Warner or FOX. While these MSOs must pay fixed franchise fees to operate in their
regions, they are able to pass these fees along to all subscribers (Pioneer Communications 2008). I assume that cable subscribers have already paid these fixed fees and they do not, therefore, enter into the subscribers’ preferences for additional cable programming. To simplify the model further, I do not address any other fixed costs that may be encountered by the MSOs. I assume that the MSOs have already paid upfront fees to both Time Warner and FOX, permitting them to purchase any additional programming from either company at the monthly per-subscriber costs.

To assess the effect of upstream bundling on consumer surplus, I assume that the MSOs add a constant markup to the monthly per-subscriber fees that they are charged by the CPTSPs. The MSOs then charge their subscribers this monthly amount (the sum of the markup and the fee which the MSOs must pay the CTPSPs). To simplify the model, I set the markup equal to zero, so that the subscribers’ monthly payments are equal to the MSO monthly payments. In other words, the consumers must pay a monthly fee to the MSOs, who then pass this payment directly along to the CPTSP. Thus, as consumer preferences are reflected directly in the MSO’s purchase decisions, the MSO’s intermediary role is taken out of the model.

For simplicity, I also assume that consumers cannot purchase both FOX News and CNN. Of course, if consumers’ willingness to pay for both channels is greater than the price charged by the MSO, there would be no reason to prevent them from purchasing both channels. However, this simplification suggests a more conservative approach to the model. If consumers could choose both FOX and CNN, bundling would be an even less effective entry deterrent. Therefore, if the model’s results show that bundling can, in some cases, facilitate FOX’s entry, this result is only made more robust by the fact that the model limits consumer choices in this way.

*After the fixed fees of entry have been paid, additional fixed fees do not present major hindrances for the MSOs, as monthly per-subscriber programming fees constitute their greatest expenses (JRANK Encyclopedia).
The model is based on the Whinston model, in which a first firm (TW) operates as a monopoly in two markets (the market for premier movie channels and the twenty-four hour news market), and a second firm (FOX) plans to enter one of these markets (the twenty-four hour news market). The timeline of the model is as follows:

1. Time Warner chooses whether or not to bundle HBO with CNN.

2. FOX chooses whether or not to launch FOX News.

3. Dependent on FOX’s entry decision, either TW or both firms set prices.

My analysis builds upon the Whinston model with two important adjustments that make it more applicable to the Time Warner-Turner case. First, I allow for heterogeneous preferences among consumers for the goods in both markets. Second, I assume that in the twenty-four hour news market, all consumers derive a constant additional utility, \( \partial \), for purchasing FOX News instead of CNN. Marginal costs of production are zero for both firms. Consumers’ preferences, \( \alpha \), for HBO are uniformly distributed from zero to one. Consumers’ preferences, \( \beta \), for CNN are also uniformly distributed from zero to one. Thus, consumers’ preferences for FOX News are uniformly distributed from \( \delta \) to \( 1 + \delta \). If the price of the channel is set lower than a particular consumer’s valuation of it, that consumer will purchase the channel, subject to the assumption that no consumer purchases both CNN and FOX News.

**Time Warner chooses not to bundle**

**FOX does not enter**

By using backward induction, I can see what the optimal prices for each firm would vary according to TW’s bundling decision and FOX’s entry decision. If TW chooses not to bundle
HBO and CNN and FOX chooses not to enter the twenty-four hour news market, then TW will set monopoly prices for both HBO and CNN because it faces no competition in either market. Thus, TW’s profits in both markets as a function of price will be

\[ p \int_p^1 1dp \]

Quantity can be written directly as a function of price because

\[ \int_p^1 1dp \]

captures the number of consumers whose valuation of the channel is higher than the price and who are therefore willing to purchase it. The optimal price in each market is simple to obtain. The first-order condition is given by:

\[
\frac{d[p\int_p^1 1dp]}{dp} = \frac{d[p(1 - p)]}{dp} = 1 - 2p = 0
\]

Thus,

\[ p = 0.5. \]

Consumer surplus in each market is the value that participating consumers derive from their purchase \( v \) minus the price that they pay. Consumer surplus over both markets is simply two times the consumer surplus in each, due to identical prices and preference probability distributions in each market:

\[ 2 \int_{0.5}^1 (v - 0.5)dv = 0.25 \]

FOX does enter

If TW chooses not to bundle and FOX decides to enter the twenty-four hour news market, TW still faces no competition in the premium movie channel market and thus continues to sell HBO at the monopoly price, 0.5. In the twenty-four
hour news market, TW and FOX engage in price competition. If FOX News and CNN were non-differentiated goods, they would compete according to Bertrand competition until prices reached zero, the marginal cost of each firm.* However, because consumers all prefer FOX News to CNN by \( \delta \), the consumer with zero valuation for CNN would be indifferent between paying zero (the marginal cost) for CNN and paying \( \delta \) for FOX News. In FOX’s attempt to capture the market, the additional marginal utility gained from purchasing FOX News mirrors the effect that a lower marginal cost of production for FOX News would have. Instead of FOX and TW both lowering prices until they reach zero, FOX can set its price \( P_{2B} \) equal to \( \delta \) and capture the entire market.

The profit that FOX News makes for capturing the market is:

\[
\delta \int_{0}^{1} d\beta = \delta
\]

Consumer surplus is given by:

\[
\int_{P_{2B} - \delta}^{1} [\beta + \delta - P_{2B}] d\beta + \int_{0.5}^{1} [\alpha - 0.5] d\alpha = \int_{0}^{1} [\beta] d\beta + \int_{0.5}^{1} [\alpha - 0.5] d\alpha = 5/8
\]

By backwards reasoning, if Time Warner chooses not to bundle, FOX will enter the market only if \( \delta > F \), in which case consumer surplus will be \( 5/8 \). If \( \delta < F \), FOX will not enter, and consumer surplus will be 0.25. Table 1 summarizes the results of the no-bundling case.

* Under Bertrand competition for two non-differentiated goods, firms set prices at the higher of the two marginal costs of production, and the firm with the lower marginal cost steals the market.
Time Warner chooses to bundle

FOX does not enter

If TW bundles and FOX does not enter the news market, TW will set its bundle price to maximize profits. To determine what that price will be, I find the probability distribution function of consumer preferences for the CNN-HBO bundle. Thus, I carry out a convolution of the probability distribution functions of consumer preferences for each individual good in the bundle. Since preferences $\alpha$ and $\beta$ are uniformly distributed from 0 to 1, the distribution of the sum of these preferences will range from 0 to 2. The optimal price for the bundle, $P_{AB}$, is equal to $\sqrt{2}/3$. TW’s profits are $\Pi_{TW} = 2/3\sqrt{2}/3$. Consumer surplus is $1 - 8/9\sqrt{2}/3$. The calculations for these results can be found in Appendix A.

FOX does enter

If TW bundles and FOX enters the news market, TW sets its bundle price and FOX simultaneously sets the price of FOX News. To compete with FOX for consumers, the price TW sets for its bundle must be lower than the price it would set without competition. Thus, $0 \leq P_{AB} \leq 1$. TW’s consumers will consist of those who value the bundle more than $P_{AB}$ and who obtain greater surplus from purchasing the CNN-HBO bundle than they do from purchasing FOX News alone. In other words, those who purchase the bundle will have preferences such that $\alpha + \beta \geq P_{AB}$ and $\beta + \alpha - P_{AB} \geq \beta + \delta - P_{B2}$. Those who purchase FOX News instead of the bundle have preferences such that $\beta \geq P_{B2}$ and $\beta + \alpha - P_{AB} \leq \beta + \delta - P_{B2}$. The profit functions for TW and FOX are written in Appendix B, where $f_z(z)$ is the probability distribution function for consumers’ valuations of the bundle (derived in Appendix A). Using the best reaction functions for both firms, each firm’s optimal prices can be found for different values of $\delta$. 
Furthermore, for each set of optimal prices, consumer surplus can be derived. The functions for consumer surplus and each firm's profits are derived in Appendix B.

If, for example, $\delta = 0$, four possible price sets satisfy the best response functions. However, only one of these price sets satisfies the necessary conditions for it to be an optimal price set for both profit-seeking firms. Because $0 \leq P_{AB}^* \leq 1$ and because FOX would not set its price such that $\Pi_{FOX} < 0$, only one equilibrium price set remains, and it yields a profit for FOX of $\geq 0.0487$ and a consumer surplus of $\geq 0.458$. This result is illustrated in Table 2 for the case when $\delta = 0$.

Using Microsoft Excel's Solver add-in, I found optimal prices, profits, and consumer surplus at different values of $\delta$ and compiled this information into two graphs. The graph in Appendix C compares FOX's post-entry profits in the bundling and non-bundling cases. FOX will only choose to enter if $\delta$ is high enough that expected profits under entry are greater than the fixed cost of entry. Thus, the graph in Appendix C can be interpreted as showing the maximum fixed cost, $F$, at which FOX would be willing to enter at different values of $\delta$. The graph in Appendix D shows consumer surplus under each of the four bundling and entry combinations. Table 3 also summarizes the effect of Time Warner's decision to bundle.

Analysis of the Model

Results

The intent of the model analysis is two-fold. First, it is meant to discuss the circumstances under which bundling may facilitate, rather than deter, entry from FOX News. Second, it should explore how bundling can act as an effective price discrimination tool.
Entry Deterrence

The graph in Appendix C makes it clear that for values of $\delta$ below 0.06, FOX would prefer Time Warner to bundle. For low levels of differentiation between FOX News and CNN, entry by FOX can occur at higher fixed entry costs under TW bundling than it can without TW bundling. In the model, heterogeneous valuations mean that TW is not able to price the bundle so that every consumer prefers to purchase it over FOX News. As $\delta$ goes to zero, the extra amount consumers are willing to pay for FOX over CNN in the non-bundling case goes to zero, so FOX profits disappear. Thus, allowing for heterogeneous preferences leads to the result that bundling facilitates entry at very low levels of $\delta$. This conclusion supports Whinston’s findings that bundling can facilitate entry when differentiation between the common goods (in this case, the news channels) is small and preferences for the unique good (in this case, HBO) differ widely across consumers (Whinston 1990). Graphically, the $y$-intercept for the profit curves in the nonbundling case is zero: if CNN and FOX News are completely identical (i.e., $\delta = 0$), FOX’s entry will lead to Bertrand competition until prices are pushed down to the marginal cost of zero, at which point neither firm makes a profit. On the other hand, the $y$-intercept for the profit curve under bundling is greater than zero: even if CNN and FOX News are completely identical, some consumers will place such little value in HBO that they prefer to purchase FOX News alone rather than purchase the bundle. However, as $\delta$ increases, FOX’s profit under nonbundling catches up to and quickly surpasses its profit under bundling. Prices under bundling grow more slowly in $\delta$, as shown in the best response functions (from Appendix B):
\[ P_{AB} = \frac{2P_{B2} - \delta - 3P_{B2}^2 + 4\delta P_{B2} - \delta^2}{1 - 2P_{B2} + \delta} \]

\[ P_{2B} = \frac{-1 + 2P_{AB} + \delta + 3/2P_{AB}^2 + 2P_{AB}^3 - 3/2P_{AB}^2}{1 - 3/2P_{AB}^2} \]

While an increase in \( \delta \) has the direct effect of increasing \( P_{2B} \), it also has the direct effect of decreasing \( P_{AB} \). \( P_{AB} \) has a positive relationship with \( P_{2B} \). Thus, an increase in \( \delta \) has a direct, positive effect on \( P_{2B} \) and an indirect, negative effect on \( P_{2B} \) through \( P_{AB} \). While \( P_{2B} \) still grows in \( \delta \), these two effects counteract each other to some extent, so the price rises at a rate slower than \( \delta \) under bundling. Similarly, an increase in \( \delta \) means that consumers derive greater utility from purchasing FOX but also pay a higher price. Thus, though FOX’s profit under bundling increases in \( \delta \), it rises slower than the linear rate at which FOX’s profit grows in \( \delta \) in the nonbundling case.

For \( \delta > 0.06 \), FOX News can enter at higher fixed costs under non-bundling than under bundling. However, this does not mean that bundling necessarily deters entry at these levels of \( \delta \). Whether or not FOX will enter ultimately depends on the fixed cost, \( F \), as well as the bundling decision.

**Consumer Surplus**

With heterogeneous consumer preferences, bundling can also act as an effective price discrimination tool with important implications for consumer surplus. The simplest way to see this effect is to compare consumer surplus without bundling to consumer surplus with bundling if FOX chooses not to enter in either case. In the no bundling and no entry setting, TW sets monopoly prices, which leads to the lowest possible consumer surplus, as is apparent in the graph in Appendix D. In the bundling and no entry setting, TW is able to maximize its profits at a bundle price that is lower than the sum of the monopoly prices for each good. Nalebuff calls this the “bundling discount” (2004). It is a consequence of the price discriminatory effect of
bundling, which allows TW to capture a greater percentage of both markets at a lower price. This effect increases consumer surplus, as shown in the graph in Appendix D.

To see why this effect exists, consider a consumer who values HBO at \( \alpha = 0.4999 \) and CNN at \( \beta = 1 \). Recall that I am comparing the two simplest cases, in which FOX does not enter no matter what TW’s bundling decision is. If TW decides not to bundle, the consumer will purchase CNN at the monopoly price of 0.5 and receive a surplus of 0.5. If TW bundles, the consumer will purchase the bundle at the optimal bundle price of \( \sqrt{2}/3 \) and receive a surplus of approximately 0.68. In this case, bundling means that the consumer’s surplus rises and that HBO gains an additional subscriber.

Of course, the price discriminatory effect of bundling does not have positive implications for every consumer. Perhaps a different consumer would purchase CNN in the non-bundling case, but does not purchase anything if CNN is bundled with HBO (if for example, this consumer’s valuation of CNN is \( 0.5 \leq \beta < \sqrt{2}/3 \) and valuation of HBO is \( \alpha \leq \beta - \sqrt{2}/3 \)). Yet, when TW faces no competition in either market, this “negative” effect of price discrimination on consumer surplus is not as powerful as the “positive” effect. A consumer willing to purchase either CNN or HBO independently in the non-bundling case is more likely to purchase them both in the bundle at the “bundling discount” than to purchase nothing at all. This result is briefly explained in a more quantitative manner in Appendix E.

If FOX does choose to enter the market, consumer surplus increases no matter which bundling decision TW has made. Competition depresses prices and prevents TW from setting either monopoly prices in both markets or the monopoly bundling price. In the case of entry and no bundling, consumer surplus is constant, as TW sets the monopoly price on HBO, and FOX sets the price \( \delta \) on FOX News. As \( \delta \) increases, the price on FOX rises, but the value consumers derive from purchasing FOX rises equivalently, so consumer surplus is constant in \( \delta \).

In contrast, if TW chooses to bundle and FOX decides
to enter, consumer surplus rises in $\delta$. As $\delta$ grows larger, the bundle price offered by TW is suppressed by competition from the increasingly attractive FOX News network. Similarly, as $\delta$ increases, the price of FOX does not increase by the same amount because of the threat posed by competition with the bundle price (which has been further repressed by the increase in $\delta$). Thus, as $\delta$ rises, the consumers who purchase FOX receive higher levels of surplus caused by increased utility at a price that is held back by competition. Furthermore, when compared to FOX entry without TW bundling, entry with bundling allows consumers to benefit from the bundling discount without suffering the same drawbacks of price discrimination. Those consumers who would purchase CNN independently but purchase nothing at all if it is bundled would enjoy an additional option: instead of an “all or nothing” choice under bundling, entry means that consumers may purchase FOX instead of the bundle.

A summary of the effect of bundling on FOX’s entry decision and consumer surplus is given in Table 4. If the differentiation of FOX and CNN is at either extreme (either very high or very low), the bundling strategy that facilitates FOX’s entry is not the bundling strategy that maximizes consumer surplus if entry occurs. If, for example, CNN and FOX are nearly identical programs (e.g., $\delta = 0.05$), bundling facilitates FOX entry. Yet, if TW chooses not to bundle, consumer surplus is maximized only if FOX enters. If FOX does not enter, then consumer surplus without bundling is lower than consumer surplus with bundling because a no-bundling policy gets rid of the price discriminatory effects. This result reinforces earlier findings by O’Brien and Shaffer: “a policy of prohibiting bundled discounts may lead to ... lower welfare if it fails to induce additional entry” (O’Brien and Shaffer 2005, 575). The decision to bundle eliminates the risk of observing the lowest possible level of consumer surplus. From a comparison of both graphs, it is clear that for the extreme values of $\delta$, a trade-off exists between facilitating entry and maximizing surplus if entry occurs. Thus, the FTC should take into
account that bundling may not necessarily deter entry, and that even in the cases where it makes entry more difficult, bundling may still lead to higher consumer surplus. For the intermediate range of $\delta$, a trade-off exists between the benefits of the non-bundling policy, i.e., facilitating FOX entry and maximizing consumer surplus in the case of entry, and the benefit of the bundling policy, i.e., eliminating the risk of observing the lowest possible level of consumer surplus if entry does not occur.

III. LIMITATIONS

In my presentation of the model, I discuss several simplifications employed to make the model more mathematically viable. In this section, I point out additional assumptions made for the model that limit its realism and applicability. In order to compare the bundling of TW and CNN to the Whinston model for bundling independent goods, the assumption that HBO and CNN are monopolies in their respective markets prior to FOX entry must hold. In the twenty-four hour news market, CNN indeed acted as a monopoly when FOX announced its intention to enter. However, just months prior to FOX's final entry decision, MSNBC entered the twenty-four hour news market. In the premium movie channel market, HBO's role as a monopoly can be debated on the grounds of market definition. While it served a particular niche of consumers in the early 1990s, it is possible to redefine HBO's market such that it has a few small competitors. An interesting extension to the model would be to see how the effects of bundling change if CNN and HBO operate in oligopolies with differentiated goods. Nonetheless, for the purposes of this paper, modeling TW as an oligopoly would only strengthen my hypothesis that bundling does not always deter FOX entry. Without full market power, TW's decision to bundle would act as an even less potent entry deterrent than is found in the model. Thus, the assumption that TW operates as a monopoly in both markets indicates a more conservative approach to the
model. The assumptions made for consumer preferences in the model are more troublesome. Clearly, consumer preferences for HBO and CNN are unlikely to be uniformly distributed. Yet, the assumption that preferences are uniformly distributed and calibrated from zero to one is mathematically clean and provides a nice structure for analysis. Similarly, the assumption that every consumer prefers FOX to CNN by a constant value works as a way of simplifying the model to provide clear intuition but does not hold true in reality. Certain regions within the U.S. may prefer FOX to CNN for sociopolitical reasons, but it is not likely that every consumer would prefer FOX to CNN. It is even less believable that each would prefer FOX to CNN by the same amount. However, because marginal costs are zero (an accurate assumption for the extra cost CTPSPs incur in order to provide programming to one additional region), the model must include some type of differentiation between FOX and CNN in order for the effect of Whinston’s results to be assessed and developed. Otherwise, FOX would never enter in the non-bundling case. Thus, the simplest form of differentiation, in which FOX is always preferred by a constant amount to CNN, is used. While it would be more realistic to incorporate preferences between differentiated goods according to the Hotelling model, a Hotelling model of preferences for differentiated news programs would be much more complex when combined with heterogeneous valuations for HBO under bundling.

Another aspect of consumer surplus that our model does not discuss is content variety. Consider a consumer who is indifferent between purchasing the CNN-HBO bundle and purchasing FOX News alone (at a certain level of \( \delta \)). A more accurate model may describe consumers as deriving an additional utility from purchasing the bundle rather than purchasing FOX alone because the bundle provides the subscriber with more content variety. However, if the model allowed consumers to purchase both FOX News and the bundle, the extra utility derived from content variety would be highest if the consumer purchased content from both FOX and TW. Thus, ignoring the
content variety of utility has ambiguous implications for the entry-deterrent effect of bundling. Perhaps modeling an extra utility from content variety would make bundles more attractive to consumers, but it would also imply that some consumers want to purchase from both FOX and TW, which my model does not allow. Nonetheless, if a consumer does not value an additional channel enough to purchase it without the content variety utility, it seems unlikely that the consumer would purchase it in order to receive the additional utility of having an extra channel. The relationship between content variety and consumer surplus is central to the current debate over à la carte channel pricing.*

IV. CONCLUSION

Though many of the model’s assumptions limit its use, it provides insight into the FTC’s decision to forbid Time Warner from bundling CNN and HBO. In particular, the model reveals that bundling can facilitate entry when differentiation between the tied goods (those in the market that faces entry) is very low. Of course, it is unlikely that FOX News and CNN would be nearly identical programs, as news programs could be differentiated by a focus on international or local news or by an alignment with conservative or liberal politics. The implication of the model is that bundling becomes more helpful for entry as FOX News and CNN become less differentiated. Furthermore, the model reveals a trade-off between facilitating entry and maximizing consumer surplus. Under the fixed cost and δ values at which bundling makes entry more difficult, bundling may also maximize consumer surplus if entry occurs. If entry by FOX is deterred independently of TW’s actions, bundling also maximizes consumer surplus. Thus, if the entry costs faced by FOX appear too high to overcome no matter which bun-

*For example, the Besen case study groups HBO together with Showtime and other movie channels in a premium network market, though none of HBO’s competitors in this market have nearly as large a subscription base as HBO (Besen et al., 1999).
dling decision TW makes, the decision to bundle leads to price
discriminatory effects that ultimately benefit consumers. It is
important to note that advertising benefits, lower marketing
costs, and lower administrative fees usually accompany a decision
to bundle (Crawford 2005, 42), which is another incentive for
bundling not discussed in the model.

These results imply that while the FTC’s decision to for-
bid Time Warner from bundling CNN and HBO may be prima
facie pro-competitive, a policy that forbids bundling may not
be the best thing for consumers. Levels of entry costs and dif-
ferentiation between FOX News and CNN must be assessed
alongside TW’s bundling decision to evaluate both the like-
lihood of entry and the consequences entry has for consumer
surplus. If the FTC fears that bundling HBO with CNN would
deter FOX’s entry, then, by virtue of the fact that the FTC’s
regulations are meant primarily to protect consumers, this fear
should be rooted in the belief that entry increases consumer
surplus. Yet, both the price discriminatory effect of bundling and
the fact that bundling may facilitate entry under certain condi-
tions must be taken into account when considering the effect
of bundling on consumer surplus.

Shortly after the FTC’s decision to forbid the newly
merged Time Warner-Turner enterprise from bundling, Time
Warner’s CNN faced another important turn of events. On Oc-
tober 7, 1996, just weeks after the consent agreement for
the Time Warner-Turner merger was publicized, FOX News
launched (Besen et al. 1999). It is unclear how foreknowledge
that TW would not be allowed to bundle CNN with HBO
factored into FOX’s decision to launch FOX News. The effec-
tiveness of the consent agreement must be viewed in light of the
fact that another CTPSP, the National Broadcasting Compa-
y, launched its own twenty-four hour news program (MSNBC)
in the summer of 1996. When the consent agreement was fi-
nalized, Time Warner no longer had monopoly power in the
twenty-four hour news market because MSNBC had already
entered. Thus, CNN-HBO bundling would not have discou-
aged FOX entry as much as it would have if MSNBC had not already entered the market. Even ignoring the effect of MSNBC’s entrance on FOX’s decision to enter the market, it is far from certain that forbidding bundling will always aid the consumer. Although the model itself cannot be used to accurately predict differences in consumer surplus, it has shown that bundling has the potential to increase consumer surplus. Thus, insofar as the FTC regulates the cable industry to protect consumer’s interests, these results cast doubt on the merit of the FTC’s interference in TW’s bundling decision.

*As a final thought, FOX’s reaction to another stipulation set forth by the consent agreement adds an interesting layer to the case. In addition to barring TW from bundling CNN with HBO, the consent agreement decreed that TW would be forced to offer one other, independently produced twenty-four hour news network to its subscribers by 2001 (FTC v. Time Warner et al., Consent Agreement, 1997). When TW chose to carry MSNBC, FOX sued, stating that it would not be able to survive if it did not receive carriage from TW MSOs in particular regions of the United States (especially the densely populated and lucrative region of New York City). Although TW successfully blocked the suit from going to court, both companies decided to settle. After accepting what was reported to be a very profitable deal from FOX, Time Warner began to offer FOX News to its subscribers. Rennhoff and Serfes provide a good summary of the literature on bundling in oligopolistic markets, where bundling policies tend to lower firms’ profits (Rennhoff and Serfes 2008, 551). However, Stole finds that bundling in an oligopoly setting may still capture price discriminatory effects that make it profitable and may lead to foreclosure of entrants (Stole 2003).
Table 1: TW chooses not to bundle

<table>
<thead>
<tr>
<th></th>
<th>$\delta &lt; F$</th>
<th>$\delta &gt; F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOX’s entry decision</td>
<td>No entry</td>
<td>Entry</td>
</tr>
<tr>
<td>FOX’s profits</td>
<td>0</td>
<td>$\delta - F$</td>
</tr>
<tr>
<td>Time Warner’s profits</td>
<td>$1/2$</td>
<td>$1/4$</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>$1/4$</td>
<td>$5/8$</td>
</tr>
</tbody>
</table>

Table 2: Finding optimal prices from reaction functions

<table>
<thead>
<tr>
<th>$P_{2B}^*$</th>
<th>0.212</th>
<th>0.384</th>
<th>0.582</th>
<th>0.833</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{AB}^*$</td>
<td>0.503</td>
<td>1.40</td>
<td>-0.900</td>
<td>0.624</td>
</tr>
<tr>
<td>$\Pi_{2B}^*$</td>
<td>0.0487</td>
<td>$P_{AB}^*$ should be $\leq 1$</td>
<td>$P_{AB}^*$ should be $\geq 0$</td>
<td>-0.0291</td>
</tr>
<tr>
<td>$\Pi_{AB}^*$</td>
<td>0.312</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer Surplus</td>
<td>0.458</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P_{2B}^*$ = Optimal FOX price

$P_{AB}^*$ = Optimal bundle price
### Table 3: TW chooses to bundle

<table>
<thead>
<tr>
<th></th>
<th>$[\Pi_{2B}^*(\delta) \mid entry] &lt; F$</th>
<th>$[\Pi_{2B}^*(\delta) \mid entry] \geq F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOX’s entry decision</td>
<td>No entry</td>
<td>Entry</td>
</tr>
<tr>
<td>FOX’s profits</td>
<td>0</td>
<td>$\Pi_{2B}^*(\delta) \mid entry$</td>
</tr>
<tr>
<td>Time Warner’s profits</td>
<td>$\frac{2}{3}\sqrt{\frac{2}{3}}$</td>
<td>$\Pi_{AB}^*(\delta) \mid FOX entry$</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>$1 - \frac{8}{9}\sqrt{\frac{2}{3}}$</td>
<td>$CS(\delta) \mid entry$</td>
</tr>
</tbody>
</table>

### Table 4: Effect of bundling on entry and consumer surplus

<table>
<thead>
<tr>
<th>Which outcome...</th>
<th>$0 \leq \delta \leq 0.06$</th>
<th>$0.06 \leq \delta \leq 0.5$</th>
<th>$0.5 \leq \delta \leq 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...facilitates FOX entry?</td>
<td>Bundling</td>
<td>No bundling</td>
<td>No bundling</td>
</tr>
<tr>
<td>...maximizes consumer surplus?</td>
<td>No bundling, entry</td>
<td>No bundling, entry</td>
<td>Bundling, entry</td>
</tr>
<tr>
<td>...minimizes consumer surplus?</td>
<td>No bundling, no entry</td>
<td>No bundling, no entry</td>
<td>No bundling, no entry</td>
</tr>
</tbody>
</table>
Appendix A: Bundling Without Entry

Calculations for finding the probability distribution of consumer preferences for the bundle and thereby finding the optimal bundle price:

\[ f_\alpha(x) = f_\beta(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases} \]

where \( x \) is a particular value of \( \alpha \) or \( \beta \).

\[ f_Z(z) = \int_{-\infty}^{+\infty} f_\alpha(z - y) f_\beta(y) \, dy, \]

where \( y \) is a particular value of \( \beta \) and \( z \) is a particular value of \( \alpha + \beta \).

\( f_Z(z) \) is zero unless \( 0 \leq z - y \leq 1 \), in which case it is one.

Since \( 0 \leq y \leq 1 \), if \( z < 0 \) or \( z > 2 \), \( f_Z(z) = 0 \).

If \( 0 \leq z \leq 1 \),

\[ f_\alpha(z - y) = \begin{cases} 1 & \text{where } y \leq z \\ 0 & \text{otherwise} \end{cases} \]

thus,

\[ f_Z(z) = \int_0^2 dy = z \]

If \( 1 \leq z \leq 2 \),

\[ f_\alpha(z - y) = \begin{cases} 1 & \text{where } z - 1 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

thus,

\[ f_Z(z) = \int_{z-1}^1 dy = 2 - z \]

\[ f_Z(z) = \begin{cases} z, & \text{if } 0 \leq z \leq 1 \\ 2 - z, & \text{if } 1 < z \leq 2 \end{cases} \] and \( 0 \) otherwise.
Time Warner will set the price of its bundle to maximize profits:

$$\Pi_1 = P_{AB} \int_{P_{AB}}^{2} f_Z(z) dz$$

If $P_{AB} < 1$,

$$\Pi_1 = P_{AB} \int_{P_{AB}}^{1} (z) dz + P_{AB} \int_{1}^{2} (2 - z) dz$$

$$= P_{AB} \left[ \frac{1}{2} - \frac{P_{AB}^2}{2} + \frac{1}{2} \right]$$

To find $P^*_{AB}$

$$\frac{d\Pi_1}{dP_{AB}} = 0 = 1 - \frac{3}{2} P_{AB}$$

$$P^*_{AB} = \sqrt{2/3}$$

$$\Pi_1 = 2/3 \sqrt{2/3}$$

If $P_{AB} > 1$,

$$\Pi_1 = P_{AB} \int_{P_{AB}}^{2} (2 - z) dz$$

$$= P_{AB} \left[ 2 - 2P_{AB} + \frac{1}{2} P_A \right]$$

In this case,

$$\frac{d\Pi_1}{dP_{AB}} = 0 = 2 - 4P_{AB} + \frac{3}{2} P_{AB}$$

$$P^*_{AB} = 2/3$$

$$\Pi_1 < 2/3 \sqrt{2/3}$$

$P^*_{AB} = \sqrt{2/3}$ maximizes profits and is therefore the optimal bun-
dling price.

Consumer surplus in this case would be:

$$\int_{\sqrt{2/3}}^{2} (z - \sqrt{2/3}) f_z(z) dz$$

$$= \int_{\sqrt{2/3}}^{1} (z - \sqrt{2/3}) dz + \int_{1}^{2} (z - \sqrt{2/3})(2 - z) dz$$

$$= 1 - 8/9 \sqrt{2/3}$$

Appendix B: Bundling with Entry

$$\Pi_{TW} = P_{AB} \int_{P_{AB}+\delta-P_{B2}}^{2} \int_{P_{AB}}^{2} f_z(z) dz \alpha$$

$$\Pi_{TW} = P_{AB} \int_{P_{AB}+\delta-P_{B2}}^{1} \int_{P_{AB}}^{1} (z) dz \alpha + P_{AB} \int_{P_{AB}+\delta-P_{B2}}^{2} \int_{1}^{2} (2 - z) dz \alpha$$

$$\Pi_{FOX} = P_{B2} \int_{0}^{P_{AB}+\delta-P_{B2}} \int_{P_{B2}-\delta}^{1} (z) d\beta \alpha$$

By differentiating the above functions with respect to $P_{AB}$ and $P_{B2}$, respectively, and setting the resulting functions equal to zero, the following best response functions are found:

$$P_{AB} = \frac{2P_{B2} - \delta - 2P_{B2}^2 + 4\delta P_{B2} - \delta^2}{1 - 2P_{B2} + \delta}$$

$$P_{B2} = \frac{-1 + 2P_{AB} + \delta + 3/2P_{AB}^2 + 2P_{AB}^3 - 3/2P_{AB}^2}{1 - 3/2P_{AB}^2}$$

Consumer surplus is the sum of the surplus of consumers who purchase the bundle and the surplus of consumers who purchase FOX News alone:

$$\int_{P_{AB} - P_{B2} + \delta P_{AB}}^{1} \int_{P_{AB}}^{2} (z - P_{AB}) f_z(z) dz \alpha + \int_{P_{AB} - P_{B2} + \delta}^{1} \int_{\text{MAX}(P_{B2} - \delta, 0)}^{1} (\beta + \delta - P_{B2}) d\beta \alpha$$

$$= \int_{P_{AB} - P_{B2} + \delta}^{1} \int_{P_{AB}}^{1} (z - P_{AB})(z) dz \alpha + \int_{P_{AB} - P_{B2} + \delta}^{1} \int_{1}^{2} (z - P_{AB})(2 - z) dz \alpha$$

$$+ \int_{0}^{P_{AB} - P_{B2} + \delta} \int_{\text{MAX}(P_{B2} - \delta, 0)}^{1} (\beta + \delta - P_{B2}) d\beta \alpha$$
Appendix C

Comparative anti-competitiveness of bundling at different values of $\delta$

Value of $\delta$

Maximum Level of $F$ at Which Firm 2 Enters
Appendix E

The price discriminatory effect of bundling does not have positive implications for every consumer. Consider a consumer who would purchase CNN in the non-bundling case, but does not purchase anything if CNN is bundled with HBO. This consumer’s valuation of CNN is $0.5 \leq \beta < \sqrt{2}/3$ and valuation of HBO is $\alpha \leq \beta - \sqrt{2}/3$.

Although bundling lowers this particular consumer’s surplus, the overall price discriminatory effect of bundling is an increase in total consumer surplus. When TW does not face any competition, a consumer is only hurt by bundling if this consumer’s preferences are such that purchasing one independent good at the monopoly price is valued more than purchasing nothing, which is in turn valued more than purchasing the bundle. Mathematically this consumer’s preferences would only satisfy:

\[ \alpha - P_{HBO}^M > 0 > \alpha + \beta - P_{AB}, \]
\[ P_{AB} - P_{HBO}^M > \beta, \text{ or } 0.5 - \sqrt{2/3} > \beta \]

which is approximated by

\[ 0.3 \approx \beta > 0 \]

Analogously, the consumer’s preferences would only satisfy:

\[ \beta - P_{1B}^M > 0 > \alpha + \beta - P_{AB} \]
\[ 0.5 - \sqrt{2/3} > \alpha \]

which is approximated by

\[ 0.3 \approx \alpha > 0 \]

In contrast, a consumer would have preferences

when

\[ \alpha + \beta - P_{AB} > \alpha - P_{1A}^M \]
\[ 1 > \beta \approx 0.3 \]

(with the analogue for $\alpha$). Due to the uniform distribution of preferences, $1 > \beta \approx 0.3$ occurs more frequently than $0.3 > \beta \approx 0$, and $1 > \alpha \approx 0.3$ occurs more frequently than $0.3 \approx \alpha > 0$. Thus, a consumer willing to purchase either CNN or HBO independently in the non-bundling case is more likely to purchase them both in the bundle at the “bundling discount” than to purchase nothing at all.
Bibliography


Cui, Richard, Jeffery Sean Davis, and Joe Zeff. "Disrupting the Viewing Landscape: How Internet Television Will Change TV As we Know It." MIT Sloan School of Management (2005).


Stole, Lars A. "Price discrimination and imperfect competition." University of Chicago
Graduate School of Business, 2003.
  http://www.newyorker.com/talk/financial.
Whinston, Michael D. "Tying, Foreclosure, and Exclusion." The American Economic 
Microfinance is a concept that has greatly expanded in meaning and practice since its inception. As a result of its diversification, there are a variety of ways to differentiate between microfinance institutions (MFIs). One possible method, to which this paper subscribes, is to distinguish MFIs based on their particular approach to service delivery in terms of either products and services they choose to offer to clients; why they make this choice; or how they administer those services. One group of MFIs chooses to provide only financial products to their clients, primarily in the form of small loans, in order to maximize operational efficiency. This strategy contrasts with MFIs that choose to consider additional means of economic and social development within their purview. These MFIs offer additional services beyond standard credit products, such as life insurance products, educational programs, and direct healthcare. For the purposes of this study I term this the “integrated” approach to microfinance.

This study takes as its basis the premise that achieving a balance between the financial sustainability and the social impact of an organization is one of the most significant goals of microfinance.¹ Based on this assumption, the primary research question of this study is: what approach to service delivery is most effective in helping an MFI achieve a balance between financial well-being and social impact? Financial well-being is a broad category; thus this paper focuses on financial sustainability, defined as an organization’s ability to reach its target population while covering administrative and other costs.² In seeking to answer this question, I

¹ In context of microfinance, Financial Sustainability is essentially the same as Operational Self-Sufficiency (OSS). OSS=total operating revenues/total administrative and financial expenses. If the resulting figure is greater than one hundred, the organization under evaluation is considered to be operationally self-sufficient. Operationally sustainable microfinance institutions are able to cover administrative costs with client revenues.
conducted research for two months in Peru and Bolivia, surveyed a wide variety of integrated and non-integrated MFIs, and collected qualitative data from field observations and interviews with microfinance stakeholders. I sought their perspective on their organization’s specific approach to service delivery as well as their views on the tradeoff between solvency and social mission.

Research focused on the various approaches to service delivery in the microfinance sectors of Peru and Bolivia, both of which are well-developed and rapidly growing, and whose MFIs serve as models worldwide. As of 2009, MixMarket reported that Peru had sixty-five total MFIs, a gross loan portfolio of 5.5 billion dollars, and 3.1 million active borrowers. In 2005, the microfinance sector in Peru represented only 5 percent of the total financial sector, but the country’s MFIs had as much as 40 percent of the total number of borrowers in the financial system. As of 2009, MixMarket reported that Bolivia had twenty-six total MFIs, a gross loan portfolio of 1.9 billion dollars, and 872,655 active borrowers. These two countries were selected because they are neighbors, microfinance exemplars, and contain a wide mix of purely financial MFIs and service integrated MFIs.

Current literature has called into question the previously touted effectiveness of microfinance and many in the microfinance community acknowledge the difficulty of fighting poverty through micro-loans alone. A new approach is now necessary in order to “put development back into microfinance.” Based on contemporaneous literature and on the perspectives of MFI leaders and employees, I conclude that the integrated service approach provides the best framework through which to balance the demands between financial sustainability and social impact. Opinion on the manner, however, is still mixed: while many practitioners and experts hold that no inherent conflict exists between these two aims, others believe a tradeoff inevitable.

It must be acknowledged that there is no one correct way to practice microfinance and that bundling a variety of services into MFIs is not always the most effective approach. The appropriate

*The subtitle of Fisher and Sriram’s 2002 book is “Putting development back into microfinance.”
context is required to make integration effective, and it is certainly possible to practice integrated microfinance in a way that has little added effect on clients and MFI operations. However, contrary to common criticisms of the service integration approach, my data demonstrate that providing additional non-financial services to clients does not necessarily lead to insolvency. In fact, my primary data reveal cases in which service integration not only failed to diminish sustainability and organizational well-being, but actually lead to improvements in these areas.

BACKGROUND

Many entities that are today microfinance NGOs began with a stricter focus in development and gradually moved into the microfinance sector for various reasons. Pro Mujer—an MFI active in five countries in Latin America, among them Peru and Bolivia—began providing financial services to its clients not because the organization was losing clients, but rather because those it served wanted additional financial services in order to further their progress. Carmen Velasco explained that when she and her co-founder began the organization, they provided only basic business training to Bolivian women. The women enjoyed the training and were eager to make business plans, but the directors realized that without funding, their clients could not ‘take the next step’ regarding their business initiatives. Upon the receipt of a U.S. Agency for International Development (USAID) donation, Pro Mujer’s directors were able to launch the credit component of the organization’s operations.

As a percentage of the global network of MFIs, which as of 2007 served approximately 155 million clients,* there are still only a small number of MFIs that have pioneered the integrated approach and established successful models. While experts and practitioners do not yet fully understand the financial impact of service integration, they generally accepted that going beyond basic credit and savings products adds cost. Because of the willingness to in-

---

* This number comes from the Microcredit Summit Campaign.
cur this extra cost, the majority of MFIs attempting the integrated approach are not-for-profit. Non-profit MFIs often have a more explicit social mandate than their for-profit counterparts, and the integration of services supports the achievement of the organizations’ mission (and is sometimes even central to it). Another common characteristic is the use of some form of the village-banking model in which clients meet in groups on a regular schedule, as group settings allow for the more efficient delivery of social services like educational sessions.5

The two major approaches to service integration are the parallel approach and the unified approach.6 In the parallel system, MFIs forge partnerships with external service providers that deliver social services to their clients. In the unified approach staff members within the MFI provide these services. One way in which unified MFIs sustain the cost of their non-financial services is the cross-subsidization of non revenue-earning activities with margins earned from financial services. For example, an MFI can use the revenue it earns from interest charged on loans to pay for an education component. This approach can help organizations offset the costs of non-financial services.

LITERATURE REVIEW

A growing number of studies have attempted to quantify the effects of microfinance on clients’ lives. In 2005, the MIT Poverty Action Lab undertook a randomized evaluation of the impacts of introducing microcredit in a new market. The researchers opened MFI branches in half of the slums in one Indian region, while the other half did not receive the intervention. They used only the standard group loan (first used by the Grameen Bank in Bangladesh), in which a group of ten women received a loan and were all jointly responsible for its repayment. The results revealed that the introduction of microcredit has some varied and significant effects on business outcomes and the composition of household expenditure. While the study showed that microcredit could enable households to borrow and invest, it revealed no short-term (fifteen to eighteen
months) impact on social measures such as health, education, or women’s decision-making.\textsuperscript{7}

A small body of literature has also looked into whether microfinance can be an effective tool to reach the Millennium Development Goals (MDGs) and, if so, in what way. In a 2003 paper, researchers from the Consultative Group to Assist the Poor (CGAP) argued that microfinance plays an important role in meeting the MDGs in the following areas: eradicating poverty, promoting children’s education, improving health outcomes for women and children, and empowering women.\textsuperscript{8} The authors give the rationale that many other interventions such as employment, education, and healthcare depend on people’s access to financial services, and that improvements in these other areas can only be maintained when households have increased control over financial resources.

While research focusing specifically on the effects of integrated microfinance is relatively rare, the work done has largely confirmed the positive nature of its impact.\textsuperscript{9} Institutions like Freedom from Hunger (FFH) and Innovations for Poverty Action (IPA) have performed several useful studies. Led by Dean Karlan and Martin Valdivia, these two organizations conducted an impact study in 2006, in which a randomized control design was used to examine the MFI Finca-Peru. Clients receiving only credit services were compared with those who received financial services and FFH’s business education program on topics like business strategy and marketing. The results indicated that recipients of both credit and education services displayed greater business acumen and higher revenues than those who received only credit services. From the perspective of the MFI’s financial health, the results showed higher repayment and retention rates among education clients; this is supposedly because the clients had improved their businesses and were able to repay more reliably, as well as because the clients valued the training they received and thus felt loyal to the institution. The higher retention rate translated into increased revenue for the MFI. It was calculated, however, that adding this education to the program had a marginal cost of 6 to 9 percent. Critics of the integrated model or those who look to minimize costs might
argue that this marginal cost is imprudent. But proponents of integration would argue that this increased cost is not significant enough to justify turning down this approach, especially because the increased revenue due to heightened rates of repayment and client retention exceeded the 6 to 9 percent cost.

One of the primary factors deterring MFIs from taking this approach is the trend toward commercialization and profit-seeking activities within the microfinance community. In a 2010 paper, “Weaving straw into gold,” Rodrigo Canales writes,

As NGOs, MFIs initially could rely only on donors—and retained earnings if available—to fund their growth (Marulanda and Otero 2005). This generated a dual pressure—from their need to fund their own growth and their dependence on donors—to adhere to strict norms of efficiency and profitability with a narrower focus on microcredit, the most profitable of their services (Dugan and Goodwin-Groen 2005).

As microloans are the most profitable product that MFIs offer, some organizations are being pressured into taking a narrow approach and leaving out non-credit products. There are also additional pressures for MFIs to “install complex accounting and managerial systems.” This explains why integrated microfinance, which has a more complex accounting system due to difficult-to-quantify costs, would be difficult to justify in the new, conservatively minded financial environment. The increase in international ratings organizations that specifically track MFIs, such as MicroRate, Planet Rating, and Micro Finanza Rating, in turn increases the pressure on MFIs to be profitable, attract investors, and “conform to international standards and best practices.” These pressures squeeze current practitioners of the integrated approach and discourage new MFIs from adopting service integration because of the difficulty of standardizing non-credit services such as education and healthcare initiatives. Yet, as I will discuss later, my interviews also indicate that a strong social mandate and well-defined organizational policies could allow MFIs to continue to pursue their stated goals rather than face “mission drift,” i.e., the gradual abandonment of the poverty-reduction aims of the microfinance sector.
Julia Battilana and Silvia Dorado recently examined the history of this growing pressure in a recent paper, “Building sustainable hybrid organizations.” They write:

These new organizations combined two previously separate logics: a development logic that guided their mission to help the poor, and a banking logic that required profits sufficient to support ongoing operations and fulfill fiduciary obligations... The pioneering commercial microfinance organizations thus faced the double challenge of having to survive as new ventures while striking a delicate balance between the banking and development logics they combined, and avoiding mission drift.

I argue that this sought-after equilibrium between the banking and development aspects of MFIs can be translated into a more general balance between financial sustainability and social impact in the microfinance sector. While agreement among those interviewed was not unanimous, I believe that this balance is possible and the tradeoff is not inevitable.

PROJECT METHODOLOGY

The original aim of my project was to obtain a general picture of the microfinance landscapes in the neighboring countries of Peru and Bolivia and to discover how their approaches to service delivery compare. I intended to collect management perspectives on the qualitative relationship between financial sustainability and social impact, based on whether an MFI was offering purely financial services or a mix of financial and non-financial services. Finally, I hoped to synthesize all of the data collected and determine whether one particular microfinance service model, organizational experience, or theoretical approach to microfinance stood out as particularly noteworthy.

To conduct the field research for this project, I traveled in Peru and Bolivia for two-and-a-half months in the summer of 2009. The project’s scope required a sample size that was limited in relation to the entire market. To ensure a general amount of representation of the two sectors and a variety of service delivery approaches, I chose
geographic locations and specific MFIs using purposive sampling. I selected a mix of for-profit MFIs that are regulated by banking authorities,† as well as unregulated, non-profit MFIs with non-governmental organization (NGO) status.‡

In order to analyze the various MFIs in my study, I used snowball and convenience sampling§. In total, I visited thirteen MFIs¶ and conducted interviews with thirty-one different MFI managers and employees, three professionals working in institutions that provide external support and funding to MFIs, three academics studying microfinance, one microfinance rating agency employee, and three MFI clients. I interviewed MFI employees who held a variety of positions, from executive director to director of product development for thirty to ninety minutes. I chose to include a wide variety of microfinance topics rather than focusing the conversation solely on my primary research questions and potentially biasing the interviewee in a particular direction. I also visited several MFI client centers, conducted participant observation of loan repayment meetings and group education modules, and accompanied loan officers on their daily routines of visiting the homes and/or

* Purposive sampling is a non-probability sampling technique in which subjects are selected for a study because of some characteristic they possess.
† The for-profit category includes non-bank financial institutions (NBFI), commercial banks that specialize in microfinance, and microfinance departments of full-service banks. (MicroRate)
‡ In addition to NGOs, the non-profit category also includes credit unions and other financial cooperatives, and state-owned development and postal savings banks. It should be noted that this project gives NGOs a disproportionate amount of attention in relation to their actual role in the entire sector. This decision was a result of my prior knowledge that NGOs were the primary entities conducting integrated microfinance (as integration was an important innovation to consider as an approach to service delivery).
§ Snowball sampling is a non-probability sampling technique in which existing subjects of a study recruit future subjects from among their acquaintances. It is akin to networking, and it was an appropriate form of sampling in my study because members of the Latin American microfinance community are very familiar with one another. Convenience sampling means that research subjects are in part selected at the convenience of the researcher, which was necessary for my study because I did not have access to every leader in the microfinance field.
¶ Lima, Peru MFIs included in my study: the NGO Finca-Peru, the NGO ADRA Peru, the NGO Prisma, and the for-profit EDPYME Credivision. La Paz, Bolivia MFIs included in my study: the bank BancoSol, the bank Fie, the NGO Pro Mujer Bolivia, the umbrella NGO Pro Mujer International, and the unregulated Non-Bank Financial Institution Sartawi. Cochabamba, Bolivia MFIs included in my study: Pro Mujer Bolivia, Sartawi, and the NGO AgroCapital. Puno, Peru MFIs included in my study: Pro Mujer Peru, the regulated Non-Bank Financial Institution CRAC Los Andes, and the NGO Prisma. Cusco, Peru MFIs included in my study: the NGO Asociacion Arariwa.
work places of clients. In addition to spending between one and five days at each single organization while I conducted interviews and field visits, I gathered information about microfinance issues and the specific MFIs in my study from MFI literature and external sources. These supplementary sources helped me integrate quantitative data into my analysis and confirm MFI claims regarding their financial and social development-related performance.

RESEARCH FINDINGS: KEY TRENDS IN PERUVIAN AND BOLIVIAN MICROFINANCE

The field research revealed a wide range of organizational regulatory frameworks, missions, target populations, types of services offered, and methodologies. Credit and savings deposits were still the primary products associated with microfinance. To administer such services, NGOs tended to use a system of group banking, although some of them had also expanded into individual loan offerings that clients could access once they had completed several group loan cycles (e.g., Pro Mujer Bolivia and Pro Mujer Peru). In terms of practices around integrating financial and non-financial services, I found that MFIs that are also NGOs tended to lean more toward some model of service integration. The majority of the NGOs in my study offered their clients some form of training or education, while only a small number of regulated entities offered education. Some MFIs I encountered focused primarily on business training to help clients improve in areas like business planning, marketing, and revenues; others taught clients specific trades; a third subset stressed broader life topics such as family life, hygiene, nutrition, and maternal and child health.

To deliver the training, the MFI loan officers taught short, structured modules to clients at the beginning of loan repayment meetings. While directors and managers of an MFI claimed that training modules were delivered to loan groups at every meeting, I observed that such sessions occurred unpredictably. MFIs often lack the resources to follow up with every loan officer to ensure that the education was being offered. Moreover, some NGO cli-
ents expressed dissatisfaction with long group meetings, inducing the loan officer to shorten the meeting by eliminating the training. However, my experience indicates that group meetings were often delayed not by long educational modules but by participants who were late due to making payments. In speaking informally with clients, I learned that many enjoyed the group meetings because they were a rare opportunity to participate in an activity for their own personal enrichment. These outings fulfill a social function as well: many knew the other women in their solidarity groups, which worked to fashion miniature social networks.

The poor healthcare statistics in Latin America* would suggest that a number of MFIs might offer a healthcare component. However, I found that most MFIs did not emphasize the need to combine credit and healthcare, and only one organization in my study, Pro Mujer, offered direct healthcare to clients. The financial and educational components are essentially the same in Pro Mujer Peru (PMP) and Pro Mujer Bolivia (PMB), but the way in which the two organizations deliver health services does vary. PMB uses the unified approach to integration, running its own medical clinics from which it offers clients primary healthcare as well as maternal and childcare. These clinics are located in the same centers where clients repay their individual or group loans and receive training sessions. Clients and their immediate families are able to use the health clinic whenever they need, and, to cover the cost, Pro Mujer charges a mandatory payment of sixty cents per month (which is added to the loan quota). PMP, on the other hand, practices the parallel approach to service integration, which focuses on establishing partner linkages for all health services rather than offering in-house services. These differences are partially explained by the fact that Peru’s public sector and health institutions are generally stronger than those in Bolivia, so PMP can trust outside providers to care for their clients whereas PMB must generally provide the services themselves.

* The World Health Organization calculates life expectancy at birth in Peru at sixty-one years, and in Bolivia at fifty-five years. See: http://www.who.int/countries/per/en/.
DISCUSSION

The Case for Service Integration

My findings demonstrate a broad range of practices among Peruvian and Bolivian MFIs. While this paper does not endorse specific MFI methods over others, it concludes that the integrated approach—using either education or health services in addition to a range of financial services—makes it more likely that an MFI will achieve the sought-after balance between solvency and social services. Velasco, co-founder and director of PMB, emphasized how important it is for MFIs to focus on financial stability: “We have to see ourselves as efficient, as responsible, and as using the resources we have in the best way possible to have the largest margin possible to reinvest in our clients.” Thus, in order to prove their feasibility, integrated MFIs need to have theoretical standards of efficiency that minimally match those of commercial entities.

Some practitioners of microfinance, however, have entered the sector motivated by profits rather than social development aims. According to a Finrural interviewee, “commoditized” microfinance is when you have a commercial emphasis and you focus on selling your product. This approach, however, can lead to a narrow focus on profit maximization and even promote system abuses because of microfinance’s potential to be both high profit and low-risk. If microfinance is to succeed in its goal of lifting people out of poverty, it must be an integrated, collaborative approach, involving the public and private sectors. If its focus is entirely on the business end, there exists a higher risk of clients failing to receive substantive assistance. Practitioners on both ends of the microfinance spectrum must take into account the potential long-term retention benefits of providing comprehensive services for their clients.

My interviews revealed that the majority of Peruvian and Bolivian microfinance practitioners believe that it is possible to achieve the balance between an organization’s financial well-being

* The example of Pro Mujer’s healthcare delivery through establishing alliances shows the importance of working with the private and public sectors.
and its social impact, and there is only a small minority of people in my study who believe that there is an inherent tradeoff between the two. Naldi Delgado, the director of PMP, says, “I do not believe there is a conflict. If an entity is efficient, it is going to have self-sustainability. The more efficient and financially sustainable it is, it is going to continue carrying out its activities so it will have an impact. I believe this is important for an entity to first be sustainable because if not, you don’t guarantee your permanence.”

Nadia Alcazar, an employee of the Bolivian MFI funding organization Fundapro, echoed this view:

I think that every entity has a cycle between sustainability and a great social impact. Initially you have to get sustainable, and you can’t search for both things at the same time—not when you are starting or when you’re at the stage of growth. If there ever comes a time where you are sustainable, you have an obligation to think about how to contribute to the people. So I think it is a process because you can’t demand the entity to take responsibility initially, because it won’t have anything to give. But over time, social impact has to become a part of an entity’s work.

These statements suggest the feasibility of achieving the desired balance.

Organizational Benefits to Service Integration

1) Financial Sustainability and Cost-Effectiveness

Secondary data confirms the cost-effectiveness of the service integration approach. One metric is an organization’s financial sustainability. Data from the Consortium of Private Organizations for the Promotion of Small and Micro-Enterprise (COPEME), a microfinance umbrella organization, shows that every Peruvian integrated organization studied had reached financial and operational sustainability. In addition, nearly all of the Bolivian MFIs studied are financially sustainable, with a small minority of exceptions due

* Where financial sustainability is defined as total revenues greater than total expenses and operational sustainability institutions as the ability to cover administrative costs with client revenues.
to the recent financial crisis. Another indicator of financial viability are the ratings issued by MicroRate, a microfinance ratings agency that rates MFIs based on financial performance. MicroRate reports on the institutions in the study reveal that NGOs doing service integration have received financial performance ratings as high or higher than their regulated, non-integrated counterparts. A December 2009 report from COPEME compiling financial analysis of a multitude of Peruvian MFIs shows that out of the fifteen Peruvian MFIs that are unregulated MFIs, thirteen were financially sustainable (defined as a positive net income after operating and finance expenses are considered). Thus, 85 percent of the unregulated MFIs were financially sustainable, and because unregulated institutions tend to provide integrated services, this strongly suggests that integration can be financially savvy.

One cause of this fiscal strength is that the field agent must now provide both financial and educational services. This is cost-effective because the MFI does not have to train additional staff and can leverage already existing space and resources. While it is reasonable to expect a qualified staff member to be able to manage both the financial and educational components of the group, organizations should be aware of individual employee limitations. To provide effective education, the loan officer should be well trained in the specific modules as well as in broader educational techniques. FFH claims that the training for the facilitation skills of credit and non-credit learning sessions can be done at minimal extra cost. Even if there were increased marginal costs from the training, practitioners claim that the cost must be weighed against the benefits gained from having staff capable of delivering education and financial transactions in the same session.

As a pioneer of the integrated education and health model, Pro Mujer has had to defend itself from those who question the decision not to focus solely on financial efficiency. To answer these doubts, Pro Mujer conducted a cost allocation study of the integration models in their three largest organizations, Pro Mujer Nicaragua (PMN), Pro Mujer Bolivia (PMB), and Pro Mujer Peru (PMP). The study revealed that the marginal cost of providing
health services was relatively low because overhead costs were dispersed throughout the organization. This low marginal cost of non-financial services suggests that integration can be cost-effective. Moreover, cost allocation showed that the Pro Mujer MFIs have a significant amount of incoming donations specifically for health services, and thus the non-financial component of the entity was more sustainable than initially realized. The cross-subsidization practices mean that funds are generated reliably from the built-in profit engine of interest rates and fees.

2) Competitiveness

Integrated services may improve the competitiveness of an MFI through differentiation from competitors. My interview with Luis Calvo, a manager of Pro Mujer Bolivia, confirmed this idea of improving competitiveness:

I believe that the social mission of Pro Mujer is what gives our organization a competitive factor. Why? Because many of the rest of the microfinance institutions do not offer you health services, or they do not offer you training services. So the goal for Pro Mujer, in addition to offering high quality financial services, is that the non-financial services will be of such good quality that this is our differentiating competitive factor.23

In addition, socially motivated donors (who may value the holistic approach to clients’ development over a simply financial focus) also seem more attracted to the integrated model.24 These increased external donations are also helpful in covering the costs of education or health coverage. This competitive advantage is especially significant for MFIs in today’s crowded microfinance markets, where clients must choose between a plethora of organizations offering savings and credit. The push for a ‘competitive advantage’ should be beneficial to the industry, as it will encourage innovation within MFIs seeking to better understand the needs of their target population.
3) Client Loyalty

The third major benefit observed was increased loyalty and retention of clients. In the experience of those microfinance practitioners in this study, offering clients an extra service such as education on a variety of life topics creates a more intimate relationship between clients and the MFI. The idea that clients become more emotionally tied to such organizations is in agreement with findings from Karlan’s study on Finca-Peru. In an interview discussing BancoSol, Patricia Rojas, a products and services analyst at BancoSol’s central branch in La Paz, expressed her belief that the training component of the bank would improve the fidelity of clients: “If you are a client and you have your financial entity that gives you increased opportunities, you’re going to feel more trusting and will not be as likely to leave the bank.” Various employees of Pro Mujer corroborated such claims, reporting that they have received positive feedback from clients who received the additional services. These programs are generally low-cost and contribute to the perception that an MFI is addressing the needs of its clientele in a comprehensive manner.

Social Benefits of Service Integration

1) Client Health

Several of the organizations in my study offered health education to clients (e.g., Pro Mujer, Finca-Peru, and Arariwa), and Pro Mujer was the only MFI to directly connect clients to healthcare. Pro Mujer has performed thousands of Pap smears on women and has given thousands of children vaccinations, thus playing a role in improving maternal and child health in Latin America. My interviews strongly suggest that health training and direct health interventions improve client health. Many Peruvian and Bolivian women struggle to access public health clinics and often do not trust the medical staff. When an organization like Pro Mujer combines the health clinic with the same location providing microfi-
nance services, there is a greater likelihood that women will seek preventative care for themselves and their families.

My research indicates that health education and direct delivery of health services are not only beneficial for clients, but also for the organization delivering them. According to Carmen Velasco, “Even if you don’t care about the social component of the program of what you do, if you’re clever, you don’t want to kill your clients. You want them to be prosperous, and if it’s not for social reasons, it’s for financial reasons.” This quote effectively sums up the logic of why focusing on the health and well-being of clients can be motivated by social and financial factors.

2) Education of Clients

My study did not attempt to gather quantitative evidence of the positive social effects of providing education to microfinance clients, but my field experiences and outside literature do suggest that integrating education into microfinance creates positive externalities. In an interview with Jose Ramon, a financial officer at Finca-Peru, he explained the process of how many clients undergo changes in their personal attitude, knowledge, and life circumstances throughout their participation in the integrated credit and education program:

Generally, when people arrive at our organization, they come with a lot of uncertainty about their future ... So what Finca does is in the first months that a client is with them—they teach them to value themselves, especially women because they have to work so hard, and to have a lot of dedication in their work, and to be persistent. With this training, we see that in the beginning, the women have small businesses, and then they grow with their loans, purchasing capital, and little by little, they grow ... And this begins to change the vicious cycle, helping the women be more persistent, have a goal, and improve their lives and their families.30

While anecdotal, this idea is supported by a recent study at a Peruvian MFI not included in my study. In 2007, a randomized
control trial was conducted with the clients of PRISMA, a microcredit organization in Pucallpa, Peru. Half of the loan groups in the study were randomized to receive a health education intervention, in addition to the existing microcredit services. The remaining loan groups continued to receive only microcredit services. The results demonstrate that clients had improved health knowledge on issues related to child health after receiving a health education intervention in their monthly loan groups. Whether education programs focus on entrepreneurship, family, self-esteem, or health, many of the practitioners I studied agreed that an education component to microfinance simply makes sense because the target population of MFIs is widely undereducated and often illiterate.

3) Economic Improvement

The final social benefit of integrated microfinance is the added economic progress for the clients. This element can be particularly difficult to measure, given the need to distinguish between economic improvements through solely financial means and the externalities of the social component of the mixed model. A piece of evidence that demonstrates the economic progress of clients in integrated MFIs is the fact that clients in solidarity loan groups are graduating out of their small group loan products, demonstrating that they are financially prepared to access a larger individual loan for their business. In rare cases, clients have succeeded in formalizing their informal businesses, requiring them to move beyond the NGO and work with banks.

UNDER WHAT CIRCUMSTANCES DOES MICROFINANCE INTEGRATION WORK?

Integrated microfinance will not work effectively in every context. In implementing integrated microfinance, it is important that an MFI price the service appropriately and cover these costs with adequate interest rates. Integrated microfinance can work well in countries that have consumer protections in the form of interest
rate caps, but that have the flexibility to allow the rates required by integrated programs. Integration may not be possible in countries with rigid interest rate regulations, such as Venezuela, where strict caps decrease the likelihood that higher interest rates could be levied to subsidize the cost of social services.

The case of Pro Mujer also suggests that integrated microfinance can work effectively where a country’s public services are weak, allowing an MFI to address a widespread need in the population through the unified approach. Similarly, a country with relatively effective social service provisions should allow an MFI to forge partnerships with other service providers.

CONCLUSION

Microfinance began as a tool for economic development, and many in the field still view poverty alleviation as its primary goal. With increased competition among MFIs and a greater emphasis on profits and shareholder returns, however, many MFIs are potentially losing sight of the sector’s original mission of alleviating poverty. This paper argues that one of the most effective ways for MFIs to maintain this original social mission while also remaining financially sustainable is through a deliberate mix of financial and non-financial services. Critics of the integration approach argue that other institutions exist for the purpose of delivering social services to underserved populations. What I wish to stress, however, is the existence of tools within integrated microfinance institutions that can be leveraged in ways unavailable to those focused solely on financial products. According to Velasco,

It is extremely expensive to cover the overall costs of non-financial programs. But if you piggy-back on the overall costs that you have to incur no matter what if you are a financial institution, (you have to have a general director, a CFO, etc.), you can take advantage of this installed capacity, paying only your direct costs. So it’s a perfect marriage, and it assures that your health and training intervention stays on board together with a sustainable and sometimes profitable microfinance approach.
In defending the integrated approach, this paper contributes to the literature on microfinance and suggests a possible direction in which the field may indeed be evolving. It does not suggest that all MFIs should seek to emulate the integrated model, nor that MFIs operating more traditionally are not valuable. Nevertheless, the international microfinance movement should consider alternate approaches towards microfinance as a possible tool that can accelerate the process of improving lives.

List of Microfinance Institutions Included in Field Research

<table>
<thead>
<tr>
<th>Location</th>
<th>Type of Institution</th>
<th>Type of MFI</th>
<th>Organization</th>
<th>Person/s Interviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lima, Peru</td>
<td>MFI</td>
<td>NGO</td>
<td>Finca-Peru</td>
<td>Director and Financial Manager.</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>MFI</td>
<td>NGO</td>
<td>ADRA- Peru</td>
<td>Manager of Microfinance Portfolio</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>MFI</td>
<td>NGO</td>
<td>Prisma</td>
<td>Director and Head of Social Development</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>MFI</td>
<td>EDPYME</td>
<td>Credivisión</td>
<td>CEO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Director of Marketing Unit, Manager of Marketing, Manager of Research and Development,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Director of Health Services, Loan Officer, and a client Coordinator of Financial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Services, Director of Human Resources, and Founder/Director</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>NGO</td>
<td>Pro Mujer Bolivia</td>
<td>Regional Director</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>NGO</td>
<td>Pro Mujer International</td>
<td>Manager and Loan Officers</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>Bank</td>
<td>BancoSol</td>
<td>Manager of Product Development</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>Bank</td>
<td>Fie</td>
<td>Manager</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>NGO</td>
<td>Pro Mujer Bolivia</td>
<td>Regional Director</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>NGO</td>
<td>AgroCapital</td>
<td>Manager and Loan Officers</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI</td>
<td>NGO</td>
<td>Pro Mujer Peru</td>
<td>Director</td>
</tr>
<tr>
<td>Puno, Peru</td>
<td>MFI</td>
<td>NGO</td>
<td>Prisma</td>
<td>Branch Director</td>
</tr>
<tr>
<td>Puno, Peru</td>
<td>MFI</td>
<td>NGO</td>
<td>CRAC Los Andes</td>
<td>Director, Business Manager, and Manager of Risk Unit</td>
</tr>
<tr>
<td>Casco, Peru</td>
<td>MFI</td>
<td>NGO</td>
<td>Asociación Arariwa</td>
<td>Director</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>Rating Agency</td>
<td>n/a</td>
<td>MicroRate</td>
<td>General Manager/Senior Analyst</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>MFI Fundeer</td>
<td>n/a</td>
<td>Fundapro</td>
<td>Consultant</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>MFI Umbrella</td>
<td>n/a</td>
<td>Fintrural</td>
<td></td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>MFI Umbrella</td>
<td>n/a</td>
<td>Instituto Educación</td>
<td>Qualification and Supervision Specialist</td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>Research Institution</td>
<td>n/a</td>
<td>MicroRate</td>
<td>Senior Researcher</td>
</tr>
<tr>
<td>Casco, Peru</td>
<td>Research Institution</td>
<td>n/a</td>
<td>Poverty Action</td>
<td>Researcher</td>
</tr>
<tr>
<td>La Paz, Bolivia</td>
<td>University</td>
<td>n/a</td>
<td>Universidad Católica</td>
<td></td>
</tr>
<tr>
<td>Lima, Peru</td>
<td>University</td>
<td>n/a</td>
<td>Universidad Peruanas</td>
<td></td>
</tr>
</tbody>
</table>
Notes


6 Dunford, “Building better lives,” p. 20.


11 Ibid.


15 Carmen Velasco, interview held during meeting with PMI, La Paz, Bolivia (July 2009).

16 Velasco (June 2009).

17 Naldi Delgado, interview held during meeting with Pro Mujer Peru, Puno, Peru (October 2009).

18 Nadia Alcazar, interview held during meeting with Fundapro, La Paz, Bolivia (July 2009).


20 “Microfinanzas en el Peru.”


23 Luis Calvo, interview held during meeting with Pro Mujer Bolivia, La Paz, Bolivia (July 2009).
24 Berry and Junkin, “Healthy Women,” p. 28.
25 Karlan and Valdivia, “Teaching entrepreneurship.”
26 Calvo (July 2009).
28 Calvo (July 2009).
29 Calvo (July 2009).
30 Velasco (July 2009).

Bibliography