

## Introduction

It's June 2001 and Israeli officials are in shock over statements by the United States Department of State (DOS). Public Security Minister Uzi Landau has called the DOS information outright "inaccurate,"<sup>1</sup> while somewhat more conciliatory, Mark Regev, the spokesman for the Israeli Embassy in Washington, has stressed that "Israel takes the issues raised very seriously." The Internal Security Minister advisor Hagai Herzl has called an emergency conference on setting the matter as a top policy priority.<sup>2</sup> Those near the key players describe the reaction as "hysteria" and "fireworks."<sup>3</sup> *The Jerusalem Post* has reported that Dan Ben-Eliezer of the Foreign Ministry called the international repercussions for Israel "severe" and added that "steps must be taken to remove Israel from the *unflattering* category [emphasis added]."<sup>4</sup>

What upset and embarrassed Israel so?

Fast forward to June 2005 in Jamaica. United States (US) officials are worried about the silence from Kingston. They worry about the "clock running," and fret that Jamaican officials do not want to be seen as "answering to instructions from Washington."<sup>5</sup> On June 24, after what

<sup>1</sup> Gilbert 2001b.

<sup>2</sup> Gilbert 2001a.

<sup>3</sup> Efrat 2012, 204.

<sup>4</sup> Gilbert 2001a.

<sup>5</sup> 05KINGSTON1531. Note that all ID numbers like this refer to a US Department of State Cable. These all follow the format TWO-DIGIT YEAR, EMBASSY CITY, CABLE NUMBER (without commas or spaces). Each ID is unique and sufficient to identify the document following the US Department of State identification system. Most can be brought up with a simple Internet search. They are also all stored in a database that will be made available on the book's resources site ([www.cambridge.org/ScorecardDiplomacy](http://www.cambridge.org/ScorecardDiplomacy)).

the embassy characterizes as “sensational media coverage that reported (feigned) surprise and disbelief on the part of many [government] officials,”<sup>6</sup> National Security Minister Peter Phillips finally summons US embassy officials to the Ministry of National Security. They arrive for what turns out to be a 90-minute meeting with 15 Jamaican top officials. The embassy officials later write to Washington describing the meeting’s “public ministerial disingenuousness,” and noting that “[r]eporters from the Jamaica Information Service appeared before and after the meeting with photo and video equipment, and as Phillips clearly intended, the meeting received prominent coverage in the weekend news.”<sup>7</sup> Kingston Mayor Desmond McKenzie tells the press that the issue has “jerked this country” at the highest levels.<sup>8</sup>

What were the Jamaican officials so keen to discuss, and to be *seen* doing so, with the Americans?

Now fast forward once again, this time to June 2008. A dispute between Oman and the US is making headlines in diplomatic cables from Muscat, which describes Oman as “indignant.” Sayyid Badr al-Busaidi, the Secretary General of the Ministry of Foreign Affairs, goes to the US embassy on June 9 and, visibly agitated, warns the ambassador that Oman might be “forced to reassess all aspects of its relationship with the US,” due to an incident that he calls “a ‘knife in the back’ of a friend” and describes as a personal insult. The shaken US ambassador writes Washington noting that he has never before known a senior Omani official to question the long-standing US–Omani relationship, which the Omanis usually view as “strategic.” He is shocked that the Omani Secretary General has suggested that the Free Trade Agreement, which is a personal initiative of the Sultan, could even be in jeopardy. The ambassador notes that as an indication of just how seriously Oman views the matter, Sayyid Badr al-Busaidi has canceled an upcoming meeting with a US delegation coming to Oman to discuss civil nuclear cooperation.<sup>9</sup> Reporting that the Sultan feels “dishonored” and that Oman’s “national honor has been impugned,” the embassy laments to Washington: “We therefore are caught in a dispute in which there is little common ground, and with a partner that has indicated its willingness to wager the relationship on the outcome of the matter.”<sup>10</sup>

<sup>6</sup> 05KINGSTON1531.

<sup>7</sup> 05KINGSTON1611.

<sup>8</sup> 05KINGSTON1531.

<sup>9</sup> 08MUSCAT425.

<sup>10</sup> 08MUSCAT431\_a.

Why were the Omanis so upset?

SCORECARD DIPLOMACY AND THE POWER OF  
REPUTATIONAL CONCERNS

The answers to all three of these questions revolve around states' concern for their reputation, a central theme of this book. Understanding power and influence among states is one of the most enduring issues in international relations.<sup>11</sup> It is central to global governance, order, and peace. Coercive uses of power like interventions or sanctions get considerable attention, partly because they are high profile and have traditionally monopolized the concept of power. Subtler uses of power, such as institution building or appeals to shared norms, are often overlooked because they are difficult to trace and their effects less blatant.

This book focuses on one such subtle type of power, namely the power to shape the reputations of states. Nowadays, when information is more easily disseminated and protest more easily coordinated, the reputation of states in the eyes of their citizens and the world at large matters more than ever.<sup>12</sup> The word reputation here is used in its broad, conventional linguistic sense: States want social recognition and their governments care about how they are viewed by their own citizens and the global community.<sup>13</sup> Because states value their reputation, the ability to influence it is a form of power.<sup>14</sup> This book shows how eliciting states' concern for their reputation, broadly defined, can influence their behavior – a crucial insight for how we govern our increasingly interdependent world.<sup>15</sup>

To explain what the situations above in Israel, Jamaica and Oman above have to do with the power to shape reputations, some background is needed: In 2000, the US Congress adopted the “Victims of Trafficking and Violence Protection Act” (TVPA) to fight human trafficking, also called trafficking in persons (TIP). Human trafficking is the trade in human beings usually for sexual or labor exploitation. In recent decades, such trade has flourished to create a multi-billion dollar industry that exploits millions of human beings in unfathomably degrading ways.<sup>16</sup>

<sup>11</sup> Dahl 1957, Baldwin 2016.

<sup>12</sup> Nye 1990, 100, Grant and Keohane 2005.

<sup>13</sup> Wendt 1999, Ch. 5.

<sup>14</sup> Barnett and Duvall 2005, 42.

<sup>15</sup> Throughout this book I sometimes refer to “states” or “countries” as unitary actors. This is not an evisceration of individual agency, which this book affirms. These terms are used to refer to the aggregate elites that drive decision-making within a state.

<sup>16</sup> For more discussion of the nature and extent of the problem, see Chapter 3.

Sadly, while crime gangs are the primary perpetrators, government officials are involved in human trafficking in nearly one-third of countries worldwide.<sup>17</sup> This makes the problem thorny to tackle both logistically and politically. The international community has become concerned and, also in 2000, adopted the Palermo Protocol to Prevent, Suppress and Punish Trafficking in Persons, especially Women and Children to supplement the Convention against Transnational Organized Crime (the Palermo Protocol).

The US had been one of the leaders on the Palermo Protocol, and the new US policy signaled US intent to become a global leader on this issue.<sup>18</sup> The nature of the problem and the fact that officials are so often involved in this crime led the US to a government-centered approach. The TVPA created the Office to Monitor and Combat Trafficking in Persons (US TIP Office) within the Department of State (DOS) to issue an annual report describing the efforts of other governments to combat human trafficking.<sup>19</sup> These were to be tied to some aid allocations, although the president could easily waive any repercussions.

The act and the report might have gone unnoticed in the sea of government information were it not for the fact that the US TIP Office had an additional mandate: to *grade* countries on their anti-TIP efforts. This monitoring and grading exercise, which I call *scorecard diplomacy*, was intentionally public. The drafters believed that “countries would only get serious about their failure to address human trafficking if their deficiencies were publicly identified.”<sup>20</sup> Although the World Bank and others had produced global indices and rankings on various topics, *countries* had never graded *all* other countries in the manner proposed by the TVPA. Other US reports had been less comprehensive. The Special 301 Report on intellectual property laws, for example, published since 1989, focuses only on problem countries, leaving others alone. Not so with the TIP Report: Since 2001, the report has come out with fanfare every summer and assessed governments’ efforts on prevention, protection, and prosecution of human trafficking.<sup>21</sup> In addition to criticizing countries and recommending various policy actions, it has also rated countries on “tiers,”

<sup>17</sup> Police and government officials have been identified as sources of trafficking in no less than 68 countries. Their participation is topped only by organized criminal gangs. Protection Project 2014, 41.

<sup>18</sup> DeStefano 2007.

<sup>19</sup> This office was originally named the G/TIP office and later renamed the J/TIP office. Often I refer simply to the US TIP Office.

<sup>20</sup> US Congress 2002, 8.

<sup>21</sup> Later a fourth “P,” for partnerships, was added.

with Tier 1 being the best and Tier 3 the worst – and clearly failing – grade. Importantly, the tiers reflect government *efforts*, not outcomes.<sup>22</sup>

### Scorecard Diplomacy and the Broader Grading Phenomenon

The TIP Report is not an isolated phenomenon. Grading countries' performance is becoming an increasingly common way to try to exert influence. The US itself uses this strategy in areas ranging from aid to religion. The US Millennium Challenge Corporation (MCC) developed scorecards to determine eligibility criteria for foreign aid. The aforementioned Special 301 Report reviews the global state of International Property Rights protection and enforcement and places US trading partners on a Watch List or a Priority Watch List. Similarly, the US Bureau of International Narcotics and Law Enforcement Affairs publishes the annual International Narcotics Control Strategy Report, which identifies countries with failing counternarcotics strategies. More recently the US Department of Labor has begun to place countries into performance categories in the annual Findings on Worst Forms of Child Labor, and the DOS has begun to flag the most restrictive countries in the International Religious Freedom Report. Notably, the model of the TIP Report is gaining favor. In July 2015, new legislation was introduced in the Senate to apply a “tier” grading system for countries' anti-corruption efforts explicitly modeled after the Anti-TIP Report.<sup>23</sup>

The US is not alone in its use of ratings, rankings, and blacklists. Going back as far as sovereign risk ratings, and gaining in popularity with the introduction of reports such as the Freedom in the World, by Freedom House, a range of actors has started to use global performance indicators as a tool of governance.<sup>24</sup> Today, non-governmental and inter-governmental organizations (IGOs) and even private actors rate and rank countries in different issue areas – for example, the World Bank uses the “Ease of Doing Business Index” to motivate governments to improve their business environments.<sup>25</sup> Ratings and rankings are also used at the subnational levels and may assess cities, firms, non-governmental organizations (NGOs) and other entities.<sup>26</sup> Illustrating this boom in indices,

<sup>22</sup> For a comprehensive discussion of this policy, see Chapter 3.

<sup>23</sup> United States Senate Committee on Foreign Relations 2016.

<sup>24</sup> Davis et al. 2012b, Broome and Quirk 2015, Kelley and Simmons 2015.

<sup>25</sup> Chapter 2, note 24 lists some of the scholarship that has evolved around this emerging phenomenon.

<sup>26</sup> One example is firm level audits on labor standards as discussed in Locke 2013.

recent studies have uncovered over 150 efforts to rate, rank, or benchmark countries on various dimensions.<sup>27</sup> While many focus on economic issues, these run the gamut from the environment, to health, gender issues, development, peace and security, and so forth.

As students and teachers know, grades can be powerful motivators, especially if they are public and recurring. Highly comparative and easy to understand numbers or categories stigmatize low performers because they provide an easy basis for others to point a finger.<sup>28</sup> Grades also facilitate competition. Global rating and ranking is clearly something that has caught on, but research on whether it works is nascent; nobody really knows. This makes the US efforts on human trafficking particularly interesting.

### Back to the Cases

So what had the Omani, Jamaican, and Israeli officials so upset? They were all angry about the “grades” their countries had received in the recent reports. They called the meetings to discuss their grades with the US, to express their disappointment, and to show their citizens and the world that they were taking the criticism seriously.

But why did they react so strongly? Why would they care about a US report on a narrow human rights issue? Some countries like Oman routinely sign human rights treaties and violate them, so why worry about this? The US had criticized countries for human trafficking issues in the larger DOS human rights report for years without provoking such reactions,<sup>29</sup> so why were the officials reacting so strongly to US criticism now?

Perhaps their reactions were just showmanship. Surely these officials might puff themselves up to impress the US officials, but what does that matter? Is there anything to show for all this huffing and puffing? This book will argue that it *does* matter, that many countries *do* change their policies, and that this reveals something interesting about the nature of influence in the international system.

Let’s revisit the cases above just briefly. First, Israel: Here the US becomes a steadfast participant at the TIP policy table. The Israelis engage on the highest levels through meetings between officials such as Attorney General Alberto Gonzales and Prime Minister Ehud Olmert.<sup>30</sup>

<sup>27</sup> Bandura 2008, Kelley and Simmons 2014, Broome and Quirk 2015.

<sup>28</sup> Kelley and Simmons 2015.

<sup>29</sup> For an elaboration of the comparison between reactions to the human rights report and the trafficking report, see Chapter 5.

<sup>30</sup> 06TELAVIV2620.

The momentum and focus on trafficking changes drastically with the TIP Report. Political insiders call it a “shakeup” and a “complete turn-about.”<sup>31</sup> Israel eventually passes anti-trafficking legislation that aligns with US preferences on both sex and labor trafficking and the annual reports receive ample coverage in the media, which often notes the US impetus behind government action.

Next, Jamaica: In the meeting described above, Phillips demands to know “definitively what further steps would be required for Jamaica to receive a ‘passing grade.’” The Kingston mayor publicly credits the 2005 TIP Report with “focusing attention on the issue.”<sup>32</sup> Between June and September, when the Tier 3 designation would be reassessed, the government undertakes several reforms.<sup>33</sup> By 2007 attitudes and behaviors have changed. Whereas in 2005 no Jamaican official even acknowledged the problem, two years later officials discuss it routinely, the police are investigating cases, and the Ministry of Justice is cracking down on employment ads used to lure women into prostitution.<sup>34</sup> When Jamaica eventually earns an upgrade in the TIP Report in 2007, the government proudly issues a press release touting “[T]he improved Tier 2 status [as] a welcome recognition by the international community in general and the United States Government in particular, of the intense efforts being undertaken by the government to tackle this growing problem.”<sup>35</sup>

Finally, Oman: The US meets frequently with Omani policymakers and helps focus attention on human trafficking in Oman. Embassy officials confront Omani officials and society with the nature of the problem, particularly the practices of using small, trafficked boys for camel racing. Omani officials literally take notes in meetings with US officials about what they needed to do to get a better grade.<sup>36</sup> Eventually, the US directly advises on the text of new anti-trafficking legislation, which passes.<sup>37</sup> The debate around human trafficking significantly changes how the problem is perceived and defined and, although trafficking problems persist, the camel racing issue eventually is actually eliminated through new technologies.

<sup>31</sup> Efrat 2012, 204.

<sup>32</sup> 05KINGSTON1531.

<sup>33</sup> Ribando 2005, 19.

<sup>34</sup> 07KINGSTON927.

<sup>35</sup> Jamaican Information Service 2007.

<sup>36</sup> 08MUSCAT409.

<sup>37</sup> 08MUSCAT830\_a.

This pattern has repeated itself in several countries around the world. The US has influenced policies in Argentina, the Dominican Republic, Japan, Malaysia, the Philippines, Indonesia, Armenia, Cambodia, Madagascar, United Arab Emirates, Jordan, Mozambique, Nigeria, and Ecuador, among others. Figure 1.1 maps one of the changes that the US has helped bring about, the domestic criminalization of human trafficking. Criminalization matters because countries need domestic statutes that allow them to arrest and prosecute offenders. The modern nature of the crime is such that by the end of the 1990s, most countries relied on a hodgepodge of unrelated statutes to piece together prosecutions and sometimes they lacked ways to charge offenders despite what was so obviously heinous and wrongful behavior. Thus, criminalization is not as a cure-all, but a *sin qua non* of anti-trafficking efforts. Furthermore, as later chapters will show, often criminalization has been connected to subsequent government efforts. The top map shows that few countries had adopted anti-trafficking legislation when US scorecard diplomacy began in 2001, while the bottom map shows the progress just 13 years later. The US has not been alone in pushing for these policies,<sup>38</sup> but this book will argue that its use of scorecard diplomacy has influenced the definition and norms and motivated and shaped many policy responses.

The term “influence” does not imply that scorecard diplomacy has reduced human trafficking, which is currently unknowable because of the poor data, or that its approach has been unproblematic. What US scorecard diplomacy *has* done is shape how many governments tackle this issue, including legislation, treatment of victims and other policies. Thus “influence” is the ability to change how a country behaves: to get it to pay attention and to adopt – and hopefully also implement – the recommended policies.<sup>39</sup> The approach doesn’t work everywhere, and the influence is subtler than the blunt Cold War arm-twisting, but it is pervasive and consequential, and, given the relatively low cost, rather efficient. Sometimes scorecard diplomacy has allowed the US to influence which laws countries pass, train domestic security officials, comment on domestic administrative personnel decisions, and force issues on the domestic agenda. Indeed, the human trafficking issue illustrates

<sup>38</sup> Foot et al. 2015.

<sup>39</sup> This accords with the definition of social power. Baldwin 2016, 24. Influence does not require that actors are made to act *against* their own interest. Rather, it can include empowerment, facilitation, and encouragement to get others to progress toward behaviors that are in their interests. If a policy encourages adoption of a behavior faster than would otherwise occur or in a different form, the policy is influential.



Full criminalization
  Partial criminalization
  No criminalization

2001



2014

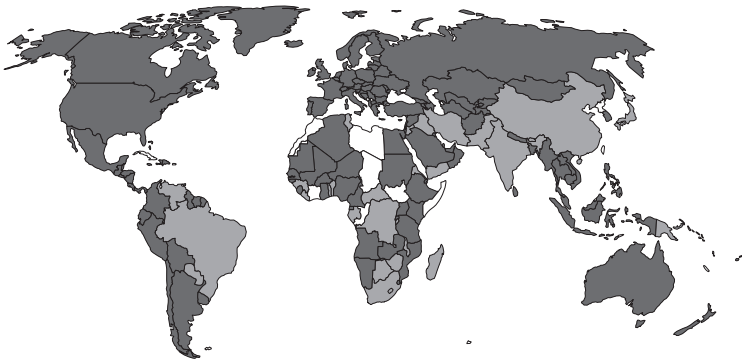


FIGURE 1.1. The spread of domestic laws criminalizing human trafficking.  
Source: Author's data.

that the US is far more intrusive and influential in the domestic politics of many countries than is commonly understood. While the TIP Report is public, the engagement it elicits goes largely under the radar, but it has been strong.

So to return to the larger question about power and reputation: Why has the US had such influence and how has it wielded it? Why did the officials above react as they did? How did the US policy bring about changes? And, most importantly, what can this teach us about international relations, state behavior, and the power of reputational concerns?

## THE ARGUMENT IN BRIEF

This book argues that the power to elicit states' concern about their reputation can be used to influence states. It contends that states care about their reputation in terms of how others perceive their performance relative to a broad set of norms and standards, and that, consequently, external actors can influence states by eliciting these concerns about their reputations – and indeed have found new ways to do so.

Specifically, the US has exercised such influence through what I call *scorecard diplomacy*. Scorecard diplomacy is the embedding of recurring monitoring and comparative grading of countries in traditional diplomacy. This contrasts in several ways with conventional “naming and shaming” or criticism of state misconduct. First, whereas shaming singles out individual countries, scorecard diplomacy gains validity by explicitly focusing on *all* countries, not just offenders. This reinforces the sense that the norms and standards are global and that everyone is being held accountable, which boosts the legitimacy of the monitoring and grading. In contrast with shaming, this inclusive approach facilitates *comparisons*, which can be powerful in the context of reputations. Second, scorecard diplomacy works not just by pointing out negative behaviors, but also by identifying desirable behaviors and plans of action. Moreover, whereas shaming is *ad hoc*, scorecard diplomacy is *recurrent*, which facilitates long-term engagement and subsequent anticipatory pressures, or what I call “status maintenance” effects. Scorecard diplomacy is thus much more than shaming. Countries are literally assigned periodic, and highly comparable, performance scores. These scores, derived by simplifying complex information, take on symbolic value and can be employed by others as well as the creators.<sup>40</sup> This allows NGOs, IGOs, and the media to augment the effect of the scores. Thus grades have outsized ability to shape states' reputations.

The US use of scorecard diplomacy is part of a larger phenomenon of using rating and rankings to influence states as a broader exercise of authority in global governance. The use of grades, rating or rankings is a particularly potent way to elicit reputational concerns. Grades are powerful symbols that shape perceptions about the performance of the graded.

<sup>40</sup> In *Economic Statecraft*, Baldwin references Harold Lasswell's work *Politics: Who Gets What When?* which referred to four different types of influence techniques, one of which was “words” or information, or symbolic means, also sometimes called propaganda. Lasswell 1958. Based on this Baldwin defines propaganda as “influence attempts relying primarily on the deliberate manipulation of symbols.” Baldwin 1985, 13. For a discussion of the politics of numbers as symbols, see also Broome and Quirk 2015.

Grades are far from neutral, however. They reduce a complex reality to a preferred interpretation and in so doing select what to call attention to and designate that as meaningful.<sup>41</sup> Symbols such as grades are thus a political exercise to label and therefore shape perceptions of reality.<sup>42</sup>

A scorecard report or a rating or ranking may capture and issue space by propagating its definitions and norms until they dominate discourse on the issue. The annual reporting and related meetings open conversations with policymakers about how to define and frame the problem. In this sense, it can define discourses, through what some have called “productive power.”<sup>43</sup> Successful scorecard diplomacy allows creators to become opinion leaders in the international system, which can change how policymakers in other countries define their interests and preferences.<sup>44</sup> A good example of this is the World Bank’s Doing Business Report, which has framed the discourse on regulation. The issuance of grades becomes a form of standard-setting activity that constructs “scripts for action” and defines “legitimate social practice.”<sup>45</sup>

States may become concerned about their reputation on the graded issue for instrumental or normative reasons. They may worry about their image or legitimacy for its own sake, or they may worry about practical implications of a poor reputation.<sup>46</sup> Practical concerns could be about material consequences, but they could also be about states’ need to be able to justify their actions to be seen as legitimate for electoral or other purposes. Thus, the use of reputation as a tool of influence is neither limited to the idea of soft power nor does it deny hard power.<sup>47</sup> Rather, it acknowledges that power is multidimensional and can work “in and through social relations.”<sup>48</sup>

The argument, however, is not a simple narrative about reputation as the lone driver of change. Rather, concern about reputation is *catalytic*; it facilitates other engagement. In the case of human trafficking, scorecard

<sup>41</sup> Bourdieu 1989, 20.

<sup>42</sup> Bourdieu 1989, 22. Eagleston-Pierce draws on Bourdieu’s notion of symbolic power to explain how relatively weaker actors can frame a situation to enhance their position in bargaining with the WTO. Eagleston-Pierce 2013.

<sup>43</sup> “Productive power concerns discourse, the social processes and systems of knowledge through which meaning is produced, fixed, lived, experienced, and transformed.” Barnett and Duvall 2005, 55. Merry et al. have a related concept called “knowledge effects.” Merry et al. 2015.

<sup>44</sup> Barnett and Finnemore 2004.

<sup>45</sup> Hansen 2011, Büthe 2012, Davis et al. 2012a.

<sup>46</sup> Erickson 2015.

<sup>47</sup> Nye 2004.

<sup>48</sup> Barnett and Duvall 2005, 42.

diplomacy combines the symbolic use of grades with traditional diplomacy and assistance that can influence state behaviors. These include things like building productive coalitions with other stakeholders to pressure for change or information exchanges that can shape understandings and habits. Once states worry about their reputation, they become more receptive to these other efforts and interactions. Combined, the scorecard and the diplomacy aid institution building and learning, which generates further reputational concerns, which motivate countries to improve in anticipation of the next cycle. The iteration is crucial; it reinforces the norms and motivates action.

Whether scorecard diplomacy works depends on three factors: the degree and credibility of the exposure of the gap between its performance and the ideal; its sensitivity to this performance gap, which will depend on the instrumental and normative salience of this gap; and, finally, its ability to prioritize the issue sufficiently to respond.

Scorecard diplomacy is interesting not just as a story about US influence in human trafficking, or even about the wider use of rankings or ratings, but for what it reveals about state behavior more generally. The officials above reacted to the ratings because they cared about their personal reputations and the reputations of their countries. As this book will show, officials shun stigmatization and find public criticism embarrassing, upsetting, and sometimes infuriating; often they seek advice on how to improve their countries' ratings. The ratings are powerful because they invoke global norms and facilitate comparisons with other countries. Officials worry about their country's relative standing in the international community; they dislike being grouped with states they perceive as worse offenders, and they don't want to lag behind their neighbors or peers. They boast when they are praised. In today's interconnected and dense information environment, policymakers react to criticism or denouncement of their country.

Thus, the particular exercise of scorecard diplomacy elaborated in this book may be unique, but it demonstrates the importance of reputation to states and its potential as a tool of influence more generally. This validates yet another facet of power, which many scholars have come to understand as not simply an artifact of capabilities or direct force, but also as a multifaceted product of institutions, structures, and discourse.<sup>49</sup> It also demonstrates the oft-invoked – but seldom systematically examined – claim that states worry about their standing and image in

<sup>49</sup> Barnett and Duvall 2005, 44.

the international community of states, a claim that is fundamental to so many other arguments about how the world works.

This book will show that US scorecard diplomacy has influenced state policies by making states concerned about their reputation in the area of human trafficking. The primary tools have been the recurrent monitoring, comparative grading, and engagement. The use of sanctions has been marginal, although the possibility has been present. The US has not been alone in the anti-trafficking fight nor has the US accomplished everything it wants. Human trafficking is an ugly and deep-seated problem. Driven by entrenched poverty, increased trans-border mobility, and unscrupulous demand, trafficking will not cease; at best it can only be managed.<sup>50</sup> Most governments still have a long way to go in addressing it adequately, including the US. Yet, this book shows that the US policy has been a major factor in this fight. It has defined the international and national discourses, engaged and empowered NGOs and IGOs, and motivated and shaped policy responses around the world. That the US has accomplished this primarily by making states worried about their reputations underscores that more subtle methods of interstate diplomacy can influence states.

One might object that this is just another story about the predominance of US power.<sup>51</sup> Clearly, its strong position is surely an advantage, and may even be a prerequisite, although examples of weak actors wielding scorecards effectively also exist, as discussed in the conclusion. However, attributing everything this book reveals to US strength alone would overlook important insights about how scorecard diplomacy works.<sup>52</sup> Indeed, the US is not quite the master of ceremonies that it used to be; alternative narratives of more diverse worldviews are emerging everywhere, from India to Turkey and China to Russia.<sup>53</sup> Furthermore, even if the US still commands considerable conventional power, such power does not obtain results by itself; it must be wielded effectively. The status and strength of the US have facilitated scorecard diplomacy, but the story is not purely a function of power asymmetry. It is about the *way* that a particular strategy has been used to exert influence. Dictating national policies is not easy, even for the powerful. Scorecard diplomacy *harnesses reputational*

<sup>50</sup> Kapstein 2006.

<sup>51</sup> On network position as an important condition for exercising influences through assessments, see Kelley and Simmons 2016.

<sup>52</sup> Or, in the words of Ikenberry and Kupchan, "Power is not reducible to coercive capacities." Ikenberry and Kupchan 1990, 289.

<sup>53</sup> I thank Peter Katzenstein for making this point.

*concerns* and enables the user to wield power more effectively. It uses reputation like a sculptor might use a chisel: to target and deliver her power more effectively than the use of her hammer alone.<sup>54</sup>

#### HOW DOES IT WORK? SCORECARD DIPLOMACY IN FIVE STEPS

Scorecard diplomacy combines traditional diplomacy with recurring monitoring and public, comparative grading of the performance of countries around the world. This approach is gaining prevalence; both the US and other actors increasingly use related approaches to rate and rank countries.

Importantly, however, scorecard diplomacy amplifies simple ratings and rankings in several ways. First, it embeds them in a web of regularized diplomacy and pushes the issue to the highest domestic players. Second, it can link the issue to other issues in the diplomatic relationship, including aid or trade. Third, localized diplomacy makes it easier to request information directly from states and local actors and to exchange ideas about possible solutions. Finally, funding to local and international actors can reinforce the central message. Thus, scorecard diplomacy creates an environment of continual policy engagement that exceeds the practices of most global performance indicators.

Figure 1.2 lays out the basic dynamics of what I call the *cycle of scorecard diplomacy*. The entire cycle is embedded within an existing normative environment. The public monitoring and comparative grading combines with ongoing diplomacy and assistance and is augmented by the indirect pressure created by media and other organizations. The grading and pressure generate concern about both present and future grades, which opens further engagement between policymakers and US diplomats. This concern and interaction increase countries' efforts to improve their ratings, thus they become more receptive to outside advice and practical assistance and keen to communicate actions taken so they can be considered for the next report. The most important feature of scorecard diplomacy, however, is its *cyclical* nature. While the steps might overlap,

<sup>54</sup> As Harold Lasswell notes, there are many tools or instruments of influence. Lasswell 1958. Cited in Baldwin. Baldwin 1985, 13. Sometimes these add up to more than the sum of their parts. As even Morgenthau noted, "[A] competent diplomacy can increase the power of a nation beyond what one would expect it to be in the view of all the other factors combined." Morgenthau 1950, 105.

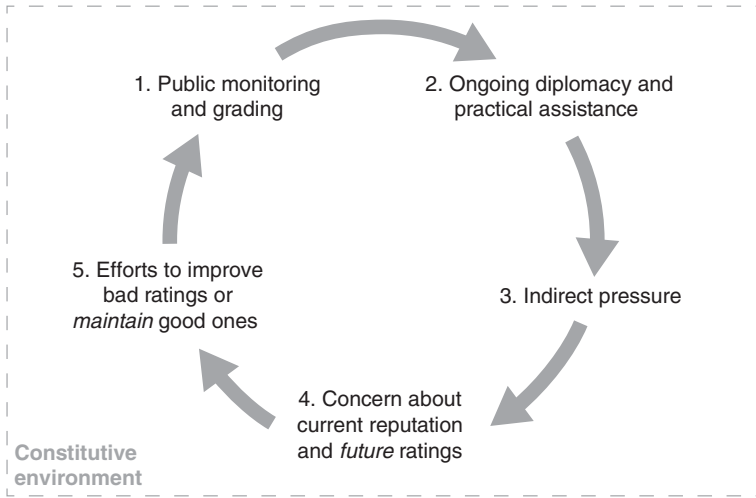


FIGURE 1.2. The cycle of scorecard diplomacy.

it is their recurrence that makes them powerful. The next section explains the components briefly.

### The Constitutive Environment

Scorecard diplomacy is rooted in and depends on prevailing standards and expectations.<sup>55</sup> Users of scorecard diplomacy may either tap into the existing norms, be part of a prior effort to establish such norms, or, through the use of scorecard diplomacy and appeals to reputation, be part of redefining and shaping these norms.<sup>56</sup> Actors who publicly assess the performance of others are actively engaged in a debate around the definition of the basic norms, and sometimes one goal of the scorecard diplomacy may be not only to shape state behaviors directly but also to shape this evolving environment. Indeed, as part of a larger conversation, the many indices that gained visibility during the 2000s and contained the word “sustainable” have likely contributed to the framing of the “Sustainable” Development Goals finalized by the United Nations in 2015. In Figure 1.2 a surrounding box represents this environment. Although for ease it is omitted in later uses of the figure, it is an essential component.

<sup>55</sup> For a broader discussion of the role of a constitutive environment, see Wendt 1999.

<sup>56</sup> See Chapter 2, note 24.

### Step 1: Public Monitoring and Grading

The heart of scorecard diplomacy is the regular publication of public reports that includes ratings or rankings of countries. These grades are often accompanied by recommendations for policy solutions. Recurrence is important for generating concerns about future grades.

Grades reduce complex reality to simple symbols that resonate easily with audiences and that other actors can employ easily to assess performance relative to a global set of norms and standards. These ratings reward or punish the non-conformant novice, and either mark or devalue their status.<sup>57</sup> Ratings and rankings mean that countries can be compared easily and movement relative to previous periods is obvious, a point elaborated by the recent body of research on global governance indicators.<sup>58</sup> This makes the information easy to process and magnifies the comparative element of status and reputation, especially when the reporting and monitoring are accompanied by concerted efforts to publicize the reporting. The monitoring furthermore has the potential to induce reflectivity, the concept that individuals change their behavior when they are aware of being observed, an idea also identified with the famous “Hawthorne effect”<sup>59</sup> and underscored in many recent experiments that show how people act more responsibly when they think someone is watching. Finally, the reporting may include narratives that help spread ideas and practices across countries.

### Step 2: Ongoing Diplomacy and Practical Assistance

Engagement is a crucial step in scorecard diplomacy and separates it from the use of rating or rankings alone. Public criticism may be sufficient to get the attention of national officials, but not enough to produce policy reforms, which may have to compete with other priorities or engender opposition. That is why scorecard diplomacy is as much about diplomacy as it is about scoring. Importantly, the two are connected: the ratings and

<sup>57</sup> Johnston defines “social influence” as “a microprocess whereby a novice’s behavior is judged by the in-group and rewarded with backpatting or status markers or punished by opprobrium and status evaluations.” Johnston 2008, 24. For the power of symbols, Bourdieu 1989, 20. On status, see Dafoe et al. 2014.

<sup>58</sup> For a discussion of the literature on ratings and rankings, see Kelley and Simmons 2015. For general references, see Hansen 2011, Büthe 2012, Davis et al. 2012a, 2012b, Merry et al. 2015.

<sup>59</sup> Adair 1984.



rankings provide what some scholars have called “external inducement” for policymakers to engage in dialogue.<sup>60</sup>

The engagement that goes along with scorecard diplomacy takes many forms. The diplomacy is often intentionally less visible than the report. The diplomacy may consist of meetings with national officials in various agencies, where diplomats can call attention to the problem, persuade policymakers of the nature of the problem, flesh out recommendations, and work with domestic officials to formulate solutions. Meetings may also bring together stakeholders to help form coalitions for reform.

Scorecard diplomacy may also link aid or other practical consequences to the ratings. When such direct or indirect issue linkage is salient, scorecard diplomacy approximates more traditional forms of conditionality. The extent to which this occurs likely depends on the issue area, but also on the actor practicing scorecard diplomacy. More powerful actors are better positioned to link issues.

If scorecard diplomacy is linked to funding, training or know-how, this can build important capacity when states possess the will but not the means to change.<sup>61</sup> Such programs can also increase coordination and collaboration among NGOs, IGOs, and the government and contribute to domestic institution building.<sup>62</sup> Diplomacy and assistance thus boost attention to the issue and provide opportunities for interaction, institution building, and information transfers.<sup>63</sup>

### **Step 3: Indirect Pressure by Third Parties**

The creators of scorecard diplomacy do not operate in a vacuum; other actors join in. Because governments worry about their reputations with multiple audiences, the greater environment and broader scope of actors are important to scorecard diplomacy. When other actors use the grades and reports, they increase the pressure on the target state.

Media is particularly keen to cover information packaged as rating or rankings rather than mere narrative reports. News stories often lead with the ratings. Sometimes they simply reprint the entire content of the

<sup>60</sup> Ikenberry and Kupchan discuss external inducement as one venue for socialization and learning. Ikenberry and Kupchan 1990, 290. In their scenario, coercive measures are used to induce elites into adopting new policies that they later internalize. In this case, the ratings and rankings may produce a non-coercive form of inducement.

<sup>61</sup> It is a long-standing argument that many countries want to comply with various international standards, but lack the capacity. Chayes and Chayes 1993, 1995.

<sup>62</sup> Finkel et al. 2006.

<sup>63</sup> Risse and Sikkink 1999, Johnston 2001, Checkel 2005.

reports more or less unedited. Other times the media itself may blame the government for underperforming. Of course, strong governments may prevent media criticism, but often both government and opposition figures comment on the accounts.

IGOs and NGOs can also boost scorecard diplomacy. If they get funding to implement related projects or use or promulgate the information in the reports, they legitimize and augment the central message. They can seize on poor ratings to pressure on their governments to reform and use information to inform their demands. Similarly to how NGOs mobilize around international legal commitments, they can use scorecard diplomacy to hold officials accountable.<sup>64</sup> Finally, NGOs also gain influence by becoming information *sources* for the reporting. This type of “information politics”<sup>65</sup> increases the influence of NGOs: When governments realize that NGOs have some input into the rating, they are likely to take NGOs more seriously.

Depending on the issue, other actors might also exert pressure on the government because of the ratings. This could include lenders, investors, or other market mechanisms. In such ways, third parties play a crucial role in the promulgation of scorecard diplomacy. Indeed, the public nature of scorecard diplomacy combined with the in-country engagement and resource provision is designed to empower such actors.

#### **Step 4: Concern About Current Reputation and *Future* Ratings**

The central step in the cycle of scorecard diplomacy is the generation of concern about the reputation generated by the tier ratings. Without this, the motivation to respond is absent. As the next chapter discusses, scorecard diplomacy can give rise to reputational concerns at both the level of the state or government, and at the level of the individual policy maker or bureaucrat responsible for a given policy area.

On a national level, governments may worry about practical consequences such as sanctions or loss of aid, trade, or other benefits. They may also worry that criticisms can damage their international or domestic legitimacy and harm their standing in the international community. A government’s concern about its reputation on a given issue may increase because the scorecard diplomacy or the associated assistance increases attention to the issue or even changes the national position.

<sup>64</sup> Simmons 2009.

<sup>65</sup> Keck and Sikkink 1998.

If they worry they will be held personally accountable or fear for job security, individual elites may also be concerned about practical consequences of poor performance.<sup>66</sup> They may also be concerned morally if they identify with the normative issues but know that their conduct, or that of their state, is contradictory. On a personal level, they may become concerned about the issues as they interact with and learn from the creators of the scorecard. Whatever the source or reason, the ratings can ignite concern that incentivizes state actors.

### Step 5: Efforts to Improve Bad Ratings or *Maintain* Good Ones

The goal of all this activity, of course, is to encourage policy reform, the last step in the cycle of scorecard diplomacy. Scorecard diplomacy usually offers many recommendations for policy actions and may stipulate what is required to improve a given grade. It thus also brings resources and know-how to the issue. As attention to the issue increases and domestic institutions begin to engage with the issue, capacity grows and becomes more institutionalized. The issue has an easier time making it onto policy agendas, and the creation of concrete capacity and programs may find more support.

This is not to say that the process ends here with perfect outcomes. Reforms are possible but not certain. If there is funding for programming, this may lead to *some* implementation, but a country's framework may need to evolve further or implementation may remain a challenge. Some countries may not respond, or perhaps a satisfactory solution for a problem does not exist. Further, countries may backslide, especially if the problem is unwieldy. Progress may occur in one cycle, regression in another. This is why scorecard diplomacy is a cycle, and why its recurrent nature is so important. Iterative practices help stabilize meaning and action,<sup>67</sup> and game theorists argue that iteration helps establish the "rules of the game" and create norms through expectations.<sup>68</sup>

### WHY STUDY SCORECARD DIPLOMACY?

Scorecard diplomacy brings a fresh perspective to the age-old, but challenging quest to study influence. Because diplomacy usually is distinctive

<sup>66</sup> Not surprisingly, survey experiments show that those most responsible for a policy are held most responsible. Renshon et al. Forthcoming.

<sup>67</sup> Pouliot and Cornut 2015, 306.

<sup>68</sup> Schelling 1980, 107, 168–169.

to each country or even each situation, most research consists of valuable – but idiosyncratic – case studies. Scholars are rarely able to examine diplomatic efforts across countries on a single topic and observe the relative merits of the various tools or differences in state responses to any given tool. This is why, although scholars have long argued that states care about their standing and image in the international community, studies have not shown this systematically. Studies that engage these ideas typically invoke them as explanations for correlations in cross-national studies but struggle to document the causal mechanisms across multiple cases. We lack rigorous investigations of how government officials respond to public criticism, whether those reactions translate into behavioral changes, and what factors facilitate or hinder such changes.

The US promotion of anti-trafficking policies offers an opportunity to overcome some of these challenges to the study of influence because of its cross-national scope, which makes it possible to compare responses to the pressure on the same issue across many countries. This is facilitated by an unprecedented availability of primary documents that makes it possible to study the causal mechanisms and derive insights about how evoking concern for reputation can work as a tool of influence.<sup>69</sup>

A good understanding of power and influence is central to the study of international relations. This book is by no means the first to argue that power is not primarily about force and coercion, but that it is also normative and symbolic: power can flow from shaping and invoking conceptions of what is normal.<sup>70</sup> However, this book brings novel evidence to bear on this argument and provides unprecedented micro-level evidence of how elites react to monitoring and to criticisms and how this connects to outcomes.

While not its main focus, the book also addresses the ever-debated role of the US in the system of global governance. Claims of the decline of US influence have become common,<sup>71</sup> but these miss subtle channels of US influence. This book shows that much consequential diplomacy happens in the background: the provision of grants that empower local actors, meetings proffering detailed advice that often gets followed, legislative council, funding for international organizations to carry out programs

<sup>69</sup> A case for testing of mechanisms in the context of randomized controlled trials is made by Mullaninathan et al., and the general insight applies. Mullaninathan et al. 2011.

<sup>70</sup> Note that Manners defines normative power as “ability to shape conceptions of ‘normal.’” Manners 2002, 240. Scorecard diplomacy holds countries to standards of behavior and shapes their reputation in terms of conceptions of these.

<sup>71</sup> See Layne 2012, 203. For a discussion of this debate, see Nye 2010.

aligned with US preferences, etc. These are all ways that the US continues to exert influence, albeit in subtler ways.

Although the book focuses on US scorecard diplomacy on human trafficking, the systematic use of reputation as a tool of power is also worth studying because the use of ratings and rankings, benchmarking, and the like is gaining popularity. The US uses it in many different issue areas, and many other actors, including IGOs and NGOs, use related approaches such as rating and ranking countries' performance across a range of different topics, some which trigger clear material payoffs and others less so. The findings may therefore provide insights into a broader range of global efforts to exert influence.

In addition to exploring the nature of influence, the book also brings useful attention to US anti-trafficking policy. IGOs and NGOs all over the world are engaged in fighting TIP,<sup>72</sup> but the US program has been one of the leaders. Opinions of the program vary greatly.<sup>73</sup> Both the US Government Accountability Office and the US Inspector General's Office have pointed out flaws in the program. Other countries have berated it as inconsistent or arrogant. Some commentators have criticized its relationship to international law<sup>74</sup> or questioned its accuracy or effectiveness.<sup>75</sup> Others accuse the policy (as well as the international Palermo Protocol itself) for harming victims, while some credit the policy with reinforcing the core provisions of the Palermo Protocol.<sup>76</sup> Yet Mark Lagon, a former US Ambassador at large to Monitor and Combat Trafficking in Person, has testified before Congress that: "[I]n case after case, we have seen how the report and rankings have worked, even among allies unused to prodding from the United States ... When some say this 'tough love' has not worked, it is flatly untrue. The US TIP Office and the report focus the mind of other governments on the problem."<sup>77</sup> This book doesn't provide definitive evidence one way or the other, but it casts more light on the subject and offers insights on the US efforts.

In sum, US scorecard diplomacy on human trafficking offers a unique opportunity to learn about the central question of influence in international relations while assessing a contested diplomatic effort on an important topic.

<sup>72</sup> Foot et al. 2015.

<sup>73</sup> DeStefano 2007.

<sup>74</sup> Chuang 2013.

<sup>75</sup> Chacon 2005–2006, Chuang 2005, Nathan 2005, Wooditch 2011, Horning et al. 2014.

<sup>76</sup> Gallagher 2015.

<sup>77</sup> Lagon 2010, 5.

## OUTCOMES, SOURCES, AND RESEARCH METHODS

This book uses the example of human trafficking to explore how scorecard diplomacy can generate reputational concerns that can bring about change. But change in what? What is the outcome of interest?

Importantly, the empirical focus of this book is neither the moral uprightness nor the efficacy of national trafficking policies, but the exercise of influence. It is not about whether the US has promoted the “best” policies, which is a matter of opinion, or whether those policies have reduced trafficking, which poor data renders elusive to assess.<sup>78</sup> Rather, this book focuses on diplomatic efficacy: has US scorecard policy been able to get governments to change their behavior – in terms of both policy and practice – by accepting the problem and undertaking reforms the US was promoting? The goal is to understand the nature of influence, what drives state behaviors, and bring us closer to that holy grail of international relations research: to understand the ability of one nation to influence another, which is the foundation of global order.<sup>79</sup>

To this end, the book studies multiple types of outcomes at various levels. As the Methods section describes further, given the complexity of the outcomes, they are explored in depth for some cases and in other cases measured cross-nationally over time. The outcomes examined are as follows:

**State criminalization of human trafficking:** Have countries criminalized human trafficking? One of the major foci of the Palermo Protocol was for countries to criminalize human trafficking in domestic laws. The measure and the appropriateness of its use and the extent to which it is meaningful are discussed more fully in Chapter 6, but essentially this captures whether states have sufficient legal measures criminalizing all forms of human trafficking with appropriately stringent penalties.

**Policy implementation:** To what extent do countries take practical measures to implement anti-TIP policies? This includes consideration of how the criminalization measure is implemented, for example, whether countries arrest and prosecute traffickers. It also includes attention to

<sup>78</sup> Tyldum and Brunovskis 2005. Much scholarship calls for greater attention to whether anti-trafficking policies effectively address root causes to reduce trafficking levels. This research is both important and much more voluminous than can be covered here, but Chuang is a good starting point for learning about much of this interesting and relevant work. Chuang 2006.

<sup>79</sup> An important reason for understanding power is, as John Harsanyi has pointed out, to understand policy options for influence. Harsanyi 1971, 80.

protection of victims as well as prevention efforts and other implementation efforts.

**Institutionalization:** Does scorecard diplomacy influence domestic designations of agency operations, change how agencies operate, or build new institutions and practices such as training academies or regularized data gathering?

**Adoption of new definitions of trafficking:** Do the norms and practices around the problem change? Does the government become more open and accepting of the problem of human trafficking, do officials change the way they discuss the problem, do laws adopt new definitions, and do attitudes and practices towards victims change?

**The reactions of officials to US grading and diplomacy:** How do officials react to the grading and issuance of reports on their government's performance on human trafficking? What responses do they have and what types of questions or concerns do they raise? Do officials compare their countries with others?

**The extent of media coverage:** Does media coverage respond to the issuance of the report and does it magnify the criticisms in the report or do officials use it to defend themselves against those criticisms?

**The behavior and views of NGOs and IGOs in the field:** Are NGOs and IGOs engaged by scorecard diplomacy? Do they discuss and use the report? Do they collaborate with the embassy, and what are their views about US efforts?

The research relies on extensive original data collection and combines multiple methods in an eclectic approach that includes the following techniques:<sup>80</sup>

**Document analysis:** The project analyzes thousands of media accounts and hundreds of other primary documents from intergovernmental organizations, the US Department of State, and other sources. The research is enhanced by the new and unprecedented availability of the quarter-million diplomatic documents from the 2012 Wikileaks release, about 8,500 of which refer to human trafficking from 2001 to early 2010.<sup>81</sup> These are US diplomatic exchanges between Washington and embassies around the world. Many of these documents reveal interactions between

<sup>80</sup> Barnett and Duvall 2005, 67–68, Sil and Katzenstein 2010, 10.

<sup>81</sup> Note that my use of these cables was revealed in my initial application to the NSF for funding and that the NSF passed the proposal through a review to ensure that it was in the national interest (as required at the time by law following an amendment to the budget from Senator Coburn (Rep-OK)). After this review, the NSF decided to sponsor this research, full knowing the use of the Wikileaks cables.

embassy staff with national officials on trafficking and report discussions in detail, as experienced by the local US diplomats. The discussions are interesting because they were not intended for publication and thus contain some frank observations and cannot easily be dismissed as public posturing.<sup>82</sup> This unconventional evidence is therefore unusually rich and promising. While the record is much fuller for some countries than others, even this incomplete archive reveals information that – combined with other evidence – is extraordinarily useful. Based on the embassy cables and media reports, I coded all official reactions to the release of the TIP Report.

**Interviews:** The book draws on interviews with 90 people from governments, NGOs and IGOs in 19 countries from all continents, conducted between 2012 and 2016. The majority were phone or Skype interviews, with about 20 one-on-one in-person interviews, three group interviews with a combined 18 people, and a few email correspondences. The interviews generally varied in length from 20 to 90 minutes. Questions were tailored to the knowledge and experience of each interviewee so that the interviews were partly structured, but open to follow-up on new information. Several were anonymous, although all but a few people agreed to be listed as sources in the Appendix.

Interviewees were selected based on research about who had played key roles in the development of US trafficking policy and inquiries with IGOs about suitable interviewees. NGO interviewees were often people who had been invited to participate in the NGO survey and preferred a personal conversation. In addition, systematic interviews were done in a few countries where research assistants connected with the project were traveling. The list of interviewees was expanded using snowball techniques, which entails asking interviewees for other recommendations of whom to interview, a technique that was also useful for confirming the appropriateness of those interviewed. While it proved difficult to get the International Organization on Migration (IOM) to agree to interviews, perhaps because it receives extensive funding for anti-TIP efforts, and while some interviewees requested anonymity, those asked to participate were generally willing and forthcoming.

**A global NGO survey:** Between 2012 and 2014 I assembled a database of over a thousand NGOs working on TIP issues around the world. During the summer and fall of 2014, 480 NGOs working in 133 countries

<sup>82</sup> Gill and Spirling 2014, 2.



responded to a survey designed to understand their engagement with the US and the TIP Report, as well as their assessments of the role of the US in their countries and their own governments' performance. Because some NGOs working in several countries filled in the survey for each country they worked in, this provided 561 separate country-level responses.

The NGO survey garnered an unusually high response rate of 43.5 percent, and those NGOs that participated were similar in terms of whether their primary focus was on human trafficking or a broader set of issues. The survey took care to minimize signaling about its purpose and any overt focus on the US, thus reducing the likelihood that any opinions of the US and its efforts influenced NGO decisions to participate in the survey. There were only slight geographic differences in participation rates, and nearly 90 percent of respondents were based outside the US, so the US organizations did not drive the findings. Furthermore, most organizations reported being very knowledgeable about TIP policy in their country.

**Case studies:** Drawing on interviews, the NGO survey, and the document analysis, 15 case studies were used to examine the evidence for the steps in the scorecard diplomacy cycle to identify outcomes and analyze the likelihood of causality between US efforts and the observed outcomes. Each case study includes construction of a chronology to facilitate analysis of sequencing of events, examination of the congruence between recommendations and outcomes, comprehensive overviews of factors driving change, and observers' own inferences. In addition to the 15 cases, other relevant examples are drawn from other countries.

**Statistical analysis:** Original data created for this project included a measure of NGO presence, US TIP grants, a dataset of public and private reactions to TIP Reports, updated data on media coverage of human trafficking, original data on sanctions waivers by the US president, data on embassy interaction with NGOs and IGOs, and data on criminalization updated from a prior project with Beth Simmons. Other pre-existing data was also merged with the above.

The Methods Appendix explains these methodologies further, describes the relevant data and lists the interviewees. It also discusses the survey methodology and participation rates, coding of data, methods for identifying and coding relevant media documents and State Department cables, as well as the case study selection and methodology. Many related materials are available on the book's resources site ([www.cambridge.org/ScorecardDiplomacy](http://www.cambridge.org/ScorecardDiplomacy)).

The analysis explores many observable implications of the argument with a variation of data. Such “triangulation” of evidence cross-validates findings in multiple ways that are not all subject to the same sets of assumptions or weaknesses.<sup>83</sup> Analysis of mechanisms at the micro level improves causal inference both by checking for evidence of the proposed mechanism and by allowing discovery of alternative explanations.<sup>84</sup> The main focus is on demonstrating the overall effects, but the methods also permit some inferences about the relative strength of the causal mechanisms.

#### THE PLAN OF THIS BOOK

The next chapter lays out the argument about reputation and scorecard diplomacy in three parts. It elaborates on the broad *definition* of reputation and discusses the notion that states hold multiple reputations in the eyes of multiple audiences and why states value a good overall reputation. It then considers the factors that *condition* whether states worry about their reputation on a given issue. It ends by exploring how the *features* of scorecard diplomacy stimulate concerns about reputation.

The chapters that follow unfold along the steps of the cycle of scorecard diplomacy. Chapters 3 and 4 focus on how scorecard diplomacy around human trafficking is produced. Chapter 3 presents the background of human trafficking to provide context for the normative environment in which scorecard diplomacy operates. It then discusses Steps 1 and 2 in the scorecard diplomacy cycle. Using the NGO survey, data about US TIP-related aid, and interviews, it describes how the US conducts scorecard diplomacy and presents data on the volume and nature of diplomatic interactions and the nature, volume, and distribution of financial assistance. The chapter also examines whether the US scorecard diplomacy treats countries differently, which is important for exploring its effectiveness.

Chapter 4 examines Step 3 in the cycle of scorecard diplomacy: indirect pressure, or how scorecard diplomacy engages other actors. Drawing on interviews and the NGO survey, the chapter highlights interactions between states, NGOs, and IGOs and illustrates how scorecard diplomacy facilitates indirect pressure from these actors. It also overviews the

<sup>83</sup> Greene and McClintock 1985.

<sup>84</sup> George and McKeown 1985, Legro 1997, 45–58.

use of funding to NGOs and IGOs and provides examples of indirect pressure in different countries.

While Part I of the book focuses on how scorecard diplomacy is produced and delivered, Part II focuses on how countries react and respond. Chapter 5 examines the evidence for Step 4 in the cycle: concern about ratings. The premise of scorecard diplomacy is that elites care about the report, take it seriously, worry about their grades etc. Is this what happens? This chapter analyzes elite reactions to tier ratings. The analysis draws on the cables from US embassies chronicling discussions with officials about their countries' tier ratings. It codes the reactions and uses case studies and statistical analysis to demonstrate the volume of reactions to the rating and the nature of concern they reveal. Finally, the chapter shows how elites react differently in private and in public, which underscores their concern with reputation.

Chapters 6 through 8 focus on how scorecard diplomacy influences state behavior, Step 5 in the cycle. Chapter 6 looks at several outcomes. It first analyzes the domestic criminalization of human trafficking, a top policy priority of US pressure. It shows that countries ramp up efforts closer to the US reporting deadline. It also shows that inclusion in the report, tier ratings, and drops in tier ratings correlate with criminalization. Importantly, it shows that countries that have documented reactions also are more likely to criminalize, indicating that reputational concern is a plausible mechanism for the established relationships. The chapter also synthesizes the case study evidence about broader outcomes such as influences on domestic institutions and implementation issues. To get at broader perceptions of the policy's effectiveness, the chapter also shares insights from the NGO survey and interviews with IGOs.

Because the effectiveness of scorecard diplomacy varies, it's important to understand why. Chapter 7 draws on the ideas developed in Chapter 2 about sensitivity, exposure, and prioritization, and uses statistical analysis and case studies to discuss the factors that have impeded or facilitated scorecard diplomacy.

The penultimate chapter uses the cases of Israel, Japan, Armenia, and Zimbabwe to illustrate how variation in the elicitation and presence of reputational concern influences how states respond to the policy demands. The cases provide a glimpse into the intensity of the diplomacy, the reactions on the ground, and some of the conditioning factors as well. Hopefully, they might entice the reader to visit the book's resources site ([www.cambridge.org/ScorecardDiplomacy](http://www.cambridge.org/ScorecardDiplomacy)) for more case study examples.

The book concludes by synthesizing the evidence and asking larger questions. What do the findings tell us about the influence of reputational concerns in today's world? What does scorecard diplomacy reveal about the nature of power and influence more broadly? What might be objections to these claims? What are the insights for US policy on human trafficking and how might these transfer to similar efforts to use ratings and rankings?