

H-Diplo | ISSF Roundtable 13-6

issforum.org

Judith Kelley. *Scorecard Diplomacy: Grading States to Influence their Reputation and Behavior*.

New York: Cambridge University Press 2017. ISBN: 9781107199972 (hardback, \$105.00);

9781316649138 (paperback, \$34.99).

14 January 2022 | <https://issforum.org/to/ir13-6>

Editor: Diane Labrosse | Commissioning Editor: Jennifer L. Erickson | Production Editor: George Fujii

Contents

Introduction by Jennifer L. Erickson, Boston College	2
Review by Alexander Cooley, Barnard College and Columbia University	5
Review by Asif Efrat, Interdisciplinary Center (IDC) Herzliya	8
Response by Judith Kelley, Duke University	10

INTRODUCTION BY JENNIFER L. ERICKSON, BOSTON COLLEGE

What convinces a country to adopt policies it might have previously eschewed as unimportant or against its interests? In practice, the global governance toolbox is notoriously limited. States, international organizations, and Non-governmental organizations (NGOs) that want other actors to change their behavior are typically reduced to selecting between the unsatisfying options of economic sanctions, military force, or some kind of ‘naming and shaming.’ Often, sanctions and military force are considered too severe, too ineffective, or too politically difficult or economically costly to adopt and implement. As a result, actors commonly use naming and shaming because they can apply it across a range of practices, whether or not they have ready access to military or institutional power capabilities, at relatively low cost to themselves. Naming and shaming is a broad category of tools that involves publicizing the normatively-unacceptable behavior of actors (usually states) in order pressure them into adopting a more normatively-acceptable behavior. Yet its effectiveness has been a frequent matter of debate.¹

Judith Kelley’s impressive book, *Scorecard Diplomacy: Grading States to Influence Their Reputation and Behavior*, investigates the use and efficacy of one specific type of naming and shaming – rankings, public grades, or what she terms “scorecard diplomacy” – in the realm of U.S. anti-human trafficking policy. She persuasively concludes that, under the right conditions, international performance indicators and public grades in particular can trigger countries’ concern for their reputations in international politics and prompt them to adopt stronger policies to penalize human trafficking. Alexander Cooley and Asif Efrat, in turn, provide insightful reviews of the book’s theoretical and empirical contributions and spark important discussions for scholars and practitioners on the politics and power of rankings, international reputation, and the promotion of anti-human trafficking law and policy around the world.

Scorecard Diplomacy begins by laying out its empirical questions related to countries’ strong – even dramatic – responses to their receipt of poor grades in U.S. Department of State’s Trafficking in Persons (TIP) reports. Kelley notes that states’ responses were not just for show; they made real, substantive changes to their anti-human trafficking policies in response to their poor TIP grades. The reason, she argues, stems directly from TIP reports’ “power to shape the reputations of states” (5). She lays out a five-step model for scorecard diplomacy more generally and provides a theoretical foundation for why it works, which is linked to states’ concern for their international reputations. She finds that reputational concerns operate – and therefore may affect state behavior – under three conditions: when governments are *sensitive* to reputational pressures; when clear, relevant information about their performance is available and *publicly exposed*; and whether elites have the political space to *prioritize* change on that policy issue (44-49). The book then examines a vast array of qualitative and quantitative data in-depth to understand how and why scorecard diplomacy can work when states have the will and capacity to improve their grades on the State Department’s annual TIP reports. Indeed, both reviews in this roundtable especially remark on the wide range of empirical evidence and sources Kelley marshals to support her arguments. Overall, the book is at once a rigorous piece of social science research, a valuable theoretical contribution on the role of reputation in international politics, and an important perspective on policy debates about whether, how, and why rankings can be a valuable naming and shaming tool in the global governance toolbox.

Both reviews highlight *Scorecard Diplomacy*’s considerable empirical, theoretical, and policy contributions. Alexander Cooley calls it “an absorbing, innovative, and richly-documented study” and focuses his assessment on the role of international performance indicators. Asif Efrat draws attention to the book’s “unprecedented, micro-level look at states’

¹ See for example: Michael P. Broache and Kate Cronin-Furman, “Does Type of Violence Matter for Interventions to Mitigate Mass Atrocities?,” *Journal of Global Security Studies* 6.1 (February 2021): 1-9; Jacqueline H.R. DeMeritt, “International Organizations and Government Killing: Does Naming and Shaming Save Lives?,” *International Interactions* 38.5 (November 2012): 597-621; Emilie M. Hafner-Burton, “Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem,” *International Organization* 62.4 (Fall 2008): 689-716; Matthew Krain, “*J'accuse!* Does Naming and Shaming Perpetrators Reduce the Severity of Genocides or Politicides?,” *International Studies Quarterly* 56.3 (September 2012): 574-589; Jack Snyder, “Backlash against Naming and Shaming: The Politics of Status and Emotion,” *The British Journal of Political Science* 22.4 (November 2020): 644-653.

worries about their image and standing” and its “wealth of empirical evidence that offers a rare glimpse into governments’ motivations and thinking.” In doing so, the book also both helps to broaden the conceptual discussion about state reputation beyond past action and credibility to capture states’ value of their reputation “as part of their state’s identity or image” (39),² and goes to great length to demonstrate those concerns in action, often behind the scenes in state decision-making.

Yet despite focusing on different aspects of Kelley’s argument, both reviews raise similar questions about the generalizability of the argument, with regard to the book’s focus on the United States as an unusually powerful international actor and to the anti-human trafficking case itself, which Efrat notes may be “easy” relative to other policy areas. While naming and shaming is often relegated to the domain of actors without access to military and/or institutional power resources, Kelley shows the United States wielding that tool effectively in the anti-human-trafficking case. Still, not many other actors have the ability to back up their social sanctions with the possibility – whether implicit or explicit – of materially-costly economic sanctions, should those social sanctions come up short. Kelley addresses the generalizability question in her response to the reviews, as well as in the book’s conclusion.

In addition, Cooley calls for more attention to authority and social hierarchies in the *ranking* system, beyond the individual grades and broader reputational concerns Kelley observes. Moreover, he asks, how will perceptions of the relative decline of the United States in recent years affect the global influence of TIP reports? Cooley’s question suggests that not only changes in the United States’ material power but also changes to its own reputation and social standing may undermine its ability to threaten other states’ reputations as a way to motivate them to change their behavior. Efrat, in contrast, highlights the importance of domestic political debates, stakeholders, and public opinion as a part of theorizing and documenting a more complete account of international rankings. Kelley notes that her book does include a discussion of US domestic politics but also acknowledges that “a more in-depth dive into the domestic dynamics could yield greater insights.”

As the reviews make clear, *Scorecard Diplomacy* has much to offer to important conversations about US foreign policy, international human rights policy and practice, international relations theory, and multi-method social science. Scholars and practitioners alike will learn much from its empirical findings and theoretical insights. It opens the door for valuable policy debates and academic research on the politics of grades, rankings, and reputation as important pieces of the global governance toolbox that may at times defy the odds to pressure states into conforming to international rules and norms.

Participants:

Judith Kelley is the ITT/Terry Sanford Professor of Public Policy, Dean and Professor in the Sanford School of Public Policy, and Professor of Political Science. Her work focuses on how states, international organizations, and NGOs can promote domestic political reforms in problem states, and how international norms, laws and other governance tools influence state behavior. Past work has focused on the International Criminal Court, the European Union, and other international organizations. Details on her election monitoring project are on the web at Project on International Election Monitoring. Her work has been published by Princeton University Press, Cambridge University Press, and in journals such as *American Political Science Review*, *American Journal of Political Science*, *International Organization*, *International Studies Quarterly*, and the *Journal of Common Market Studies*. Her first book was *Monitoring Democracy: When International Election Observation Works and Why It Often Fails* (Princeton, 2012). Her most recent book is *Scorecard Diplomacy: Grading States to Influence their Reputation and Behavior* (Cambridge University Press, 2017), and her newest work focuses

² See also Joshua William Busby, “Bono Made Jesse Helms Cry: Jubilee 2000, Debt Relief, and Moral Action in International Politics,” *International Studies Quarterly* 51.2 (June 2007): 247-275; Abram Chayes and Antonia Handler Chayes, *The New Sovereignty: Compliance with International Regulatory Agreements* (Cambridge: Harvard University Press); Jennifer L. Erickson, *Dangerous Trade: Arms Exports, Human Rights, and International Reputation* (New York: Columbia University Press, 2015); Martha Finnemore and Kathryn Sikkink, “International Norm Dynamics and Political Change,” *International Organization* 52:4 (Autumn 1998): 887-917; Alastair Iain Johnston, *Social States: China in International Institutions, 1980-2000* (Princeton: Princeton University Press, 2008)

on the global fight against human trafficking. She is leading a major research project to study the effectiveness of the diplomacy of the United States on human trafficking. She earned her Ph.D. (2001) and MPP (1997) in Public Policy at Harvard University, and a BA in Communication with honors and distinction (1995) from Stanford University.

Jennifer L. Erickson is an associate professor of political science and international studies at Boston College. Her research interests include conventional and nonconventional arms control, sanctions and arms embargoes, and the laws and norms of war. Her first book, *Dangerous Trade: Arms Exports, Human Rights, and International Reputation*, won the APSA Foreign Policy Section 2017 Best Book Award. Her current book project explores the creation of laws and norms of war around new weapons technologies. Dr. Erickson is also a faculty affiliate at MIT's Security Studies Program, an Associate Editor at *Security Studies*, and the Chair of the Board of Associate Editors for H-Diplo/ISSF. She has held fellowships at Stanford University and Dartmouth College, and was a research fellow at the Stiftung Wissenschaft und Politik (SWP) and the Wissenschaftszentrum (WZB) in Berlin.

Alexander Cooley is the Tow Professor of Political Science and Director of Columbia University's Harriman Institute for the Study of Russia, Eurasia and Eastern Europe. His research explores how external actors and organizations have impacted the sovereignty and governance of the post-Soviet states. His books include *Ranking the World: Grading States as a Tool of Global Governance* (Cambridge 2015), co-edited with Jack Snyder, and, most recently, *Exit from Hegemony: The Unravelling of the American Global Order* (Oxford University Press, 2020), co-authored with Daniel Nexon.

Asif Efrat is Associate Professor of government at the Interdisciplinary Center (IDC) Herzliya. He holds a Ph.D. in government from Harvard University. His research and teaching interests include international relations, international law, and transnational crime. His work has appeared in *International Organization*, *International Studies Quarterly*, and *European Journal of International Relations*, among others.

 REVIEW BY ALEXANDER COOLEY, BARNARD COLLEGE AND COLUMBIA UNIVERSITY

What role do international ratings and rankings play in contemporary global governance and how do states respond to being publicly graded? In *Scorecard Diplomacy*, Judith Kelley provides an absorbing, innovative and richly-documented study of the global influence of the by U.S. Department of State's Trafficking in Human Persons Report (TIP), a version of which was first published in 2001, including a detailed account of how the United States leverages its high profile rating to promote advocacy on the human trafficking issue around the world, and how states react to these evaluations and judgments.

The book contributes to a rapidly growing body of research on the politics of international performance indicators that has exploded in recent years, of which Kelley has been a pioneer.¹ Scholars from a wide variety of disciplines and fields—including law, economics, psychology and political science—have charted the widespread turn to indices, rankings and ratings across nearly all sectors of public policy and global governance. The research agenda has been rich and varied, including explorations of how indicators embed themselves into global administrative law,² how their production often rests on flawed data and non-existent state capacity,³ how international organizations and other producers of indicators reproduce social hierarchies and unequal power relations,⁴ and important questions about the value systems, normative criteria and unequal power relations that underlie the judgments and use of indicators for public policy purposes.⁵

Kelley's book contributes to this diverse literature on indicators, but also provides critical insights into how U.S. power, in public and private bilateral settings, is routinely and inextricably bound to the influence of this knowledge production. All too frequently, theories of international organization contrast explanations that privilege state power or hegemonic leadership with those that emphasize the importance of international norms and global advocacy networks. Kelley's framework effectively integrates these two approaches by outlining the "cycle of scorecard diplomacy (16-17, Figure 1.2)" a recurring sequence through which the United States issues grades and ratings of the practices of individual countries and

¹ Judith G. Kelley and Beth A. Simmons, eds. *The Power of Performance Indicators* (New York: Cambridge University Press 2020); Judith G. Kelley and Beth A. Simmons. "Introduction: The Power of Global Performance Indicators," *International Organization* 73:3 (2019): 491-510; and Judith G. Kelley and Beth A. Simmons. "Politics by Number: Indicators as social pressure in international relations," *American Journal of Political Science* 59: 1 (2015): 55-70.

² Sally Engle Merry, Kevin E. Davis, and Benedict Kingsbury, eds. *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law* (Cambridge University Press, 2015); Kevin E. Davis, Angelina Fisher, Benedict Kingsbury and Sally Engle Merry, eds., *Governance by Indicators: Global Power through Quantification and Rankings* (Oxford, Oxford University Press, 2012); and Davis, Kevin E., Benedict Kingsbury, and Sally Engle Merry. "Indicators as a Technology of Global Governance." *Law & Society Review* 46:1 (2012): 71-104.

³ Lukas Linsi and Daniel K. Mügge, "Globalization and the Growing Defects of International Economic Statistics," *Review of International Political Economy* 26: 3 (2019): 361-383; and Morten Jerven, *Poor Numbers: How we are misled by African development statistics and what to do about it*. (Ithaca: Cornell University Press, 2013).

⁴ André Broome, Alexandra Homolar, and Matthias Kranke, "Bad Science: International Organizations and the Indirect Power of Global Benchmarking," *European Journal of International Relations* 24: 3 (2018): 514-539; and Ann E. Towns, and Bahar Rumelili, "Taking the Pressure: Unpacking the relation between norms, social hierarchies, and social pressures on states," *European Journal of International Relations* 23:4 (2017): 756-779.

⁵ Péter Érdi, *Ranking: The Unwritten Rules of the Social Game We All Play* (New York: Oxford University Press, 2020); Debora Valentina Malito, Gaby Umbach, and Nehal Bhuta, eds. *The Palgrave Handbook of Indicators in Global Governance*. (Cham: Springer International Publishing, 2018); and Alexander Cooley and Jack Snyder, eds. *Ranking the World: Grading States as a Tool of Global Governance* (New York: Cambridge University Press, 2015).

then proceeds to raise the issue of compliance and performance within regular diplomatic channels and routinized interactions. The result? Countries that are moderately or poorly rated, even initially reluctant ones, come to anticipate future U.S. follow-up and diplomatic pressure and fear the reputational costs of not improving their practices, while countries with good ratings are incentivized to maintain them.

Kelley makes the case for the powerful transformative influence of scorecard diplomacy with a wide range of evidence and sources. A major one is the leaked State Department cables, disseminated by Wikileaks beginning in 2010, which Kelley painstakingly probes and codes to reveal the high level of meetings that U.S. officials have in local embassies about the TIP, how scorecard diplomacy informs meetings and agendas with members of international organizations and other third parties, and the variety of reactions (angry, face-saving, embarrassed) that the annual TIP reports elicit by government officials. Further, Kelley finds that the countries that are the most verbal in reacting to the TIP reports are the more likely to pass legislation criminalizing trafficking and establishing domestic information gathering routines and dedicated agencies (179-181), while more in-depth case studies of Armenia, Israel, Zimbabwe and Japan explore the importance of sensitivity to international reputation and continuity of engagement in different country and cultural settings (Chapter 8).

The book's approach and findings raise a number of important issues about the mechanisms through which ratings and state power exert influence over global governance. The first concerns the generalizability of the scorecard cycle, especially given the intensive diplomatic engagement, via State Department principals and embassy staff, with host states and third parties about their TIP performance. It is precisely the fact the TIP scorecard is highly integrated within periodic cycles of diplomatic contacts and meetings that makes the process so effective. Iterative diplomatic engagement pushes governments in target countries to prioritize and reframe the human trafficking issue as a criminal matter, facilitates the mobilization of domestic coalitions in support reforms, and ultimately helps to anchor new legislation to prevent human trafficking. Kelley's account shows not only how U.S. scorecard diplomacy prompted several states to adopt their own domestic reporting and information gathering practices about human trafficking, but that governments now time the release of their findings around the timetable of the U.S. study (84). Importantly, Kelley also shows how these mechanisms of routine diplomatic engagement and pressure, in effect, decouple the human trafficking issue from broader human rights concerns that might be subject to different political dynamics.

Here, we should underscore the distinctness of the TIP's scorecard diplomacy cycle within the broader ecology of international ratings and rankings. There are other cases where missions or delegations from countries that are being rated meet with or lobby their evaluators. For example, the Doing Business unit at the World Bank has an institutionalized process for engaging with visiting country delegations that are eager to discuss their scores, while even smaller non-for profits, especially ones like Freedom House and Heritage whose ratings are used as criteria in determining eligibility for Millennium Challenge Corporation (MCC) funding, routinely host visitors. However, since they involve only occasional trips to Washington DC, these instances of 'ratings diplomacy' are a far cry from the intensive and continuous diplomatic engagement described in *Scorecard Diplomacy*. Indeed, it is difficult to imagine how any non-hegemonic actor could muster the necessary resources, personnel, and global diplomatic reach to sustain such global efforts on a comparable governance issue. The influence of the TIP rating is premised upon U.S. state power and, importantly, is understood as such by the ranked countries themselves.

A second issue concerns the micro-level incentives and exact mechanisms that motivate states to comply with *Scorecard Diplomacy*. Here, Kelley assigns considerable weight to the reputational concerns of individual states, especially their sensitivity and potential exposure to external criticism within the international community, observing that "[s]tates can have reputations in relation to multiple actors: citizens, national elites, other governments and the global community (34)." Ratings and other forms of public assessment provide the necessary information to reveal the extent of gaps between a country's domestic practices and local norms and international ideal benchmarks and practices (35-36).

But the book's empirical findings actually reveal a range of important mechanisms, some of which might be either too nuanced or situational to be lumped together under what seems a broad 'reputational' rubric. For example, Kelley acknowledges the importance that states assign to social status, especially when international rankings vault them over a peer

competitor (128). But such groups and resulting comparators are less an ‘audience’ and more a hierarchical subgrouping, where status and prestige are conferred by directly outperforming a rival or adversary, asserting regional primacy, or demonstrating consistency or divergence from the performance of Western peers. Here, the authority of the rating entity also is critical, as some states, especially revisionist ones, might even value the social status afforded by the stigmatization from a negative or divergent rating, as Rebecca Adler-Nissen anticipated, especially in their effort to question the authority and/or credibility of the rating organization itself.⁶ Russian officials scorning their country’s poor democracy ratings as hypocritical Western judgments is a case in point.⁷

This also has implications for understanding the precise mechanism through which the global public policy process is impacted by international rankings and ratings. States with a strong positional sense in social hierarchy might well be more attuned to an international ranking rather than their actual rating or score. What matters more to them, for example, is how well they do vis-à-vis their peer competitors in combating corruption (moving up 10 places) rather than improving their own internal performance and practices for the sake of improving public policy. The distinction is important, not least because it often reveals a sleight of hand by organizations that claim to be primarily interested in using their assessments as diagnostic tools for improving policy as opposed to underscoring their own authoritative positions as issue experts. For example, after the World Bank was heavily criticized for its Ease of Doing Business rankings, a review panel appointed by it actually recommended that the high-profile ranking be jettisoned in favor of a more diagnostic set of ratings and benchmarks.⁸ Ultimately, however, the ranking was maintained, in large part because of its dramatic international prominence. This distinction, I think, also suggests that many of the normative changes that activists hope will be internalized by states might instead be responses to international image and status concerns. It also suggests that in some cases states will fully delegate or substitute their own expertise, research, and judgments about global trends and challenges for that of the ratings organization for reasons unrelated to the quality or probative value of the evaluation itself.⁹

Finally, Kelley’s fascinating account reveals a hegemon acting in the 2000s at the peak of its normative and agenda-setting power. It remains unclear how accelerating U.S. disengagement from global governance issues and oversight institutions, as well as broader perceptions of American hegemonic decline, might impact the long-term effectiveness of scorecard diplomacy and U.S. stewardship of human rights-related issues and norms. The Trump administration’s withdrawal from the UN Human Rights Council, as well as the removal of material detailing the reproductive rights of women from its annual Human Rights report, suggest that the TIP itself might potentially become entangled or even bargained away for specific countries as U.S. diplomacy becomes increasingly transactional. *Scorecard Diplomacy* is a landmark study of a convergent moment in US power and international influence. If that era has already passed, will the global influence of the TIP endure?

⁶ Rebecca Adler-Nissen. “Stigma Management in International Relations: Transgressive Identities, Norms, and Order in International Society.” *International Organization* 68:1 (2014): 143-176.

⁷ Andrei P. Tsygankov and David Parker. “The Securitization of Democracy: Freedom House Ratings of Russia.” *European Security* 24:1 (2015): 77-100.

⁸ Trevor Manuel, et al. “Independent Panel Review of the Doing Business Report.” World Bank, Washington, D.C., June 2013, <http://pubdocs.worldbank.org/en/237121516384849082/doing-business-review-panel-report-june-2013.pdf>.

⁹ See Rawi Abdelal and Mark Blyth, “Just Who Put you in Charge? We Did: Credit Rating Agencies and the Politics of Ratings,” in Alexander Cooley and Jack Snyder, eds., *Ranking the World: Grading States as a Tool of Global Governance* (New York: Cambridge University Press, 2015): 39-59 and Nehal Bhuta, “Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order,” in Kevin E. Davis, Angelina Fisher, Benedict Kingsbury and Sally Engle Merry, eds., *Governance by Indicators: Global Power Through Quantification and Rankings* (Oxford: Oxford University Press, 2012): 132-164.

REVIEW BY ASIF EFRAT, INTERDISCIPLINARY CENTER (IDC) HERZLIYA

In this pathbreaking book, Judith Kelley takes on a critical question in international relations: How do reputational concerns influence state behavior? More specifically: How can one elicit states' concern about their reputation to lead them to change their policies? Scholarly interest in reputation is, of course, not new.¹ Kelley's book, however, brings important and innovative insights to the analysis of reputation. It also provides a wealth of empirical evidence that offers a rare glimpse into governments' motivations and thinking.

Kelley defines scorecard diplomacy as involving the recurring monitoring and comparative grading of states that is embedded in traditional diplomacy. The focus on all countries, not just the offenders; the easy comparison across countries; the identification of negative alongside positive behavior; and the recurring nature of this exercise, all give the scores legitimacy and symbolic value, allowing them to elicit reputational concerns. But for scorecard diplomacy to work, states must care about their reputation. Indeed, the fundamental engine of scorecard diplomacy is the value that states place on their reputation. Many existing studies of reputation emphasize the material and instrumental reasons for reputational concerns, such as a desire to improve a country's foreign relations.² Kelley agrees that instrumental motivations are at work, but her account shifts the emphasis to normative concerns: states care about their image, standing, and legitimacy, and they wish to be seen as adhering to the norms in the society of states. In the model she proposes, the intensity of reputational concerns depends on the sensitivity of the government to reputational pressures as well as the availability and clarity of information that exposes the state's conduct. Yet reputational concerns in themselves do not necessarily translate into action: turning them into an actual shift in behavior requires that they obtain the necessary priority, which, in turn, is a function of both agenda setting and the capacity to design and implement change.

Overall, Kelley constructs a useful theoretical account for thinking about reputational concerns and how scorecard diplomacy can capitalize on them to affect policy change. This model takes something many IR scholars intuitively believe – that states want to look 'good' in the eyes of others – and develops a much deeper theoretical understanding of this intuition. But the test of every theory is the empirical evidence that supports it, and here Kelley does a masterful job of assembling rich and diverse evidence that focuses on the case of U.S. diplomacy on human trafficking. Based upon a treasure trove of U.S. government cables, interviews with policymakers, and statistical analysis, Kelly's multiple data sources and methods paint a careful, nuanced picture of governments' reputational concerns and their policy impact. Indeed, the result is an unprecedented, micro-level look at states' worries about their image and standing, and how such worries seem to bother states *more* than the material consequences of a negative reputation. This, of course, carries important implications for how we think about states' motivations and behavior in all areas of international relations.

But like every excellent book, this one has some limitations.

One concern relates to the generalizability of the findings: Do other rankings trigger similar responses from governments? Do governments care about ratings issued by actors other than the United States? In the case of human trafficking that this book examines, scorecard diplomacy worked since the pressure came from the United States. As the preeminent power in the international system, the United States holds significant leverage. States might wish to enjoy good ratings in order to secure the flow of material benefits from the United States, such as aid and trade. More importantly, as Kelley convincingly

¹ See, for example, Jonathan Mercer, *Reputation and International Politics* (Ithaca: Cornell University Press, 1996); Michael Tomz, *Reputation and International Cooperation: Sovereign Debt across Three Centuries* (Princeton: Princeton University Press, 2007); Jennifer L. Erickson, *Dangerous Trade: Arms Exports, Human Rights, and International Reputation* (New York: Columbia University Press, 2015).

² See, for example, Rachel Brewster, "Unpacking the State's Reputation," *Harvard International Law Journal*, 50:2 (Summer 2009): 231-269; Heather Smith-Cannoy, *Insincere Commitments: Human Rights Treaties, Abusive States, and Citizen Activism* (Washington, D.C.: Georgetown University Press, 2012).

demonstrates, they worry about their image, standing, and legitimacy in American eyes. Non-governmental organizations (NGOs) also take the American rankings seriously and employ them to pressure governments. Yet the translation of this leverage into policy impact was far from immediate. It required a significant investment of efforts and resources, including monitoring and data collection, the publication of an annual report – The Trafficking in Persons (TIP) Report – extensive meetings, and practical-assistance programs. Other states and organizations that issue rankings are unlikely to possess the credibility and resources that the United States enjoys and are unlikely to employ scorecard diplomacy as effectively as the United States. Kelley recognizes this point and cites several studies documenting the impact of rankings issued by other actors, such as the Organisation for Economic Co-operation and Development (OECD) and the World Bank (248-249). This does suggest that the success of U.S. scorecard diplomacy is not a unique case but may be part of a larger trend. At the same time, one cannot shake the notion that the exercise of scorecard diplomacy by the United States is a relatively easy case for demonstrating the impact of rankings. Other actors will find it far more difficult to trigger policy change by exploiting reputational concerns.

Human trafficking is an easy case for demonstrating the impact of rankings for yet another reason that is less recognized. From a political standpoint, human trafficking – at least its sex-trafficking part – is not costly for governments to tackle. While it is true that some corrupt officials may themselves be complicit in trafficking, the sex-trafficking industry is not typically a major source of employment or income for the country as a whole. The criminals involved in sex trafficking may seek to influence politicians through illegitimate means, such as bribery and intimidation, but they cannot openly participate in policy debates and do not represent legitimate interests that politicians can speak for. The clients of the sex-trafficking industry also do not organize politically to defend commercial sex. As a result, governments are unlikely to bear significant political costs for yielding to the American pressure to curb sex trafficking. This does not mean that cracking down on sex trafficking is easy or cost-free. To eliminate sex trafficking, governments need to invest resources in enforcement and assistance to victims – resources that governments would prefer to spend elsewhere. But the modest domestic political costs make governments more receptive to the American pressure to eliminate sex trafficking. In other areas, where governments have to act against significant domestic interests – such as important industries – reputational concerns will likely be less effective in prompting policy change.

While curbing the sex trade entails limited political repercussions, things are somewhat different when one looks at labor trafficking. Here the offenders are actors such as farmers, contractors, and factory owners who exploit, abuse, or use violence against their workers. Despite such conduct, they belong to legitimate industries or enterprises that governments are reluctant to harm. Indeed, governments have hard time thinking of these individuals as ‘criminals.’ The offenders’ image as ‘legitimate’ actors allows them to lobby officials and participate in the policymaking process. All this could make the elimination of labor trafficking more challenging compared with sex trafficking. Kelley’s analysis could have benefited from greater attention to the different political dynamics of these two types of human trafficking. I also would have liked to see a more complete picture of how scorecard diplomacy affects the domestic arena and shapes the domestic political debate. Understandably, the analysis focuses on government officials and their reputational concerns. Yet the rankings could also influence public opinion and may be contested by domestic stakeholders, and these domestic dynamics should be part of an account of international rankings.

These limitations aside, *Scorecard Diplomacy* is a scholarly achievement that provides important insights based on admirable empirical work. This book offers a close look at something we rarely get to observe – how states think about their reputation – and what we learn here should affect how we approach and analyze state behavior, inter-state influence, and international norms.

RESPONSE BY JUDITH KELLEY, DUKE UNIVERSITY

Which Latin American countries are floundering on environmental policies? Is the media in Russia getting more or less free? What are the overall global trends in foreign aid provisions, and which countries are become more or less generous? What countries have the best health systems? The questions go on and on, we are in need of good data to answer them. And data is indeed becoming increasingly abundant. But data alone will not suffice. This is especially true as we are inundated by a massive influx of data. Indeed, most raw data may be too messy, the sources may lack credibility, and the updating may be irregular. As our information environment densifies, we can only process so much data and we need ways to convey them succinctly to advance policies and arguments. This demand has, over the last decades, fueled a new global trend in efforts to package information in a usable way.

Ratings and rankings, performance indicators, monitoring reports, annual reviews, grades, and assessments —they take many forms— are all recurrent efforts to provide summary information about the performance or quality of various phenomenon. All these efforts to capture data and package it in a way that allows comparison between units seek to fill exactly this growing demand for usable information. And they are becoming not only increasingly common, but also influential.

Somewhere along the way the producers of these systems realized that they could be used for something potentially even more powerful than basic information provision, namely as tools of influence. Information has always been power, but these systems were powerful in new and interesting ways. They could be used to inform, for sure. But they could also be used to brand an organization, or, more enticingly, to influence the definition of ideas, norms and concepts. Most strikingly, they could, under the right conditions, be powerful tools to shape the behavior of the actors or units being assessed.

That point is the book's core focus. It argues that ratings and rankings can influence the behavior of the target actor who wants to avoid opprobrium or aspires to show off areas of strength. It uses the United States' efforts to fight human trafficking around the world through its annual State Department reports and ratings as a deep case study to examine how this works and when it changes policies.

One of the most common objections to this argument, and the lead point by both Alexander Cooley and Asif Efrat, is that it is difficult to see the argument generalizing beyond the case of the U.S. State Department's annual Trafficking In Persons (TIP) report.¹⁴ The U.S. is, after all, incredibly powerful. The diplomatic missions of the U.S. are rivaled by no other country, and the energy that that U.S. has poured into this issue is enormous. That objection makes sense. Certainly, those circumstances are unique. Furthermore, Efrat argues that human trafficking is a relatively easy issue for governments to tackle, although he notes that this is more true for sex trafficking than labor trafficking.

To consider the generalizability of the argument beyond the TIP context it makes sense to ask two questions. First, how are we seeing U.S. power play out in the cases? Second, do we have actual examples of other ratings and rankings by other types of actors and on other issues than trafficking that appear to have been influential?

I address the generalizability argument extensively in the book's conclusion under the heading "Asymmetry of US Power," which begins by acknowledging the merit of this objection (239). That notwithstanding, the conclusion notes four points in particular:

First, it does not seem that countries are simply trying to please the United States. If this were the case, why would officials be disappointed with a poor rating as long as the embassy applauds the country's progress? Why would officials fume at the embassy staff and argue with them, something that happens frequently? Sometimes officials even antagonize U.S. diplomats by accusing the country of being arrogant, paternalistic, or even hypocritical. They also at times denounce the U.S. efforts

¹⁴ <https://www.state.gov/reports/2020-trafficking-in-persons-report/>.

publicly. While it may well be that they would care less if the criticism came from a less significant country, it is certainly not the case that most countries are simply marching to the tune of the dictates of the United States.

Second, the statistical analysis provides no evidence that countries with economic ties with the U.S. – regardless of how funds are delivered, measured, or restricted – are more sensitive to being in the report or being pressured by the United States. While that does not prove that money or other unmeasured forms of “clout” do not matter, such lack of support is not what one would expect if it were all about power.

Third, the U.S. also has disadvantages that undermine its authority. It is criticized for acting like a global policeman, for having its own sizable trafficking problem which undermines its credibility, and for having no authority to author the report. Indeed, national officials repeatedly level these charges in their efforts to dismiss U.S. pressure. Furthermore, because of its political entanglements and changing administrations, the U.S. implements the policy inconsistently, which weakens its impact.

Finally, it also seems that U.S. diplomats think that this is not merely a game of power asymmetry. If it were only a case of the U.S. flexing its muscles, this laborious exercise would be an extraordinary waste of time. However, despite protests from some in the State Department, the U.S. has deliberately used scorecard diplomacy to bolster its influence and is even seeking to copy this strategy to other areas. If it was simply about raw power, the State Department would not need to engage so thoroughly and go to the elaborate annual exercise of allocating scores.

So yes, the U.S. has advantages in terms of its agenda setting power and its voice as a world leader, but there is a lot to suggest that this is not only a game of power asymmetry.

In a related point, Cooley notes that few other actors have the capacity to do “diplomacy” on the level of the U.S. embassy. This is true. However, I would not dismiss the level of interaction an organization like the World Bank has around its “Ease of Doing Business Report.”¹⁵ Both from primary documents as well as interview with World Bank staff, I have come to appreciate just how extensively the Bank interacts with many countries around the report. Furthermore, while the interaction certainly is helpful, it need not be as intensive as that of the U.S. embassies to have some effect. What matters is that the rated believe that the rater has the ability to shape the ratings. Some rating organizations merely compile pre-existing data on outcomes that are not actionable. That will not have much effect. But organizations like the non-governmental organization (NGO) Publish What You Fund, which creates the Aid Transparency Index,¹⁶ for example, has some modest interaction with the aid agencies it rates, and the aid agencies realize that the rating is one that they can influence through their actions. What is needed to do so is fairly transparent, and many aid agencies therefore respond. So yes, a vast diplomatic core is certainly useful, but it need not always be necessary.

But can we find other examples of ratings and rankings working? While we know that many ratings and rankings work domestically (think the “US News and World Report” rankings for best universities), are there good examples of other non-U.S. actors that have exerted meaningful influence through ratings and rankings of countries’ performances on various issues?

Yes. While there clearly are many ratings and rankings that are inconsequential, we do have examples of others that do matter. The book makes cursory mention some of these (248), but for those seeking a more in-depth treatment and further examples, Beth Simmons and I recently published an edited volume full of examples of ratings and rankings by different

¹⁵ <https://www.doingbusiness.org/en/rankings>.

¹⁶ <https://www.publishwhatyoufund.org/the-index/2020/>

actors, ranging from NGOs to international organizations as well as states.¹⁷ These ratings and rankings span issue areas, including topics such as business and environmental regulations, education and foreign aid. In sum, while the *sui generis* objection deserves appropriate consideration, there are other examples of similar systems working for other types of actors and on other issues.

Efrat's point about the relative ease of using ratings and rankings to influence human trafficking deserves some more thought, not the least because Efrat surely knows what he is talking about. As the author of outstanding work,¹⁸ especially on the role of Israel in this area, he is a top expert on the matter. The book problematizes precisely the choice of human trafficking as the topic in the section on page 71 entitled "Doesn't Everyone Want to Fight Human Trafficking?" As I note there, certainly there are harder issues, like nuclear armament, or more political rights such as freedom of speech, which might threaten power holders more directly. It's worth noting, though, that human trafficking as a topic also *lacks* some attributes that might have made influence *more* likely. First, there are few direct material linkages that can be leveraged. Secondly, as Efrat rightly notes, the domestic politics of trafficking matters too. However, there are few voting constituents that can be engaged on this issue to advocate for trafficking victims, who often lack legal standing. The issue, furthermore, is not a simple isolated policy issues. Rather, its deeply intertwined with crime and poverty, and "[i]t involves border issues, labor regulations, investigations and increased law enforcement, training, shelters, and inter-agency cooperation, which can create friction (73)." Some factors make trafficking hard to fight. Elites often benefit from trafficking, and in many cases certain practices that are more accepting of certain treatment of women or lower castes may be deeply embedded in different cultural, religious, or domestic traditions. All of these issues often make denial easier. Still, human trafficking is not world peace or climate change and should, all things equal, be easier to tackle. This, of course is a rather discouraging thought, since overall progress on human trafficking is far from complete.

Cooley also makes a great point in terms of the reputational mechanisms that may be at play in giving effect to ratings and rankings. My theory places a lot of weight on reputational concerns. Cooley, leaning on the evidence in the book, rightly points out that this is not merely a case of bilateral relationships, but that the reactions of states may reflect a social, often hierarchical, positioning system. This may lead a state to push back, for example, and by so doing, to position itself within its peer group, as Cooley notes Russia has done in response to the U.S.-based Freedom House's repeated downgrading of the state of Russian democracy. The point is that states may worry more about their relative position within their peer group than about their reputation, per say. The book also makes this point, but Cooley's formulation is helpful. Indeed, it is often the magic of the mechanisms of comparisons that elicits concerns about relative standing. Yes, states care about their reputations, but they are concerned about their image relative to their peer groups. Therefore, it's not just a matter of chastising states for their bad behavior, as shaming tends to do, but of directing criticisms at their identity in a community (246). This is why African countries are less concerned about receiving the middle grade of "Tier 2," while Switzerland, as the only recipient of that grade among advanced democracies, found that upsetting. Table 5.1 in the book really brings out this relative standing issue, highlighting the many cases of countries making intra-group comparisons (129).

Cooley's point about the World Bank's Ease of Doing Business (EDB) Ranking is also useful in pointing out that ratings and rankings can become important tools for branding an organization as an issue expert. Rush Doshi, Beth Simmons and I have explored¹⁹ the EDB rating and ranking system extensively. I can only agree that the immense international prominence

¹⁷ Judith Kelley and Beth Simmons, eds., *The Power of Global Performance Indicators* (Cambridge: Cambridge University Press, 2020).

¹⁸ Asif Efrat, *Governing Guns, Preventing Plunder: International Cooperation against Illicit Trade* (Oxford: Oxford University Press, 2012)

¹⁹ Rush Doshi, Kelley and Simmons. "The Power of Ranking: The Ease of Doing Business Indicator and Global Regulatory Behavior," *International Organization* 73.3 (Summer 2019): 611-643. DOI: <https://doi.org/10.1017/S0020818319000158>.

of that system has driven many states to respond to international image and status concerns and rather than internalizing the value of the standards and devising a considered domestic strategy.

Relatedly, Efrat also has an important point when he notes that the book could have paid even more attention to domestic politics. The book focuses mostly on the official government reactions, although the model certainly incorporates domestic audiences. The sections about “indirect pressure by third parties” addresses both international and domestic third parties, but it focuses mostly on NGOs, not on the general public. Thus, Chapter 4, *How Third Parties Boost Reputational Concerns*, is largely devoted to an in-depth discussion of how the U.S. funds and enables domestic NGOs, and how they in turn help increase the salience of human trafficking as an issue and build capacity to fight it, and how the NGOs and the media use the TIP rankings to hold their governments accountable. But clearly a more in-depth dive into the domestic dynamics could yield greater insights. In addition, Efrat is right that much may be learned from an analysis that makes a deeper differentiation between labor and sex trafficking, a point he has made well in his own work.²⁰

Finally, Cooley wonders whether the power of the TIP report will decline because of the detachment of the Trump administration from global U.S. leadership. This is one case in which my slowness to reply to the reviews at least provides some additional insight. Certainly, the Biden administration has sought to pivot to greater reengagement. Whether the trust in the U.S. leadership can be restored does of course remain to be seen. Indeed, this question itself underscores the importance of the related concept of reputation, and that power is so much more than physical capabilities. Certainly, in a future where knowledge and information grow denser and more diffused, those wishing to influence policies in countries around the world will need to continue to diversify their exercise of different types of power.

In closing, Cooley and Efrat raise important points and I’m grateful for their close reading of the book. I hope others will take the time to examine the evidence closely and discover one of the surprising ways in which power operates in today’s world.

²⁰ Asif Efrat, “Global Efforts against Human Trafficking: The Misguided Conflation of Sex, Labor, and Organ Trafficking,” *International Studies Perspectives* 17.1 (February 2016):34–54. DOI: <https://doi.org/10.1111/insp.12097>.