

The Politics of Rating Freedom: Ideological Affinity, Private Authority, and the Freedom in the World Ratings

Sarah Sunn Bush

The *Freedom in the World* (FITW) ratings of countries' freedom, created by Freedom House in 1972, are widely used by many U.S. audiences, including journalists, policymakers, and scholars. Why and how did these ratings acquire private authority in the United States? Furthermore, why and to what extent have they retained private authority over time and across different audiences? Contrary to previous research on private authority, which emphasizes the role of raters' expertise and independence, I advance an argument that emphasizes the role of ideological affinity between raters and users. Specifically, I argue that ratings are more likely to have authority among actors that share raters' ideas about concept definition and coding. I also argue that ratings are more likely to have authority among weak actors that depend on powerful other users of the ratings. Diverse evidence and methods—including data on the ratings' usage, an internal archive of Freedom House records, interviews with key informants, and a statistical analysis of bias—support the argument.

When the American non-governmental organization (NGO) Freedom House launched its annual rating of countries' freedom in 1972, the reverberations were swift. American political leaders cited the ratings regularly during foreign policy discussions

in Congress. At the same time, journalists routinely used the ratings to report on countries' political systems, citing them in leading U.S. newspapers more often than alternative ratings of democracy and related concepts.¹

Usage of the *Freedom in the World* (FITW) ratings remains frequent in the United States. The government employs them to determine whether countries qualify to receive economic aid via the Millennium Challenge Corporation (MCC), an initiative that has dispensed more than \$10 billion since 2004. It also uses them to evaluate its efforts to promote democracy in the developing world.² Moreover, U.S. journalists use the ratings when judging whether countries are democratic—as do risk ratings agencies, pension funds, and other NGOs.³ Similarly, the World Bank uses the ratings as a source for the Worldwide Governance Indicators, which are in turn used by donors to assess the effectiveness of foreign aid programs.⁴ These examples suggest that the FITW ratings are an example of private authority in world politics, defined as “situations in which non-state actors make rules or set standards that others in world politics adopt.”⁵ In fact, the FITW reports not only *rate* democracy but also *define* democracy for certain audiences.⁶

The FITW ratings have also been used outside of the United States, although attention to them there generally arrived later and has been less consistent. Whereas policymakers and media in other democracy-promoting states rarely use the ratings, policymakers and media in countries that are the *targets* of democracy promotion have done so more frequently in the twenty-first century. Those

Sarah Sunn Bush is Assistant Professor of Political Science at Temple University (sarah.bush@temple.edu). Her work focuses on the international sources of domestic politics. Her publications include a book, The Taming of Democracy Assistance (Cambridge University Press, 2015), as well as articles in journals such as International Organization, International Studies Quarterly, and Comparative Political Studies. Prior to coming to Temple, she held a post-doctoral fellowship at the Harvard Kennedy School and received her Ph.D. from Princeton University. For helpful feedback on earlier versions of this article, Bush thanks André Broome, Jeff Colgan, Maryam Deloffre, Jennifer Dixon, Orfeo Fioretos, Erin Graham, Julia Gray, Seva Gunitsky, Sidra Hamidi, Marina Henke, Alexandra Homolar, Jeffrey Isaac, Robert Keohane, Stephen Nelson, Erin Snider, Hillel Soifer, Alex Weisiger, Wendy Wong, Lauren Young, three anonymous reviewers, and seminar participants at Case Western Reserve University, Northwestern University, and the AGORA V workshop. She also thanks Temple University for its support of this research, Ronald Joseph and Charlotte Myer for research assistance, and the individuals who agreed to be interviewed for sharing their time and expertise.

doi:10.1017/S1537592717000925

© American Political Science Association 2017

September 2017 | Vol. 15/No. 3 711

users, however, are more likely than American users to criticize the ratings, though they also cheer good ratings and urge reforms after poor ratings.

These patterns present several puzzles. First, why and how did actors based in the United States—such as government officials, international organizations (IOs), and journalists—adopt the *Freedom in the World* ratings as global standards for democracy? Second, why and to what extent have the ratings failed to acquire a similar level of private authority in other countries?

Existing theories do not offer convincing answers to those questions. Leading accounts of private authority emphasize the importance of their creators' expertise.⁷ But as I will show, when the FITW ratings were created, the methods Freedom House (FH) used were neither particularly transparent nor systematic. Other prominent accounts emphasize the independence of standards' creators as a source of private authority.⁸ This explanation also falls short. As I will demonstrate, although Freedom House is an NGO, it has many informal ties to the U.S. government, and those ties have increased at the same time that the ratings' authority has spread globally.

What is missing from existing theories is a discussion of the roles of ideology and state power. Sociologists have previously noted that benchmarks reflect ideological commitments,⁹ but their insights have not been fully incorporated into international relations (IR) theories about private authority. Paying attention to the ideological dimensions of benchmarks—and how they relate to state power—suggests a new argument about when and why they are adopted by actors in world politics and an answer to the puzzles noted above.

I develop an original, two-stage argument in which, in the first stage, benchmarks are more likely to have private authority among actors that share ideas with raters, and in the second stage, benchmarks are more likely to have private authority among actors that depend on powerful other users. First, since benchmarks inevitably reflect subjective judgments, actors are more likely to adopt benchmarks that reflect shared values. Thus, the FITW ratings gained private authority first and primarily in the United States because of their affinity with U.S. foreign policy in terms of how to define democracy and code countries. Second, when benchmarks reflect the ideology of powerful states, weaker states are more likely to use them. Usage is more likely when powerful states formally adopt benchmarks, which makes them more consequential for weaker states. Thus, the FITW ratings gained private authority—albeit less so—in states targeted by U.S. democracy promotion in the twenty-first century.

I support these claims with diverse methods and evidence covering 1972 until 2010; this study is thus the first in-depth, multi-method analysis of the history of the FITW ratings.¹⁰ First, to understand variation in the ratings' usage, I present various examples as well as

systematic statistics of usage by several audiences, including states, IOs, the media, and the public. Where appropriate, I compare usage of the FITW ratings and plausible alternatives. Second, I use an archive of Freedom House records that contains thousands of documents—including the minutes from board meetings, papers of executive directors, and financial records—to understand how the ratings were and are made and used. A third source is FH staff's writings about the ratings' creation. I also draw on ten interviews with current or former Freedom House employees and a FITW ratings review meeting that I observed. Finally, I use other forms of detailed evidence to test my argument, including a statistical analysis of biases in the ratings.

My analysis has significant implications for users of the ratings, including scholars, journalists, and policymakers. Previous scholars have questioned whether academics ought to use the ratings, arguing that the ratings are biased in favor of U.S. interests.¹¹ However, one hypothesized source of bias is raters' reliance on American media sources to code countries in the database.¹² That source of bias likely plagues other ratings, too. My findings suggest that ideological affinity is a deeper, and likely subconscious, source of bias. This more systematic understanding of *why* and *how* such bias occurs is important for decisions about whether to use the ratings, which remain widely used in political science. Thus, and as I discuss in the conclusion, the analysis has important implications for the literature on democratization.

The analysis also has theoretical implications for IR. First, I offer an original argument about the sources of private authority in world politics. Indeed, the evidence presented here about Freedom House suggests that the explanatory factors identified by previous theories, such as expertise and independence, are neither necessary nor sufficient for private authority to exist in all cases. Instead, my analysis reveals the ideological and political commitments that underlie and empower indicators that appear objective. As I discuss in the conclusion, my argument should apply to other benchmarks, which are increasingly important tools of global governance.

Second, in offering a new argument about how ideological affinity leads to private authority, I contribute to the literature on democracy promotion. Recent research has sought to understand the conceptual foundations that guide how governments, IOs, and NGOs promote democracy.¹³ I draw on that literature to understand the ideology underlying the FITW ratings while offering an argument about how ideological affinity leads to private authority. In the conclusion, I return to this issue, suggesting that it is possible—and in fact, desirable—to integrate normative and conceptual insights with positive analysis of democracy promotion.

I proceed as follows. First, I draw on the previous literature to define and theorize private authority in world

politics. Second, I summarize two previous arguments that emphasize how expertise and virtue shape when and why benchmarks have authority. I then present the ideological affinity argument. Third, I provide a brief history of the FITW ratings and an overview of my data and research methods. Fourth, I examine the evidence in support of each argument, concluding that ideological affinity best explains the authority of the FITW ratings. Finally, I conclude and outline the implications of the study for scholars and policymakers.

Private Authority in World Politics

I seek to describe and explain variation across time and space in the private authority of the FITW ratings. I follow Jessica Green's conceptualization of private authority, which she defines as "situations in which non-state actors make rules or set standards that others in world politics adopt."¹⁴ As this definition suggests, private authority is inherently relational. The actors included in Green's definition include NGOs as well as "private firms, multinational corporations, associations, foundations, transnational advocacy networks, and other non-state actors."¹⁵ Building on this definition, I seek to understand why and how actors in world politics, such as states, IOs, and the media, adopted the FITW ratings as standards for democracy.

Green's definition of private authority builds on and further specifies extant definitions of private authority. For example, Hall and Biersteker consider private actors to be entities that are "neither states, state-based, nor-state created" and authority to be "institutionalized forms or expressions of power."¹⁶ Meanwhile, Avant et al. define authority as "the ability to induce deference in others."¹⁷ In their framework, the global governors can hold five types of authority: institutional, delegated, expert, principled, and capacity-based. In related research, Stroup and Wong use a similar definition of authority, although they suggest that delegated, expert, and principled authority are the most relevant types of authority for international NGOs.¹⁸

One point of disagreement concerns whether private authority necessarily refers to *legitimate* authority. For Green and Avant et al., authority and legitimacy are closely related, with legitimacy often contributing to authority. However, Green notes that there are "illegitimate forms of private authority, such as traffickers, mafias, and mercenaries," and Avant et al. likewise caution against conflating compliance with legitimacy.¹⁹ In contrast, others define authority as "rightful or legitimate rule," meaning that authority necessarily connotes legitimacy.²⁰ Two drawbacks to this latter approach are that it seems to suggest that greater authority is equated with greater legitimacy and elides the possibility of illegitimate authority. In response, Zürn et al. advocate for a multi-layered conceptualization of authority, in which the first layer involves

actors recognizing an authority as "functionally necessary" and the second layer involves actors recognizing an authority as legitimate.²¹ I find this approach—in which legitimacy is not necessary for authority but adds a layer to it—helpful for thinking about why ratings acquire authority. In the two-stage argument I develop, ratings are more likely in the first stage to have authority among actors who share ideas with raters and also to have legitimacy. In the second stage, ratings are more likely to have authority among actors that depend on powerful other users and are less likely to be perceived as legitimate.

It is important to note that across all of the above-mentioned definitions, private authority is considered a form of power, with power understood as the ability to get actors "to do something" that they "would not otherwise do."²² By this logic, private authority confers power, since actors are deferring to standards set by others. As such, it may lead to behavioral or discursive change among the governed,²³ although these outcomes are not the focus of my analysis. Critically, however, the power that private actors can achieve via making rules and setting standards does not exist in isolation. Indeed, it can be combined with and even enhance public authorities' power in respects.²⁴ Thus, I do not interpret the growth of a benchmark's authority as necessarily indicating a reduction in the authority or power of states or IOs.²⁵

The Sources of Global Governance Benchmarks' Authority

Why, how, and to what extent did the FITW ratings acquire private authority among certain audiences? I outline here three arguments about why benchmarks gain and maintain private authority: two drawn from existing literature and one original argument. Each argument yields observable implications about the likely users of benchmarks, the effects of government funding, and other outcomes, which table 1 summarizes.

Epistemic Quality

States typically delegate to private actors to take advantage of their expertise.²⁶ In technical fields, such as accounting and climate governance, private actors often possess the knowledge required to make and enforce global standards. In less technical fields, such as democracy promotion and human rights, private actors also help set and spread standards because they can quickly gather and generate useful information²⁷ and make costly, observable efforts to establish their credibility.²⁸ Even states with diverse preferences can be united in the belief that they need to delegate to expert and impartial actors.²⁹ Indeed, there is a "general political rationality of rule" that causes states to turn to NGOs for performing certain tasks.³⁰

An epistemic perspective thus implies that benchmarks will have more private authority when produced by

Table 1
Observable implications regarding private authority and benchmarks

Argument	Brief description of argument	Likely state and IO users?	Effect of government funding on authority?	Other observable implications?
Epistemic	Benchmarks will have more private authority when produced by experts	Diverse	No effect	Challenges to benchmarks' private authority will lead to improvements in their methods
Independence	Benchmarks will have more private authority when produced by independent raters	Diverse	Negative	Raters will avoid government connections to maintain their independence
Ideological Affinity	Benchmarks will have more authority among users that share common ideas with raters or that depend on powerful other users	Those with ideological affinity or that depend on powerful users	No effect or positive	Benchmarks will reflect ideological affinity in concept definition and coding

experts using costly, rigorous, and transparent methods. In other words, a benchmark will have more private authority when it is produced more expertly. In addition, when multiple benchmarks exist, more expertly produced ones will have more private authority.

The epistemic perspective also yields several other observable implications. With regard to users, it implies that benchmarks will enjoy private authority across diverse state and IO audiences. The logic is that diverse audiences will choose to use the same benchmarks so long as those audiences have similar preferences for epistemic quality. Next, the epistemic perspective suggests that government funding will not affect private authority. Unless government funding affects expertise, it should neither enhance nor detract from benchmarks' private authority. Finally, the epistemic perspective implies that challenges to benchmarks' private authority will prompt raters to improve their methodology. Since expert knowledge is the crucial ingredient for private authority according to this argument, it predicts that raters will try to enhance their perceived expertise when they seek to enhance the private authority of their benchmarks.

Independence

According to the literature on NGO authority, benchmarks also gain authority via other internal characteristics of their creators. For example, although "virtue" is rarely sufficient for private actors to maintain authority, it is often the foundation of such authority.³¹ Studies of transnational activists emphasize that they are motivated by their ideals,³² and that activists' deep commitment to a cause encourages certain audiences to trust them.

These insights imply that benchmarks will have more private authority when virtuous raters produce them. But what makes a rater virtuous? The fundamental dimension of virtue for many raters—including NGOs—is independence, or a perceived neutrality and absence of bias.³³ As Sending and Neumann state regarding NGO power, "it is precisely . . . the fact that [NGOs] appear to be autonomous political subjects with a capacity for political will-formation that make them key subjects of, and allies in, governmental tasks."³⁴ When raters are linked formally to state or corporate interests, their ability or willingness to evaluate states accurately comes into question.³⁵ Thus, a benchmark will have more private authority when it is produced more independently. In addition, when multiple benchmarks exist, more independent benchmarks will have more private authority.

The independence perspective also yields several other observable implications. With regard to users, it implies that benchmarks will enjoy private authority across diverse state and IO audiences. The logic is that diverse audiences will choose to use the same benchmarks so long as those audiences have similar preferences for independence. Next, the independence perspective predicts that government funding will decrease benchmarks' private authority. Since government funding undermines perceived independence, it should detract from private authority. Finally, this perspective implies that raters will avoid connections with states to promote benchmarks' credibility. Since independence from competing interests is crucial, this argument predicts that raters will take steps to protect their reputations.

Ideological Affinity

But what about those cases of ratings that acquire authority *despite* the absence of expertise or independence? To answer that question, I introduce a third argument that emphasizes the role of common ideas and ideals and how they relate to state power. Specifically, I argue that ideological affinity between raters and powerful users is a necessary and—in some cases, such as Freedom House—sufficient condition for the adoption of benchmarks.

According to this argument, benchmarks will first gain authority when suppliers and users share relevant ideas and ideals. I term this the *ideological affinity* perspective because it emphasizes the importance of similar preferences (i.e., affinity) over a set of common ideas and ideals (i.e., ideology) for the emergence and maintenance of a benchmark as an authority within a community. This perspective is rooted in the observation that benchmarks assess states on broad outcomes, such as democracy and development, and that the extent to which the benchmarks conform to a set of common ideas and ideals about those outcomes matters for who uses them and how. Specifically, rating states entails two decisions for which ideological affinity matters for potential users. We can observe the extent of ideological affinity between benchmarks' suppliers and users by examining their *core concepts* and *coding decisions*.

To begin, *concept definition* is the first of two essential elements of ideological affinity. Concepts such as development and democracy are defined in myriad ways, with raters opting for minimalist or maximalist definitions and emphasizing certain features.³⁶ Debates over meanings often reflect different cultural and ideological understandings about the concepts in question.³⁷ Potential users are more likely to view benchmarks as authoritative *if the underlying concepts are defined in ways that match their worldviews*. Illustrating this idea, among audiences that care about democratic elections, the reports of election monitors from non-democracies tend to lack credibility and thus private authority because such monitors are not perceived by audiences in democracies as wanting to promote free and fair elections.³⁸

Coding decisions are the second key element that affects the existence and extent of ideological affinity. Raters often must make subjective judgments about how to code individual countries. Even in the case of credit ratings, raters rely not just on "hard" data (e.g., inflation rates) but also on "soft" factors (e.g., "the stability and legitimacy of political institutions") that may be difficult to assess despite the existence of detailed coding rules.³⁹ When the suppliers and potential users of benchmarks share common ideas, the subjective judgments of raters and potential users are more likely to jibe. When they do not, private authority is unlikely. For example, a 2000 assessment of "global health systems" by the World Health

Organization (WHO) resulted in the United States being ranked thirty-seventh. From the U.S. perspective, this score was unexpectedly low and thus the benchmark lacked authority; the WHO subsequently abandoned the enterprise because "U.S. official and media criticism was so intense."⁴⁰ Often, disagreements between suppliers and potential users over how countries are coded reflect conceptual disagreements. But they may also reflect disagreements over how to code specific states.

The first stage of the ideological affinity perspective thus predicts that *benchmarks will have more private authority when suppliers and potential users share common ideas about concept definition and coding*. The mechanism through which users select benchmarks reflecting common ideas could be deliberate, but need not be. Indeed, users may adopt benchmarks that reflect their ideas without an immediate instrumental purpose or even a consideration of the ideological stakes. Regardless, this perspective implies that a benchmark will have more private authority when it reflects ideas that are more aligned with potential users. Moreover, when multiple benchmarks exist, better-aligned benchmarks will have more private authority.

The first stage of the ideological affinity perspective also implies that benchmarks will enjoy more authority among state and IO audiences that share ideas with raters. Notably, it implies that not all audiences will treat the same benchmarks as having authority, even if they have similar preferences over raters' expertise and independence. Instead, audiences that share common ideas with benchmarks' suppliers will be the first users to embrace the benchmarks.

Shared ideas about concept definition and coding countries are admittedly difficult variables to define and operationalize. Nevertheless, it is possible to test the ideological affinity perspective's assumptions that how benchmarks' concepts are defined and how countries are coded reflect an affinity between raters and powerful users. First, one can map out the potential conceptual models associated with the benchmark's subject and then analyze relevant texts, information from interviews, and other qualitative material to establish the "conceptual contours and discursive trends" among raters and their potential audiences.⁴¹ Second, one can map out the potential assumptions of various users about how countries will be coded and then use statistical analyses and case studies to establish whether benchmarks code countries in ways that are biased in favor of particular users' assumptions.

If benchmarks acquire private authority via ideological affinity with powerful states, they become important for weaker states because they reflect the ideas of the powerful. Thus, the second stage of my argument implies that weaker states will adopt a benchmark in the absence of ideological affinity because stronger states have adopted it. For example, regardless of whether a developing country embraces the U.S. government's ideas about the

ills of human trafficking, being included on the U.S. State Department's trafficking "watch list" may prompt it to criminalize human trafficking because the United States is powerful and the country wants to comply with U.S. standards.⁴² More generally, when powerful actors adopt a benchmark, many weaker actors will adopt that benchmark, too. They may also criticize the benchmark, an act that represents its own form of recognition.

Thus, the second stage of the argument implies that benchmarks that reflect shared ideas with the powerful will be more likely to have authority in weak states, even in the absence of ideological affinity. Moreover, benchmarks' private authority will increase within weak states when powerful states *formally* adopt them. The sanctions associated with bad performance increase for weak states when standards are adopted formally (e.g., as selection criteria for aid). Such sanctions make weak states more likely to use a benchmark.

In addition, the ideological affinity perspective predicts that the effect of government funding on benchmarks' private authority depends on the audience. In general, government funding will have no effect on the likelihood of a benchmark acquiring private authority. If, however, government funding indicates shared ideas between a rater and government, then it will enhance benchmarks' private authority among audiences that value that affinity and detract from their private authority among audiences that do not.

I will use various forms of evidence to examine support for the ideological affinity argument as well as the epistemic and independence arguments. Before doing that, however, I provide some background information on Freedom House and its ratings.

The Origins of the *Freedom in the World* Ratings

In 1941, a group of Americans established Freedom House in New York to support the United States' involvement in World War II. After the war, FH continued to operate, supporting the emerging liberal international order and fighting for U.S. civil rights through a combination of research and advocacy. During its first decades, FH adopted a self-described "centrist" stance. According to a former executive director, "Freedom House's distinction . . . is the organization's long-term objective of broadly advancing human rights while, as necessary, pursuing short-term policies regarded as controversial by other human rights groups."⁴³

Despite its initial blend of domestic and international activities, Freedom House focused exclusively on freedom overseas after the end of the Cold War. In part, the shift occurred because the organization's bipartisan board found it increasingly difficult to agree on domestic issues. A comment at a special meeting of the executive committee in 1991 is illustrative: "The foreign [problems]

Freedom House had concentrated on in its 50 years presented clear good-vs.-evil lines, but the domestic ones invariably were legitimate right vs. legitimate right. Freedom House would be hard-pressed to define [domestic] human rights issues and to take sides without venturing into a morass that was far more a policy area than a rights question."⁴⁴

Moreover, shifting to international issues opened new funding opportunities. In the 1980s, the U.S. government began funding democracy assistance, and Freedom House was an early recipient of this aid. Government grants were lucrative at a time when the organization's finances were troubled due to fundraising challenges and the costs associated with maintaining and then selling its New York headquarters building.⁴⁵ According to Adrian Karatnycky, president of Freedom House in 1994, Freedom House "could not have weathered the transition period without [government] contracts."⁴⁶

Against this backdrop, the idea for the *Freedom in the World* ratings came from executive director Leonard Sussman. In 1972, Sussman commissioned Raymond Gastil, a social scientist affiliated with the Hudson Institute, to conduct a comparative survey of freedom. Previously, Freedom House had produced a document called the *Balance Sheet of Freedom*, an annual summary of political trends. The ambition of the comparative survey was greater: first, to create a calling card for the organization, and second, to "force some harder thinking about what freedom means" among "opinion molders."⁴⁷ Freedom House, which was then funded through special events and individual donations, paid the initial cost of \$5,500.⁴⁸

From the outset, Gastil rated countries annually in terms of their political rights and civil liberties on seven-point scales. In addition, he assigned countries overall ratings of "free," "partly free," and "not free." This basic approach remains in place. During the 1990s, Freedom House also began to publish economic and press freedom ratings. The FITW ratings were originally available for purchase. By 1999, however, the ratings were freely available via the FH website.⁴⁹

As noted earlier, authority can vary across time, space, and audience. Whether private authority exists can be identified by the adoption of standards by actors in world politics, such as when the U.S. government used the FITW ratings as an indicator for determining whether states' qualify for the MCC aid program. How much private authority exists can be identified by the degree and frequency with which actors rely on and defer to a standard. Here, I briefly use the frequency of references in U.S. government records, U.S. newspapers, and scholarly articles to consider the extent of the FITW ratings' private authority in the United States relative to other ratings. I present these data here and then later analyze them alongside other forms of empirical evidence, such as usage

patterns in a number of other countries. What will be significant, then, is how usage patterns vary across audiences and time, and what explains that variation.

For now, I present data on the ratings' usage. Usage of the ratings is an important indicator of their private authority because when actors use the ratings they are adopting them as standards for democracy. Some types of usage also indicate that a user views the ratings as legitimate, such as references that endorse the ratings (e.g., by calling them the "leading" ratings of democracy) or that omit an endorsement but also omit any criticism.

Usage patterns confirm that the FITW ratings have enjoyed more authority in the United States than other democracy ratings. I measure U.S. media usage with references to the ratings in two leading newspapers, the *New York Times* and *Washington Post*.⁵⁰ As figure 1 shows, journalists used the ratings from the start. I measure U.S. officials' usage with references in the Congressional Record, finding a similar pattern in figure 2.⁵¹ Across references in all three sources, the ratings were only criticized four times. As a comparison, I measure usage of two other ratings of democracy, Polity and the Economist Democracy Index (EDI), and find no more than one reference per year (and usually none) across all three sources. I also measure usage of a related benchmark—the Economic Freedom Index produced by the Heritage Foundation—and find less than one reference per year on average. Finally, I measure scholarly usage with references

in an all-language archive of JSTOR and EBSCOHost social science journals. Figure 3 shows a steady increase in usage of the FITW ratings as well as Polity, suggesting that academics are reliant on both democracy benchmarks. In addition to establishing that the FITW ratings have at least some private authority in the United States, these data will be used here to test the ideological affinity argument as well as other arguments about private authority.

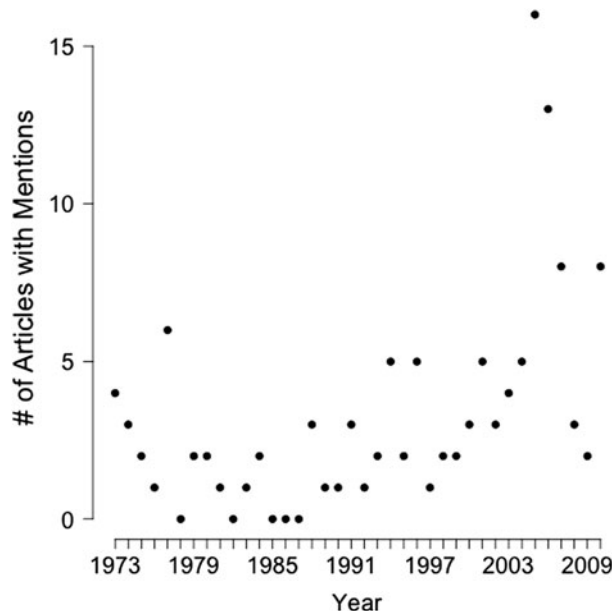
Explaining Variation in the FITW Ratings' Private Authority

Epistemic Argument

An epistemic perspective implies that benchmarks will have more private authority when experts produce them using costly, rigorous, and transparent methods. Close examination of the FITW ratings reveals the insufficiency of this argument in terms of explaining variation both in their usage over time and relative to other ratings. As I show below, U.S. journalists and officials have used the FITW ratings more than alternatives for years, even though the FITW ratings lacked rigor and transparency, including relative to alternatives.

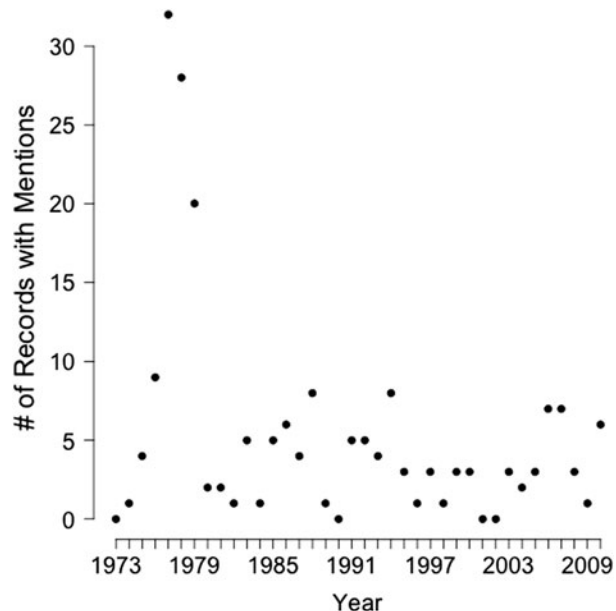
Several criticisms of the ratings' methods were common from the start. First, critics questioned their reliability,⁵² since Gastil initially worked "alone," although his wife Jeanette provided un-credited assistance. In 1990, a team of raters replaced him, but the ratings' reliability was still questioned, since Freedom House did not make public inter-coder agreement levels. Second, critics noted the

Figure 1
Authoritative references to the FITW ratings in the *New York Times* and *Washington Post*, 1973–2010



Source: ProQuest Historical Newspapers, ProQuest Newsstand.

Figure 2
Authoritative references to the FITW ratings in the Congressional Record, 1973–2010



Source: HeinOnline's Congressional Record.

ratings' unclear coding criteria.⁵³ Freedom House only released the criteria for political rights and civil liberties in 1990, and information about how the scores were derived remained private for another decade. Moreover, Gastil separated countries' numerical scores from their overall ratings as "free," "partly free," or "not free."⁵⁴ Consequently, countries' ratings did not appear to be based on clearly defined attributes. The ratings' lack of transparency arguably permitted biased judgments more easily than other benchmarks, such as Polity. Political scientists Munck and Verkuilen provide a representative critique:

Obscuring the entire exercise, very little is done to open the process of measurement to public scrutiny. Because no set of coding rules is provided, independent scholars are left in the dark as to what distinguishing features would lead a case to receive a score of 0, 1, 2, 3, or 4 points. Furthermore, the sources of information are not identified with enough precision so that independent scholars could reanalyze them. To make matters even worse, the failure to make public the disaggregated data ensures that a scholarly, public debate about issues of measurement is virtually impossible.⁵⁵

Insider accounts reinforced such critiques. According to a published article by Gastil, the ratings' "strongest claim to 'scientific' status result[ed] from the author's determination not to let current international opinion, the interests of American foreign policy, Freedom House, or personal prejudices affect survey ratings."⁵⁶ In fact, Gastil himself resisted the idea that the ratings represented an

exercise in scientifically valid descriptive inference. When he learned about research that used the ratings to understand the relationship between democracy and economic growth, he reacted skeptically and was quoted as saying, "These data were not intended to be used in economic predictions. They are not cardinal measures. They are simply rough descriptions of various political and civil rights across countries."⁵⁷

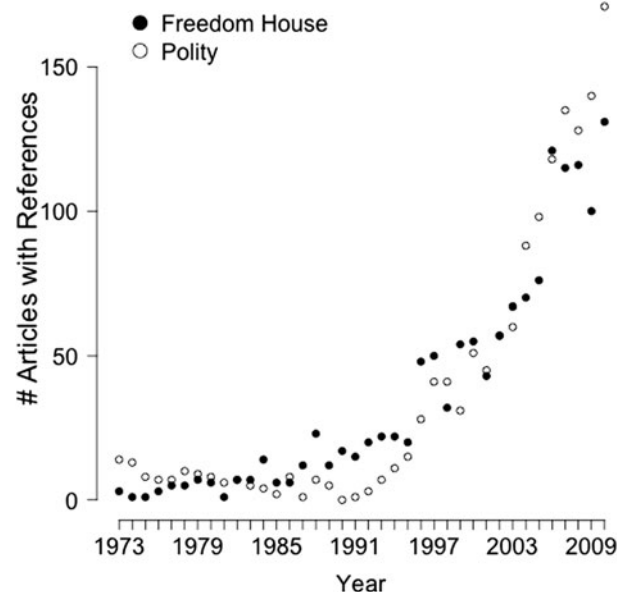
Eventually, the ratings' methodology improved. Although small adjustments occurred at various times, there were two key periods of improvement.⁵⁸ First, in 1990, a team began rating countries and a checklist of coding criteria was released. The significance of the 1990 reforms is reflected in a finding that the ratings' coding decisions diverged more from other indices, such as Polity, during the pre-1990 era.⁵⁹ Second, in 2006, the disaggregated scores (i.e., scores for the checklist items) were released. More recently, the organization has embarked on further efforts to professionalize the ratings process as part of the Governance Data Alliance.

According to the epistemic perspective, the same benchmark will have more private authority when it is produced more expertly. According to this argument, the FITW ratings should have gained authority after 1990 and again after 2006. As figures 1 and 2 show, however, American media and policy usage of the ratings did *not* increase after either 1990 or 2006. As figure 3 shows, scholarly references have increased over time. The same, however, is true for Polity, which suggests that the increased usage reflects a growing interest in democratization or usage of quantitative analyses, rather than a reaction to changes in the quality of the rating itself.

Beyond these patterns, there is considerable evidence that U.S. policymakers have consistently used the FITW ratings, including before 1990, when even the ratings' creator viewed them as subjective assessments. Starting in the 1970s, U.S. diplomats working abroad requested copies of them, and officials such as U.S. Ambassador to the United Nations (UN) Daniel Patrick Moynihan and Senators Henry Jackson and Hubert Humphrey referred to the ratings when formulating and making statements about U.S. foreign policy.⁶⁰ Moreover, in 1976, the State Department began using the ratings as a source for its human rights reports,⁶¹ and the U.S. government continued to rely on them during the 1980s and 1990s.⁶² Also during the 1990s, the U.S. Agency for International Development (USAID) began using them to assess the effects of aid on democracy.⁶³ Policymakers' reliance on the ratings culminated in the U.S. government's adoption of them in the qualification criteria for MCC aid in 2004.

The epistemic perspective implies, moreover, that benchmarks will enjoy private authority across diverse audiences—not just among U.S. audiences. The logic is that a range of users that value expertise ought to accept high-quality benchmarks. An appropriate way of testing

Figure 3
Authoritative references to the FITW ratings and Polity in academic articles, 1973–2010



Note: References to Polity, Polity II, Polity III, and Polity IV are combined.

this implication is to consider usage in the United Kingdom, another English-speaking country that promotes democracy abroad and might be expected to have similar preferences over expertise. To examine British usage, I searched two sources—*The Guardian* and the House of Commons' Parliamentary Papers—and found on average less than one reference per year between 1973 and 2010.⁶⁴ Moreover, whereas “Freedom House” appears on the same page as “United States Agency for International Development” 3,720 times on U.S. government websites, it appears on the same page as “Department for International Development” (DFID, the British equivalent of USAID) only 150 times on U.K. government websites.⁶⁵ DFID instead suggests that its staff rely on indicators such as election observers' reports and public opinion to assess the efficacy of their aid.⁶⁶ These patterns show that American media and policymakers use the FITW ratings significantly more than do British media and policymakers.

Moreover, few other democracy-promoting states and IOs use the FITW ratings, which is again inconsistent with the epistemic argument. In fact, the ratings are so unpopular at the UN that the UN Development Programme released a rival Human Freedom Index in 1991. An exception to IOs' general disuse of the FITW ratings is the World Bank, which has used them as a source for its Worldwide Governance Indicators since 1996. However, the ratings' role there may be an exception that proves the rule; since the United States plays such an important role in World Bank governance, the ideological affinity between the FITW ratings and U.S. elites that I will discuss may have contributed to their use.

A final observable implication of the epistemic perspective is that challenges to the authority of benchmarks will prompt raters to improve their methodology. As previously noted, Freedom House *has* improved its methodology. Challenges to the ratings' authority were not, however, a sufficient cause for improvements. As early as 1972, Sussman raised concerns about the ratings' methods, writing to Gastil about “the necessity for some backup information in this office justifying your assignments of political and civil status.”⁶⁷ A few months later, Sussman repeated his concerns, asking for a memo that answered the following question: “how do you derive the ratings from the checklist?”⁶⁸ Gastil did not produce the coding criteria in response to either request, later describing his approach as relying on “hunches and intuitions” and “a loose, intuitive rating system for levels of freedom or democracy, as defined by the traditional political rights and civil liberties of the Western democracies.”⁶⁹

Instead, the key improvements in the ratings' methodology coincided with other events. Gastil retired in 1989, following Sussman's retirement in 1988. When the new executive director R. Bruce McCollm reviewed Gastil's work, he discovered major problems. He conveyed

his concerns to Sussman after an outgoing presentation by Gastil:

If what we saw Friday is really all there is . . . Freedom House and I am in a difficult position concerning defending the past Surveys. The emperor has no clothes from an academic point of view . . . We can not [sic] sustain a fraud or cloak the Survey in academic obscurantism. This year's Survey should have an open methodology, which for the sake of the reading public we are putting at the end of the book. We will clearly state our assumptions as well as “subjective elements” in the criteria and what criteria are weighted more than others. The past Surveys gave greater weight to free and fair elections and freedom of expression, even though that was never explicitly stated. (In fact, the Survey was never really developed beyond the first 4 items on the Checklist).⁷⁰

McCollm's comments indicate that anticipated challenges to the ratings' authority encouraged the organization's 1990 reforms. Yet the immediate cause of the reforms was leadership turnover.

The improvements that Freedom House implemented in 2006 loosely coincided with its attempts to ensure that the ratings would be used in the MCC initiative. At a 2002 board meeting, Arch Puddington, the Vice President of Research, made a presentation that highlighted “some publication changes being discussed, some encouraged by the prospect of the Millennium Fund [MCC],” which included “reevaluating the methodology of the survey [and] expanding the methodology team.”⁷¹ Steady criticisms of the ratings' quality did not prompt the organization to significantly change its methods. Instead, a new opportunity for policy influence was what led the organization to enact further changes.

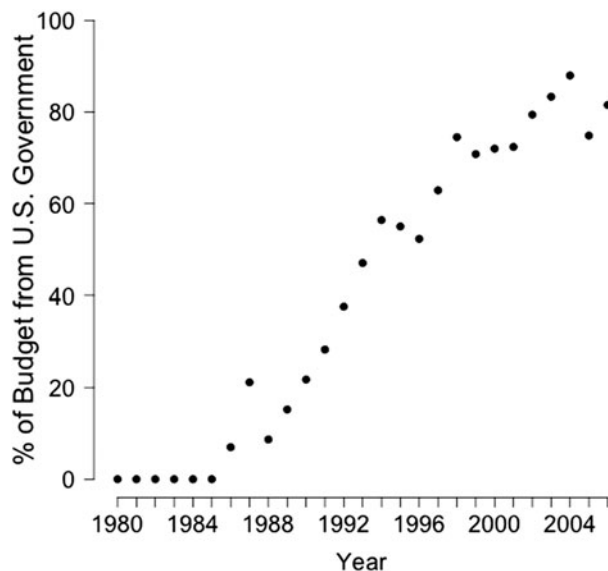
Independence Argument

The independence perspective implies that benchmarks produced by independent raters will have more authority. Freedom House, with its reliance on U.S. government funding, does not fit the image of an independent rater, particularly when compared with EDI and Polity, which have no discernible government connections. And yet, these more independent benchmarks are not used as frequently as the FITW ratings in the United States.

With regard to users, the independence perspective implies that diverse state and IO audiences will regard benchmarks as authoritative. The logic is that so long as audiences share preferences for independent raters they will use similar benchmarks. Yet as described here, the FITW ratings have more authority in the United States than in other democracy-promoting states and IOs. This pattern is inconsistent with the independence argument's prediction.

The independence perspective further implies that government funding will reduce benchmarks' private authority since it may indicate a lack of NGO independence. As figure 4 shows, Freedom House has relied heavily on U.S. government grants in terms of support for its overall budget since the mid-1990s. Yet, according to

Figure 4
Proportion of overall FH budget funded by the U.S. Government, 1980–2006



Source: From boxes 2 (Folders 17 and 19), 148 (no folder), and 181 (no folder), FH Records.

the data in figures 1, 2, and 3, the organization's increased reliance on government funding is not correlated with a decrease in the ratings' authority among U.S. media or government audiences or scholars.

Finally, the independence perspective implies that raters will avoid government connections. Because it posits that independence from competing interests promotes authority, it implies that raters will protect their independence. Consistent with that idea, Freedom House produces the FITW ratings without government funding. Moreover, FH staff worry about how their reliance on government funding affects the organization's reputation. A 2003 report noted, for example, "the current Freedom House public—private funding imbalance directly impacts perceptions of Freedom House's independence and ability to do good work."⁷²

Yet Freedom House has accepted increasing amounts of government funding for non-FITW activities, as figure 4 shows. Moreover, individuals within the organization have long enjoyed informal relationships with the U.S. government, given that board members and staff often come from or subsequently enter government. Examples include David Kramer (FH president and Assistant Secretary of State), Max Kampelman (FH board member and U.S. Ambassador), Jeanne Kirkpatrick (FH board member and U.S. Ambassador), Thomas Melia (FH deputy executive director and Deputy Assistant Secretary of State), and Jennifer Windsor (FH executive director and USAID Deputy Assistant Administrator). Thus, it cannot be argued that the organization had altogether avoided

government connections to preserve its image as an independent rater.

Ideological Affinity Argument

The ideological affinity perspective implies that benchmarks will have more private authority when they are ideologically aligned with users and when they are adopted by powerful users. I argue that the FITW ratings have been more ideologically aligned with U.S. foreign policy than alternative ratings and that this alignment has led to their relatively strong private authority in the United States as well as subsequent authority elsewhere. I consider the evidence related to: (1) how democracy is defined; (2) how countries are coded; (3) patterns of the FITW ratings' usage; (4) the effects of government funding; and (5) the origins of shared ideas.

Concept Definition: The ideological affinity perspective assumes that benchmarks' core concepts reflect raters' ideological commitments. The evidence suggests that the FITW ratings have long reflected shared ideas about democracy with U.S. foreign policy elites. To support this claim, I consider how U.S. foreign policy elites conceptualize democracy and the extent to which their ideas match the ideas of Freedom House and other raters.

Although it is difficult to operationalize "ideology," the democracy promotion literature has made important progress in identifying how U.S. foreign policy elites conceptualize democracy. The consensus is that "there is still a liberal hegemonic consensus at the heart of democracy promotion,"⁷³ with liberal democracy understood as involving "a distinctly liberal concept of man and society; values of freedom and liberty as key democratic values; a representative parliamentary system as the key site and institution of democratic politics; and a liberal capitalist model of economics as a crucial underpinning of liberal democracy."⁷⁴ This conceptualization of democracy can be contrasted with the core principles and institutions emphasized by alternative approaches, including majoritarian, participatory, deliberative, and egalitarian conceptualizations.⁷⁵ Researchers have identified U.S. democracy promotion's liberalism in diverse studies, including critical analyses of U.S. foreign policy and U.S.-led IOs,⁷⁶ close readings of the history and rhetoric of U.S. democracy promoters,⁷⁷ case studies of presidents' personal records,⁷⁸ and analysis of elite interviews.⁷⁹ Moreover, Kurki has found that U.S. democracy promotion differs from European democracy promotion, with European actors being more open to conceptualizing democracy in participatory and egalitarian ways.⁸⁰

Despite this overall liberal consensus, there are notable differences among U.S. democracy promoters about how to conceptualize democracy. For example, certain U.S. elites have become more willing to emphasize "economic justice and inequality" as elements of democracy promotion.

However, Kurki finds that this shift represents a change *within* liberal models rather than a shift to non-liberal concept of democracy.⁸¹ She suggests that the reason for the continued liberal consensus has to do with the acceptance of liberal concepts as “common sense” among democracy promoters.⁸² Supporting this point of view is the remarkable stability of USAID’s definition of democracy. The definition developed in 1990 and still in use today emphasizes elections and political processes, rule of law, governance, and civil society.⁸³ As Lincoln Mitchell explains with an emphasis on American democracy promotion institutions:

It is not possible that everybody working in democracy promotion . . . would have the same definition of democracy, but without a basic shared understanding, democracy work would be difficult. Most people would agree on a few basic components of democracy, such as freedoms of speech and assembly; free, fair, and competitive elections; and the need for a vibrant civil society.⁸⁴

Those components fit the liberal model, with its emphasis on individual rights, constraints on political power via elections, and interest groups.⁸⁵ In contrast, Mitchell notes, “the U.S. government definition of democracy . . . does not devote any attention to issues such as equality and opportunity, or a gamut of economic concerns,”⁸⁶ hallmarks of more egalitarian models.⁸⁷

Freedom House has similarly adopted a liberal conceptualization of democracy. From the start, staff were attuned to the possibility of multiple definitions of democracy; as Sussman wrote in a letter to Gastil in 1973, “suppose a communist analyst were to undertake a study like ours from the Marxist point of view; would not the black become white and vice versa?”⁸⁸ Indeed, a 1974 feature in *Worldview* magazine critiqued FITW for downplaying egalitarian concerns, with one author noting that FITW did not accurately reward countries such as China, Cuba, and Tanzania for progress in terms of “social or communal freedom”⁸⁹ and another faulting it for ignoring the struggle for “social justice in the developmental mode.”⁹⁰ Gastil responded by emphasizing that not all good things are dimensions of “freedom” and advocated for an essentially liberal conceptualization of democracy. He wrote, “As Isaiah Berlin, among others, has pointed out, freedom must be defined in largely negative terms in order to avoid its misuse by tyrannies . . . For this reason I have, in the spirit of J.S. Mill, drawn fairly tight lines around an essentially individualistic definition of freedom in the political context.”⁹¹

This philosophical orientation had meaningful implications for the FITW reports. The ratings’ checklists have consistently featured the characteristics of liberal democracy, including:

- (1) Centrality of the value of liberty.
- (2) Centrality of civil and political rights to the detriment of socio-economic rights.
- (3) Tendency to the formal rather than substantive acknowledgement of rights.

- (4) Declination of freedom mainly in negative terms (freedom from), especially referring to government intervention, and to individual (rather than social) protection, above all in the market sphere. . .
- (5) Gradual vanishing or lack of relevance of the value of equality.⁹²

In these ways, the FITW ratings’ underlying conceptualization of democracy matches the prevailing conceptualization of democracy among U.S. foreign policy elites. To further substantiate this claim, table 2 lists the first four checklist items used by Gastil,⁹³ which as noted earlier were the items given the most weight historically by coders, and matches them with a general model of democracy according to the Varieties of Democracy (V-Dem) project. Whereas electoral democracy is treated by V-Dem as embodying the “fundamental” component of democracy, each other conception (liberal, majoritarian, etc.) is associated with distinctive attributes.⁹⁴ As table 2 shows, each of the main FITW checklist items are associated either with the fundamental traits of electoral democracy or the distinctive traits of liberal democracy according to the most closely aligned V-Dem indicators. Thus, we can conclude that the FITW ratings generally emphasize the traits of democracy that are important to U.S. policymakers.

The freedom ratings’ liberal approach to democracy is not only consistent with U.S. policymakers’ approach to democracy, but it also distinguishes them from some other indicators. In particular, the Economist Democracy Index was created with the intention of providing a “thicker” definition of democracy, moving beyond political rights and civil liberties to account for equality, quality of governance, and the extent of political participation.⁹⁵ The EDI consistently rates some countries, such as Kazakhstan and Russia, as more democratic than does FITW because those countries do relatively well in terms of “political participation.”⁹⁶ As a consequence of the EDI’s more participatory conceptual framework, it is a worse match for American democracy promoters than FITW.

Coding Countries: The ideological affinity perspective assumes that raters code countries in ways that reflect their ideological commitments. Coding inevitably involves subjective judgments. When suppliers and potential users of benchmarks share common ideas, the subjective judgments are more likely to jibe. The evidence is consistent with this assumption.

As Mitchell observes, democracy promoters often conflate a country’s level of democracy with its “support for U. S. foreign policy” or its “orientation toward the West.”⁹⁷ This conflation is part of a long-standing process in which “the coding rules defining democracy are better understood as a time-bound product of America’s historical international circumstances than as [a] timeless exogenous force.”⁹⁸ Given these dynamics, coding countries in ways that fit U.S.

Table 2
Freedom House Checklist items and conceptions of democracy

Political Rights Indicators	
Freedom House Indicator	Associated V-Dem High Principle
1. Chief authority recently elected by a meaningful process	Electoral*
2. Legislature recently elected by a meaningful process	Electoral
3. Fair election laws, campaigning opportunity, polling and tabulation	Electoral
4. Fair reflection of voter preference in distribution of power; parliament, for example, has effective power	Liberal
Civil Liberties Indicators	
Freedom House Indicator	Associated V-Dem High Principle
5. Media/literature free of political censorship	Electoral
6. Open public discussion	Electoral
7. Freedom of assembly and demonstration	Liberal
8. Freedom of political and quasi-political organization	Electoral

*The associated V-Dem indicators linked to each Freedom House Indicator in the table are: (1) v2x_accex; (2) v2x_accex, v2xel_frefair; (3) v2xel_frefair; (4) v2xlg_legcon; (5) v2x_freexp_thick; (6) v2x_freexp_thick; (7) v2xcl_rol; and (8) v2x_frassoc_thick.

democracy promoters’ assumptions about American allies and non-allies is important for private authority.

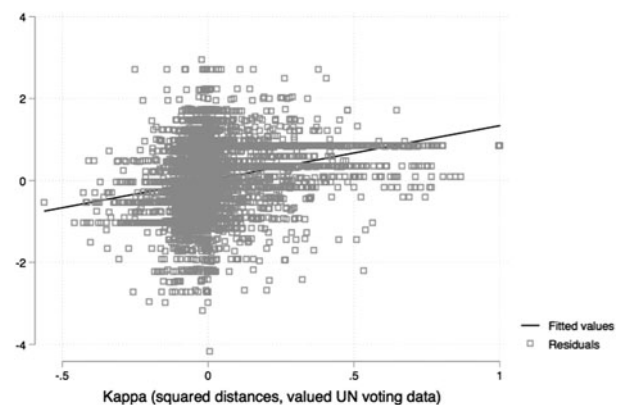
A comparison with other benchmarks shows that countries with an affinity with the United States—measured in terms of voting patterns at the United Nations⁹⁹—received consistently better ratings on average from FITW than Polity, which is another index with a liberal conceptualization of democracy. I demonstrate this pattern in figure 5 by plotting the residuals from a regression of countries’ FITW ratings on Polity scores, for which a positive value indicates a better FITW rating than was predicted. It shows that positive residuals are more likely for countries with a greater affinity with the United States. This relationship exists in every decade of the FITW ratings’ existence, which suggests that ideological affinity has existed from the start and was not caused by Freedom House’s later dependence on government funding.

Analyses of individual cases support the general trend illustrated in figure 5. There are a number of widely-cited examples of a FITW bias in favor of U.S. allies when compared to other democracy measures. They include relatively low FITW scores for Russia since the end of the Cold War and for Nicaragua under Sandinista rule, and relatively high FITW scores for junta-led El Salvador.¹⁰⁰ In some cases, of course, these differences may be the result of differences in how FITW and Polity conceptualize democracy. Yet it can also be that raters draw different conclusions based on the same information. This is what Pemstein et al. argued was the case when Polity rewarded Russia in 1998 with an improved rating when the Duma rejected President Boris Yeltsin’s nomination for Prime Minister, whereas FITW did not.¹⁰¹ Even when coders have well-specified coding rules, there are many ways that commonsense ideas about what democracy is can shape their decisions. The analysis in figure 5 suggests that

FITW ratings are more likely to incorporate U.S. foreign policy values than the Polity ratings.

Interviews further support the idea that FITW raters incorporate ideas that are consistent with American foreign policy when coding countries. A staff person at Freedom House’s regional Middle East office told me that the office sometimes found it difficult to work with local NGOs because of skepticism there about FITW coding Israel as “free.”¹⁰² Although debates about specific countries’ ratings are common, FH staff do not debate whether to code Israel as less than free according to the documents

Figure 5
Residuals from a regression of FITW scores on Polity scores, plotted against foreign policy similarity scores, 1972–2004



Note: For comparison to Polity, the two FITW scores (political rights and civil liberties) were averaged and rescaled, with 1 as the lowest score and 7 as the highest score. This analysis is modeled on Steiner 2014, 14.

in the organization's archive, perhaps because doing so is outside of their ideological framework. In contrast, for many of Freedom House's potential local partners in the Middle East, coding Israel as "not free" is commonsensical, given its treatment of Palestinians.¹⁰³ Most potential users are probably not reading Freedom House's checklists or trying to assess their conceptual foundations—but they may be scanning the reports to make sure they match their perceptions of specific countries.

Patterns of Usage: If the FITW ratings gained relatively strong private authority in the United States because of their ideological affinity with U.S. elites, we would not expect all states and IOs engaged in democracy promotion to use the ratings. Instead, we would expect audiences in the United States to use the ratings most. Thus, the ideological affinity perspective offers an explanation for the fact that U.S. journalists and government officials use the FITW ratings more than journalists and officials in other democracy promoting states.

The argument also suggests that when benchmarks acquire private authority via ideological affinity with powerful states, they become important for weaker states. Consequently, weaker state audiences may adopt a benchmark even if they do not embrace its concepts or coding. This empirical expectation is also supported. The FITW ratings have long interested officials in hybrid regimes and autocracies. For example, after its Carnation Revolution in 1974, Portugal began lobbying to improve its score. In 1976, Francisco de Sá Carneiro, head of the country's Social Democratic Party, went to New York to meet with Sussman. He argued that Freedom House was wrong to rate the Azores—one of Portugal's autonomous regions—"not free," as it was at least as free as several other territories that had been rated better.¹⁰⁴ In 1977, the Portuguese Embassy to the United States made a similar case about the Azores and Madeira.¹⁰⁵ Portugal's umbrage at its FITW ratings was, moreover, common. Within a few years of the ratings' creation, countries including Guyana, India, and the Philippines had complained to Sussman.¹⁰⁶ Appeals for higher ratings continued in the ensuing years,¹⁰⁷ although according to Sussman their frequency increased after the creation of the MCC.¹⁰⁸

Countries' ratings have also been a topic of discussion among U.S. diplomats. Such discussions do not typically become public; however, some diplomatic cables have been released and provide examples. On the one hand, U.S. diplomats have used the ratings to encourage countries to democratize (e.g., embassy officers reported after a meeting with Tajik officials, "We will continue to use the MCC scorecard as an instrument to help prod Tajikistan to reform, but recognize that aid conditioned on policy reform is a hard sell here").¹⁰⁹ On the other hand, foreign leaders have argued that their countries' poor ratings should not be taken seriously (e.g., President Aliyev

of Azerbaijan told Assistant Secretary of State Barry Lowenkron in 2007, "Partners need to be open and work closely on all the elements [of the relationship and] judge the situation not only based on Freedom House or Transparency International's assessments").¹¹⁰

References to the ratings in the global media also suggest a keen interest in countries targeted by democracy aid. Figure 6 graphs references to the ratings in articles from World News Connection, a database that translates global news articles into English.¹¹¹ Although World News Connection is not a fully representative sample of global media, as it translates articles of potential interest to the U.S. government, it is the most comprehensive database of non-English newspapers. Moreover, it is unlikely that articles were selected for inclusion because of FITW references. References in the global press take several forms: (1) boasts about positive ratings (e.g., in Taiwan in 2005, "Taiwan Ranks Among Top in Annual Freedom House Survey"); (2) criticisms of negative ratings (e.g., in Russia in 2011, "Human Rights Council Chairman Disagrees with Freedom House"); and (3) pledges to reform after negative ratings (e.g., in Ukraine in 2011, "President: We Will Take Into Account Fair Criticism By Freedom House").

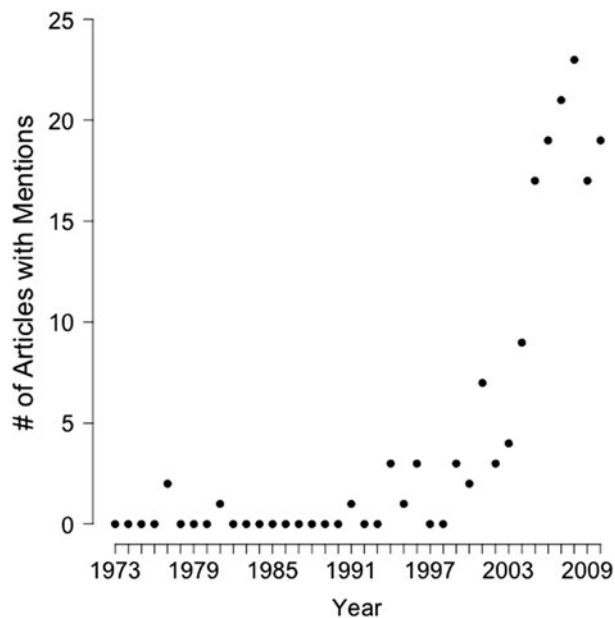
Figure 6 shows that global attention to the ratings has increased significantly since 2001. This increase is expected because the United States expanded democracy assistance after 2001¹¹² and, as noted above, adopted the ratings as criteria for MCC aid decisions in 2004. Moreover, about 30 percent of the references criticize the ratings, in contrast to the less than 1 percent in the U.S. media. That pattern underscores the ratings' relatively strong private authority in the United States, as my affinity argument predicts. At the same time, given the ratings' usage in the United States, weaker states ought to regard the ratings as important, if not necessarily as legitimate.

Government Funding: The ideological affinity approach suggests that government funding will not undermine benchmarks' authority among audiences that are using the ratings. After all, such funding could indicate shared interests, but it could also reflect a pre-existing ideological affinity. Thus, my argument is consistent with the fact that the FITW ratings' authority in the United States preceded the organization's dependence on government funding.

Origins of Ideological Affinity: Why were U.S. policy elites' ideas about democracy reflected in the FITW ratings in the first place? Although it is not my goal to establish the causes of ideological affinity, it is relevant to briefly consider its origins.

Organizational culture often derives from staff members' backgrounds. For example, an education in elite American economics departments is important for how

Figure 6
References to the FITW ratings in World News Connection articles, 1973–2010



staff at the International Monetary Fund approach their work.¹¹³ Leaders at Freedom House generally fit the mold of “establishment internationalism,” which involves several cultural and social elements, including advanced education at Ivy League universities, participation in the Council on Foreign Relations (CFR), and careers in business, law, and government.¹¹⁴ The FH executive directors on whom I found data (six) and Raymond Gastil all attended Ivy League universities, obtained graduate degrees, and worked for the government or military. In other words, they plausibly subscribed to the prevailing views about democracy held by U.S. elites. In keeping with the idea that FH is an organization firmly in the mainstream, Sussman described its mission as “to propose consensual, credible policies that can gain adherents from diverse interests.”¹¹⁵

Moreover, there is suggestive evidence that beliefs about democracy are a core element of establishment internationalism. Consider, for example, Shepherd’s comparison of the content of *Foreign Affairs* (the CFR’s flagship journal) and *International Affairs* (the flagship journal of Chatham House, the British CFR equivalent) since the 1930s.¹¹⁶ She found that democracy was one of the relatively few topics on which the two journals diverged, with *Foreign Affairs* displaying a much greater interest in political systems. Relatedly, Oren argued that American elites have promulgated a notably U.S.-centric conception of democracy over the twentieth century.¹¹⁷

As such, the origins of these shared ideas are plausibly the product both of common worldviews and more specific beliefs about democracy.

It is worth noting that all the executive FH staff as of November 2015 had previously worked for the U.S. government or for other organizations in the “democracy establishment,” the field of NGOs that implements U.S. democracy aid.¹¹⁸ As democracy promotion has professionalized, the institutions of the democracy establishment have become important places where ideas about democracy are developed and reproduced. Although linkages between practitioners and academics have long characterized U.S. democracy promotion,¹¹⁹ they have grown stronger with the field’s professionalization.¹²⁰ Though the resulting “more rigorous, evidence-based, fact-driven democracy support” field may appear neutral, it is of course still deeply—if unwittingly—informed by ideology.¹²¹

Some readers may wonder if Freedom House merely promoted a standard for democracy that had been previously articulated by the U.S. government. Though it is difficult to rule out this possibility, I identified no evidence in the organization’s archive of U.S. government officials contacting and then influencing Raymond Gastil or other FH staff in terms of how to conceptualize democracy or code countries. Although I have shown that the FITW ratings are more supportive of U.S. allies than Polity, this pattern does not prove that the ratings are an arm of U.S. foreign policy. In fact, Freedom House has occasionally sought to criticize American policymakers, such as in 2002, when various staff and board members sought to reduce the United States’ FITW rating from its traditional score of 1 (the best possible score) in protest of post-September 11, 2001 infringements on civil liberties.¹²² Thus, the evidence is most consistent with the idea that establishment internationalism, not government pressure, shaped how FH specified, expanded, and codified ideas about how to conceptualize and measure democracy.

Competing and Complementary Mechanisms

In this section, I briefly consider two alternative mechanisms. One is that the ratings possess authority owing to sequencing. The FITW ratings were one of the first comparative, global ratings of democracy.¹²³ They may enjoy relatively strong levels of private authority today because of path dependence, or the “process in which the structure that prevails after a specific moment in time . . . shapes the subsequent trajectory in ways that make alternative institutional designs substantially less likely to triumph.”¹²⁴ The preceding analysis suggests that this dynamic is at play, since U.S. media and officials continue to use the FITW ratings despite the existence of alternatives with authority in other contexts (e.g., EDI and Polity). But path dependence implies that the initial

adoption of the ratings set in motion forces that perpetuated the ratings' private authority in the United States. Contrary to that expectation, figures 1, 2, and 3 do *not* indicate that the ratings' usage is increasing, either in general or relative to other measures.

Furthermore, the U.S. government does *not* treat its use of the FITW ratings as a given. As noted above, it has long relied on them to assess aid effectiveness. But it has changed its benchmarks considerably over time. In 2008, for example, the "Master List of Standard Indicators" adopted by the Office of U.S. Foreign Assistance Resources at the State Department used eight overall benchmarks (including FITW) to assess the effectiveness of democracy aid. By 2013, none of those benchmarks were still used, having been replaced by eight alternatives, including the World Justice Project's Rule of Law Index, the Cingranelli and Richards Human Rights Dataset, and other FH indicators.

Finally, one might expect the contingent initial adoption of the freedom ratings by American media, researchers, and public officials to have launched the mechanisms of path dependence by shifting the balance of power among the relevant stakeholders such that certain constituencies supported using the FITW ratings and opposed alternatives. Yet even in the United States, where the ratings have the most private authority, the ratings' adoption has not empowered any obvious constituencies. The explanatory power of path dependence is thus limited when it comes to understanding where and why the ratings first gained some private authority in the United States and less private authority elsewhere, as well as why the ratings' private authority has had different levels of "stickiness" across difference audiences.

Another explanation is that Freedom House had a unique publicity strategy that led to the ratings' relatively strong private authority in the United States. It is well known that NGOs strategically seek to gain attention in a crowded marketplace.¹²⁵ To reach the media, FH has issued press releases, sent copies of the ratings to journalists, and hired communications staff.¹²⁶ Yet if publicity shaped usage, we would expect its publicity to have been concentrated in countries such as the United States where the ratings acquired more private authority. That was not the case. The organization's media strategy has always included Europe.¹²⁷

In contrast to that continuity, Freedom House outreach to the U.S. government increased significantly in the 1980s, coincident with its efforts to win government funding. At that time, Freedom House started to more regularly send U.S. officials the ratings and to correspond and meet with Congressional and White House staff; it also opened an office in Washington, D.C. that concentrated on "placing articles . . . in the Congressional Record."¹²⁸ Yet figure 2 does not indicate that those developments led to increased usage of the FITW ratings in Congress. Thus I conclude that

recognizing ideological affinity is necessary for understanding continuities in the private authority of the FITW ratings in the United States and variations in their private authority globally, though the organization's publicity efforts have likely also played a supportive role.

Recognizing the Ideological Roots of Private Authority

This article began with a puzzle: The *Freedom in the World* ratings enjoyed relatively strong private authority among American audiences from the start—and weaker private authority elsewhere in the world, although that is changing. The FITW ratings' early usage in the United States occurred despite limitations in terms of their epistemic quality and the existence of some plausible alternatives. The evidence presented here supports the conclusion that the ratings' affinity with the U.S. foreign policy establishment's ideas about liberal democracy and how to code countries were key to their relatively early and lasting private authority in the United States and more limited private authority elsewhere. At the same time, the uptick in usage of the ratings in the developing world after the U.S. government formally adopted them in the MCC initiative shows how ratings gain influence when they are combined with state power.

These findings have important implications for policymakers. Ironically, the ideological alignment that encourages U.S. policymakers to use the FITW ratings makes them problematic to use for certain functions. The MCC explicitly seeks to prevent states from qualifying for aid unless they make the reforms that policymakers believe are necessary for aid to succeed.¹²⁹ But if the benchmarks that are used to implement conditionality are biased in favor of U.S. allies, then the stated purpose of the program is undermined. Policymakers interested in promoting democracy should diversify their measures of democracy.

Academics working on topics related to democratization should also use the FITW ratings with care. On the one hand, the FITW ratings offer an under-appreciated measure of how U.S. elites perceive other countries' political systems. Such a measure may be appropriate for assessing whether policymakers target democracy aid according to their perceptions of countries' political systems. Thus, the FITW ratings may be especially helpful for contributions to the literature on the causes and consequences of democracy assistance.¹³⁰

On the other hand, that the FITW ratings favor countries that are aligned with the United States complicates researchers' ability to use them to make inferences about democratization. Researchers might infer that countries are democratizing when they are simply becoming better aligned with U.S. foreign policy. Moreover, using ratings that reflect a fundamentally liberal conception of democracy is not appropriate for certain academic purposes, such as testing hypotheses that are concerned

with changes in citizens' political participation or equal protection. In such cases, a participatory or egalitarian measure would be more appropriate.

A shift within political science is underway in terms of how we measure democracy, with the Varieties of Democracy project developing an approach that is "historical, multidimensional, disaggregated, and transparent."¹³¹ One of the many benefits of its approach to coding indicators of democracy in its varied forms is that it allows—and perhaps forces—researchers to think more carefully about which principles and institutions of democracy are relevant to the arguments that they seek to develop and test. After all, my examination of the FITW ratings suggests that *any* democracy measure involves ideological commitments.

In addition to contributing to policy and scholarship concerned with democratization, I advance a new argument about the ideological sources of private authority in world politics. A central contribution here is thus an original theory about what causes actors in world politics to adopt certain rules and standards. According to previous scholarship, non-state actors' authority stems from their capacity, expertise, legitimacy, or embeddedness in institutions.¹³² The same factors play important roles in previous arguments about the emergence of private standards¹³³ as well as in recent work on the political authority of ratings and rankings organizations.¹³⁴ In contrast, I offer an argument about private authority that emphasizes the effect of shared ideas and ideals among ratings' suppliers and powerful users.

Third parties now monitor states' compliance with a variety of international standards, including regarding the environment,¹³⁵ labor rights,¹³⁶ and elections.¹³⁷ Part of the rise of third-party monitors is the emergence of low-quality standards, which are influential despite not providing credible information.¹³⁸ It is difficult to understand the emergence of low-quality standards if we focus just on expertise and independence as the sources of private authority. The ideological affinity argument suggests that benchmarks need not be produced well, or independently, for academics, journalists, and policymakers to use them. Instead, they can become authoritative because powerful users share raters' ideas about concept definition and coding, and because weak actors depend on powerful other users of the ratings. In fact, the argument advanced in this article suggests that even expertly-produced or independent ratings may require some amount of ideological affinity to gain significant private authority. If my argument is correct, then the privately produced benchmarks that have the most authority globally should reflect ideas that are shared with powerful states and IOs. This proposition can and should be tested in the future.

Finally, in advancing a new argument about the sources of private authority, I drew on insights from the literature on the conceptual politics of democracy

promotion.¹³⁹ Using a variety of mainly interpretive and post-positive approaches, this literature has revealed the fundamentally liberal definition of democracy that underpins much of the American democracy promotion industry. Thus, the conceptual literature provided crucial insights into how potential state and non-state users of the FITW ratings think about democracy. This literature also suggested an empirical approach to examining the conceptual foundations of the FITW ratings.

In addition to drawing on this literature to better understand the sources of private authority in world politics, I have sought to contribute back to the conceptual literature by putting forward a positivist argument that used ideological affinity as an independent variable. The literature on democracy promotion generally falls into two categories: a critical literature that is often centrally concerned with normative questions, and a problem-solving literature that—while perhaps motivated by normative commitments—focuses on hypothesis testing.¹⁴⁰ Unfortunately, the critical literature's insights are often not well-integrated in the broader positive literature on the effects of democracy promotion. Though this article falls in the second category since it sought to develop and test an argument about how ideological affinity contributes to private authority, it has sought to take the insights of the first category seriously. Further testing this article's argument will require researchers to engage in or with conceptual analysis, identifying the core ideas that underpin how benchmarks define their fundamental concepts and code countries. The potential benefits of continuing to integrate normative and positive approaches to understand private authority in world politics are significant.

Notes

- 1 I describe and present data on the ratings' usage later in the article.
- 2 McMahon 2001, 463.
- 3 Bradley 2015, 53.
- 4 Kaufmann, Kraay, and Mastruzzi 2010, 29.
- 5 Green 2013, 6.
- 6 Guilhot 2005, 4.
- 7 For example, Mattli and Büthe 2005 and Green 2013.
- 8 For example, Hall and Biersteker 2002, 13–15.
- 9 For example, Davis, Kingsbury, and Merry 2012, 9.
- 10 As Büthe 2012 (49) notes, "academic studies, including several careful comparative analyses of alternative measures of democracy, regularly note the lack of information about the origins and aggregation process of the Freedom House data, and there appear to be virtually no independent studies of the organization."
- 11 Giannone 2010; Steiner 2016.
- 12 Steiner 2016, 7.
- 13 For example, Hobson and Kurki 2012 and Kurki 2013.

- 14 Green 2013, 6.
- 15 Ibid, 29. Non-state actors are thus *formally* distinct from states and IOs, although empirically this distinction “is not always clear-cut.” Ibid, 31–32.
- 16 Hall and Biersteker 2002, 4–5.
- 17 Avant, Finnemore, and Sell 2010a, 6.
- 18 Stroup and Wong 2016, 141.
- 19 Green 2013, 33; Avant et al. 2010b, 360.
- 20 Lake 2010, 591. See also Hurd 1999, 400–401; Hall and Biersteker 2002, 5.
- 21 Zürn, Binder, and Ecker-Ehrhardt 2012, 83.
- 22 Dahl 1957, 202–203.
- 23 Broome and Quirk 2015.
- 24 Avant, Finnemore, and Sell 2010b, 357.
- 25 Sending and Neumann 2006.
- 26 Cutler, Hauffer, and Porter 1999; Mattli and Büthe 2005, 402; Green 2013, 31.
- 27 Keck and Sikkink 1998; Avant, Finnemore, and Sell 2010a, 12.
- 28 Hyde 2012, 48–52.
- 29 Green 2013, 130.
- 30 Sending and Neumann 2006, 688.
- 31 Gourevitch and Lake 2012, 12. See also Hall and Biersteker 2002, 13–15, and Green 2013, 16.
- 32 Keck and Sikkink 1998, 30.
- 33 Gourevitch and Lake 2012, 12; Broome and Quirk 2015.
- 34 Sending and Neumann 2006, 688.
- 35 Kelley 2009, 769–771.
- 36 Coppedge et al. 2011, 248–249.
- 37 Bruner and Abdelal 2005.
- 38 Kelley 2009, 778; Hyde 2012, 46–47. Though note that other audiences may value different aspects of election monitors’ capabilities or expertise; Bush and Prather 2018.
- 39 Lake 2010, 605. See also Gray 2013.
- 40 Cooley 2015, 18.
- 41 Kurki 2013, 16.
- 42 Kelley and Simmons 2015.
- 43 Sussman 2009, 277.
- 44 Special Meeting of the Executive Committee Plus Other Board Members, August 3, 1991, p. 2, Freedom House Records, box 148 (no folder), Seeley G. Mudd Manuscript Library, Princeton University. Used by permission of the Princeton University Library. Hereafter “FH Records.”
- 45 Bush 2015, 158–160.
- 46 September 20, 1994, Board Meeting minutes, p. 6, FH Records, box 180 (no folder).
- 47 Raymond Gastil, letter to Leonard Sussman, September 2, 1971, FH Records, box 40, folder 11.
- 48 1974 Program Audit Committee Report, p. 3, FH Records, box 2, folder 5. See also Correspondence between Raymond Gastil and Leonard Sussman, FH Records, box 40, folder 11.
- 49 September 30, 1999, Board Meeting minutes, p. 5, FH Records, box 183 (no folder).
- 50 Data are from ProQuest Historical Newspapers and ProQuest Newsstand. Since analyzing the content of historical newspapers is a time-intensive endeavor, I focused on two leading newspapers. The main risk of focusing on two newspapers is that I would overestimate or underestimate the use of the FITW ratings *in certain years or relative to other types of benchmarks*, but I have no reason to believe that my inferences are biased.
- 51 Data are from HeinOnline’s Congressional Record. Figure 2 reflects especially frequent references to Freedom House from 1977 to 1979. This is because Panama’s rating was a frequent talking point in Congressional debates about the United States handing over control of the Panama Canal.
- 52 Scoble and Wiseberg 1981; Bollen 1986, 587–589.
- 53 Bollen 1986, 584–585.
- 54 Gastil 1990, 28.
- 55 Munck and Verkuilen 2002, 21.
- 56 Gastil 1990, 26.
- 57 Klitgaard 1994, 7. Others at Freedom House shared Gastil’s view. In 1983, for example, the executive committee worried about how the controversial ratings of Chile, Poland, South Africa, and Yugoslavia might be perceived. One of their conclusions was that the survey assessments were “educated judgments” and not “scientific.” See Executive Committee Meeting Minutes, December 1, 1983, FH Records, box 3, folder 15.
- 58 R. Bruce McCollm, memo on “Survey 1990,” July 23, 1989, FH Records, box 153 (no folder); Interview with staff person at Freedom House, by telephone, November 2, 2015.
- 59 Steiner 2016, 9.
- 60 U.S. Embassy Argentina, December 29, 1976, Freedom House Report on Argentina, WikiLeaks Cable 1976BUENOS080381_b; U.S. Embassy India, April 7, 1975, Freedom House Publication, WikiLeaks Cable 1975CALCUT00800_b; Leonard Sussman, Memo to Gerald Steibel, January 21, 1976, FH Records, box 2, folder 7.
- 61 Joe Wilson, letter to Leonard Sussman, April 29, 1976, FH Records, box 41, folder 4.
- 62 See, for example: R. Bruce McCollm, letter to the Board of Trustees, August 6, 1987, FH Records, box 74, folder 6; “Freedom House Activities, 1990–1991: Comparative Survey of Freedom,” pp. 1–2, FH Records, box 148 (no folder); and September 23, 1997, Board Meeting Minutes, p. 1, 6, FH Records, box 183 (no folder).
- 63 McMahon 2001, 463.
- 64 Newspaper data are from ProQuest Historical Newspapers and ProQuest Newsstand.

- 65 As of November 1, 2015. The searches I performed were completed in Google and used the following terms: (1) “Freedom House” and “United States Agency for International Development” site:.gov; and (2) “Freedom House” and “Department for International Development” site:.gov.uk.
- 66 DFID 2011, 23.
- 67 Leonard Sussman, letter to Raymond Gastil, December 19, 1972, FH Records, box 40, folder 11.
- 68 Leonard Sussman, letter to Raymond Gastil, April 12, 1973, FH Records, box 40, folder 11.
- 69 Gastil 1990, 25–26.
- 70 R. Bruce McColm, memo on “Survey 1990,” July 23, 1989, FH Records, box 153 (no folder).
- 71 May 15, 2002, Board Meeting minutes, pp. 4–5, FH Records, box 181 (no folder).
- 72 Freedom House Fundraising Strategy, October 21, 2003, p. 1, FH Records, box 181 (no folder).
- 73 Kurki 2013, 18.
- 74 Ibid, 30.
- 75 Coppedge et al. 2011, 253.
- 76 Robinson 1996.
- 77 Monten 2005; Kurki 2010.
- 78 Green 2012.
- 79 Bunce and Wolchik 2012.
- 80 Kurki 2010, 380; 2013, 147. European actors’ weaker commitment to a liberal conceptualization of democracy is why I expect the FITW ratings to have less private authority among European actors than American actors.
- 81 Kurki 2013, 145.
- 82 Ibid, 217. See also Snider and Faris 2011, 51.
- 83 Interview with former USAID official, May 4, 2010.
- 84 Mitchell 2016, 111.
- 85 Coppedge et al. 2011, 254.
- 86 Mitchell 2016, 112.
- 87 Coppedge et al. 2011, 254.
- 88 Leonard Sussman, letter to Raymond Gastil, March 13, 1973, FH Records, box 40, folder 12.
- 89 Quigley 1974, 40.
- 90 Goulet 1974, 40.
- 91 Gastil 1974, 42.
- 92 Giannone 2010, 78.
- 93 Gastil 1990, 30, 36.
- 94 Coppedge et al. 2016.
- 95 Kekic 2007, 1–2. Interview with person involved with the creation of the EDI, by telephone, June 10, 2015.
- 96 Kekic 2007, 5–6; Gunitsky 2015, 138, 148.
- 97 Mitchell 2016, 112.
- 98 Oren 1995, 153.
- 99 Häge 2011.
- 100 Goldstein 1986, 48; Mainwaring, Brinks, and Pérez-Liñán 2001, 54; Gunitsky 2015.
- 101 Pemstein, Meserve, and Melton 2010, 429.
- 102 Interview with Program Officer at Freedom House, Amman, Jordan, May 27, 2012.
- 103 Israel’s rating was not discussed at any length in the FH Records reviewed for this paper. Documents in the FH Records demonstrate that board and staff members frequently debate how to rate specific countries, such as Chile, Poland, South Africa, and Yugoslavia in the early 1980s.
- 104 Leonard Sussman, letter to Raymond Gastil, March 9, 1976, FH Records, box 40, folder 12.
- 105 Luis A. de Sousa, letter to Leonard Sussman, January 14, 1977, FH Records, box 40, folder 12.
- 106 Leonard Sussman, letter to William Evan, April 7, 1976, FH Records, box 40, folder 12.
- 107 See, for example: T.T.B. Koh, letter to Leonard Sussman, January 16, 1979, FH Records, box 41, folder 1; Leonard Sussman, letter to Board of Trustees and Advisory Council, September 8, 1980, FH Records, box 7, folder 5; Frank Calzon, memorandum to R. Bruce McColm, Re: *Freedom in the World: 1990–1991*, Chapter on Argentina, April 23, 1992, FH Records, box 75, folder 1.
- 108 Sussman 2009, 276.
- 109 The FITW ratings were also specifically cited in the report. U.S. Embassy in Tajikistan, November 14, 2007, “MCC Demarche—Tajiks Still Missing the Point,” WikiLeaks Cable 07DUSHANBE1612_a.
- 110 U.S. Embassy in Azerbaijan, January 9, 2007, “President Aliyev, DRL A/S Lowenkron Discuss Bilateral Relations, Democratic Development,” WikiLeaks Cable 07BAKU37_a.
- 111 Other benchmarks are referenced less frequently in this database, as would be expected since the U.S. government has not adopted them as benchmarks. For example, I found only five total references to the EDI and 53 total references to the Economic Freedom Index, in comparison to 235 references to the FITW ratings.
- 112 Bush 2015, 8.
- 113 Chwieroth 2009; Nelson 2014.
- 114 Busby and Monten 2008, 452.
- 115 Sussman 2002, 6.
- 116 Shepherd 2011, 42.
- 117 Oren 1995.
- 118 Bush 2011, 2015.
- 119 Guillhot 2005.
- 120 Bush 2015, 47.
- 121 Kurki 2016, 1.
- 122 September 26, 2002, Board Meeting Minutes, p. 5, FH Records, box 181 (no folder).
- 123 According to Gastil, however, they were “hardly the first.” The antecedents he cites include surveys of democracy by James Bryce, Arthur Banks and Robert Textor, Robert Dahl, Dankwart Rustow, and Charles Humana. Another contemporary was Polity,

- a project begun by Ted Robert Gurr in the late 1960s. See Gastil 1986, 5.
- 124 Fioretos 2011, 376.
- 125 Bob 2005.
- 126 Leonard Sussman, letter to Raymond Gastil, December 19, 1972, FH Records, box 40, folder 11; and R. Bruce McColm, memo on “Freedom House’s Image,” 1982, FH Records, box 3, folder 14.
- 127 Leonard Sussman, Maintaining and Developing the Map of Freedom and the Comparative Survey, no date (likely 1973), p. 1, FH Records, box 40, folder 11.
- 128 R. Bruce McColm, letter to the Board, August 6, 1987, FH Records, box 74, folder 6.
- 129 Girod, Krasner, and Stoner-Weiss 2009.
- 130 Bush 2016, 375.
- 131 Coppedge et al. 2011, 248.
- 132 Avant, Finnemore, and Sell 2010a, 11.
- 133 Bütthe and Mattli 2011; Green 2013.
- 134 Cooley 2015, 24–25.
- 135 Green 2013.
- 136 Nooruddin and Sokhey 2012.
- 137 Kelley 2009; Hyde 2012; Bush and Prather 2017.
- 138 Kelley 2009; Nooruddin and Sokhey 2012.
- 139 For example, Hobson and Kurki 2012; Kurki 2013.
- 140 Kurki 2013, xii.

References

- Avant, Deborah D., Martha Finnemore, and Susan K. Sell. 2010a. “Who Governs the Globe?” In *Who Governs the Globe*, ed. Deborah D. Avant, Martha Finnemore, and Susan K. Sell. Cambridge: Cambridge University Press.
- . 2010b. “Conclusion: Authority, Legitimacy, and Accountability in Global Politics.” In *Who Governs the Globe*, ed. Deborah D. Avant, Martha Finnemore, and Susan K. Sell. Cambridge: Cambridge University Press.
- Bob, Clifford. 2005. *The Marketing of Rebellion: Insurgents, Media, and International Activism*. New York: Cambridge University Press.
- Bollen, Kenneth A. 1986. “Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984.” *Human Rights Quarterly* 8(4): 567–91.
- Bradley, Christopher G. 2015. “International Organizations and the Production of Indicators: The Case of Freedom House.” In *The Quiet Power of Indicators: Measuring Governance, Corruption, and the Rule of Law*, ed. Sally Engle Merry, Kevin E. Davis, and Benedict Kingsbury. Cambridge: Cambridge University Press.
- Broome, André and Joel Quirk. 2015. “Governing the World at a Distance: The Practice of Global Benchmarking.” *Review of International Studies* 41(5): 819–841.
- Bruner, Christopher M. and Rawi Abdelal. 2005. “To Judge Leviathan: Sovereign Credit Ratings, National Law, and the World Economy.” *Journal of Public Policy* 25(2): 191–217.
- Bunce, Valerie J. and Sharon L. Wolchik. 2012. “Concepts of Democracy Among Donors and Recipients of Democracy Promotion: An Empirical Pilot Study.” In *The Conceptual Politics of Democracy Promotion*, ed. Christopher Hobson and Milja Kurki. London: Routledge.
- Busby, Joshua W. and Jonathan Monten. 2008. “Without Heirs? Assessing the Decline of Establishment Internationalism in U.S. Foreign Policy.” *Perspectives on Politics* 6(3): 451–72.
- Bush, Sarah Sunn. 2011. “International Politics and the Spread of Quotas for Women in Legislatures.” *International Organization* 65(1): 103–37.
- . 2015. *The Taming of Democracy Assistance: Why Democracy Promotion Does Not Confront Dictators*. Cambridge: Cambridge University Press.
- . 2016. “When and Why Is Civil Society Support ‘Made-in-America’? Delegation to Non-State Actors in American Democracy Promotion.” *Review of International Organizations* 11(3): 361–85.
- Bush, Sarah Sunn and Lauren Prather. 2017. “The Promise and Limits of Election Observers in Building Election Credibility.” *Journal of Politics* 79(3): forthcoming.
- . 2018. “Who’s There? Election Observer Identity and the Local Credibility of Elections.” *International Organization* 72(3): forthcoming.
- Bütthe, Tim. 2012. “Beyond Supply and Demand: A Political–Economic Conceptual Model.” In *Governance by Indicators: Global Power through Classification and Rankings*, ed. Kevin Davis, Angelina Fisher, Benedict Kingsbury and Sally Engle Merry. New York: Oxford University Press.
- Bütthe, Tim and Walter Mattli. 2011. *The New Global Rulers: The Privatization of Regulation in the World Economy*. Princeton, NJ: Princeton University Press.
- Chwieroth, Jeffrey M. 2009. *Capital Ideas: The IMF and the Rise of Financial Liberalization*. Princeton, NJ: Princeton University Press.
- Cooley, Alexander. 2015. “The Emerging Politics of International Rankings and Ratings: A Framework for Analysis.” In *Ranking the World: The Politics of International Rankings and Ratings*, ed. Alexander Cooley and Jack Snyder. New York: Cambridge University Press.
- Coppedge, Michael, et al. 2011. “Conceptualizing and Measuring Democracy: A New Approach.” *Perspectives on Politics* 9(2): 247–67.
- . 2016. “V-Dem Codebook v6.” Varieties of Democracy (V-Dem) Project. Available at <https://www.v-dem.net/files/42/Codebookv6.pdf>; accessed April 3, 2017.
- Cutler, A. Claire, Virginia Haufler, and Tony Porter. 1999. “Private Authority and International Affairs.” In

- Private Authority and International Affairs*, ed. A. Claire Cutler, Virginia Haufler, and Tony Porter. Albany: SUNY Press.
- Dahl, Robert A. 1957. "The Concept of Power." *Systems Research and Behavioral Science* 2(3): 201–15.
- Davis, Kevin E., Benedict Kingsbury, and Sally Engle Merry. 2012. Introduction: Governance by Indicators. In *Governance by Indicators: Global Power through Classification and Rankings*, ed. Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry. Oxford: Oxford University Press.
- Department for International Development (DFID). 2011. "Briefing Note." On file with author.
- Fioretos, Orfeo. 2011. "Historical Institutionalism in International Relations." *International Organization* 65(2): 367–99.
- Gastil, Raymond D. 1974. "Reply to the Critics." *Worldview* 17(11): 42–3.
- . 1986. *Freedom in the World: 1985–1986*. Westport, C.T.: Greenwood Press.
- . 1990. "The Comparative Survey of Freedom: Experiences and Suggestions." *Studies in Comparative International Development* 25(1): 25–50.
- Giannone, Diego. 2010. "Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case." *Democratization* 17(1): 68–97.
- Girod, Desha M., Stephen D. Krasner, and Kathryn Stoner-Weiss. 2009. "Governance and Foreign Assistance: The Imperfect Translation of Ideas into Outcomes." In *Promoting Democracy and the Rule of Law: American and European Strategies*, ed. Amichai Magen, Thomas Risse and Michael A. McFaul. London: Palgrave Macmillan.
- Goldstein, Robert Justin. 1986. "The Limitations of Using Quantitative Data in Studying Human Rights Abuses." *Human Rights Quarterly* 8(4): 607–27.
- Goulet, Denis. 1974. "Freedom Does Not Exist in a Vacuum." *Worldview* 17(11): 40–2.
- Gourevitch, Peter A. and David A. Lake. 2012. "Beyond Virtue: Evaluating and Enhancing the Credibility of Non-Governmental Organizations." In *The Credibility of Transnational NGOs: When Virtue Is Not Enough*, ed. Peter A. Gourevitch, David A. Lake, and Janice Gross Stein. Cambridge: Cambridge University Press.
- Gray, Julia. 2013. *The Company States Keep: International Economic Organizations and Investor Perceptions*. Cambridge: Cambridge University Press.
- Green, Brendan Rittenhouse. 2012. "Two Concepts of Liberty: U.S. Cold War Grand Strategies and the Liberal Tradition." *International Security* 37(2): 9–43.
- Green, Jessica F. 2013. *Rethinking Private Authority: Agents and Entrepreneurs in Global Environmental Governance*. Princeton, NJ: Princeton University Press.
- Guilhot, Nicolas. 2005. *The Democracy Makers: Human Rights and the Politics of Global Order*. New York: Columbia University Press.
- Gunitsky, Seva. 2015. "Competing Measures of Democracy in the Former Soviet Republics." In *Ranking the World: The Politics of International Rankings and Ratings*, ed. Alexander Cooley and Jack Snyder. New York: Cambridge University Press.
- Häge, Frank M. 2011. "Choice or Circumstance? Adjusting Measures of Foreign Policy Similarity for Chance Agreement." *Political Analysis* 19(3): 287–305.
- Hall, Rodney Bruce and Thomas J. Biersteker. 2002. "The Emergence of Private Authority in the International System." In *The Emergence of Private Authority in Global Governance*, ed. Rodney Bruce Hall and Thomas J. Biersteker. Cambridge: Cambridge University Press.
- Hobson, Christopher and Milja Kurki. 2012. "Introduction: The Conceptual Politics of Democracy Promotion." In *The Conceptual Politics of Democracy Promotion*, ed. Christopher Hobson and Milja Kurki. London: Routledge.
- Hurd, Ian. 1999. "Legitimacy and Authority in International Politics." *International Organization* 53(2): 379–408.
- Hyde, Susan D. 2012. "Why Believe International Election Monitors?" In *The Credibility of Transnational NGOs: When Virtue Is Not Enough*, ed. Peter A. Gourevitch, David A. Lake, and Janice Gross Stein. Cambridge: Cambridge University Press.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. "The Worldwide Governance Indicators: Methodology and Analytical Issues." World Bank Policy Research Working Paper No. 5430. Washington, DC: World Bank.
- Keck, Margaret E. and Kathryn Sikkink. 1998. *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.
- Kekic, Laza. 2007. "The Economist Intelligence Unit's Index of Democracy." In *The World in 2007*. London: The Economist Intelligence Unit.
- Kelley, Judith G. 2009. "D-Minus Elections: The Politics and Norms of International Election Observation." *International Organization* 63(4): 765–87.
- Kelley, Judith G. and Beth A. Simmons. 2015. "Politics by Number: Indicators as Social Pressure in International Relations." *American Journal of Political Science* 59(1): 55–70.
- Klitgaard, Robert. 1994. "Do Better Politics Have Higher Economic Growth?" Working Paper No. 113. College Park, MD: Center for Institutional Reform and the Informal Sector.
- Kurki, Milja. 2010. "Democracy and Conceptual Contestability: Reconsidering Conceptions of

- Democracy in Democracy Promotion." *International Studies Review* 12(3): 362–86.
- . 2013. *Democratic Futures: Revisioning Democracy Promotion*. Abingdon: Routledge.
- . 2016. "Causality, Democracy Support, and the Cult of the Factish Gods." *Journal of International Relations and Development* forthcoming.
- Lake, David A. 2010. "Rightful Rules: Authority, Order, and the Foundations of Global Governance." *International Studies Quarterly* 54(3): 587–613.
- Mainwaring, Scott, Daniel Brinks, and Aníbal Pérez-Liñán. 2001. "Classifying Political Regimes in Latin America." *Studies in Comparative International Development* 36(1): 37–65.
- Mattli, Walter and Tim Büthe. 2005. "Accountability in Accounting? The Politics of Private Rule-Making in the Public Interest." *Governance* 18(3): 399–429.
- McMahon, Edward R. 2001. "Assessing USAID's Assistance for Democratic Development: Is It Quantity versus Quality?" *Evaluation* 7(4): 453–67.
- Mitchell, Lincoln. 2016. *The Democracy Promotion Paradox*. Washington, DC: Brookings Institution Press.
- Monten, Jonathan. 2005. "The Roots of the Bush Doctrine: Power, Nationalism, and Democracy Promotion in U.S. Strategy." *International Security* 29(4): 112–56.
- Munck, Geraldo L. and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy." *Comparative Political Studies* 35(1): 5–34.
- Nelson, Stephen C. 2014. "Playing Favorites: How Shared Beliefs Shape the IMF's Lending Decisions." *International Organization* 68(2): 297–328.
- Nooruddin, Irfan and Sarah Wilson Sokhey. 2012. "Credible Certification of Child Labor Free Production." In *The Credibility of Transnational NGOs: When Virtue Is Not Enough*, ed. Peter A. Gourevitch, David A. Lake, and Janice Gross Stein. Cambridge: Cambridge University Press.
- Oren, Ido. 1995. "The Subjectivity of the 'Democratic' Peace: Changing U.S. Perceptions of Imperial Germany." *International Security* 20(2): 147–84.
- Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4): 426–49.
- Quigley, Thomas E. 1974. "Miss Freedom' Awards Are, At Best, Irrelevant." *Worldview* 17(11): 39–40.
- Robinson, William I. 1996. *Promoting Polyarchy: Globalization, U.S. Intervention, and Hegemony*. Cambridge: Cambridge University Press.
- Scoble, Harry M. and Laurie S. Wiseberg. 1981. "Problems of Comparative Research on Human Rights." In *Global Human Rights: Public Policies, Comparative Measures, and NGO Strategies*, ed. Ved P. Nanda, James R. Scarritt, and George W. Shepherd, Jr. Boulder, CO: Westview Press.
- Sending, Ole Jacob and Iver B. Neumann. 2006. "Governance to Governmentality: Analyzing NGOs, States, and Power." *International Studies Quarterly* 50(3): 651–72.
- Shepherd, Hana. 2011. "Great Decisions: The Organizational Basis of Foreign Policy Ideas in the Council on Foreign Relations." Ph.D. diss., Princeton University.
- Snider, Erin A. and David M. Faris. 2011. "The Arab Spring: U.S. Democracy Promotion in Egypt." *Middle East Policy* 18(3): 49–62.
- Steiner, Nils D. 2016. "Comparing Freedom House Democracy Scores to Alternative Indices and Testing for Political Bias: Are U.S. Allies Rated as More Democratic by Freedom House?" *Journal of Comparative Policy Analysis* 18(4): 329–49.
- Stroup, Sarah S. and Wendy H. Wong. 2016. "The Agency and Authority of International NGOs." *Perspectives on Politics* 14(1): 138–44.
- Sussman, Leonard R. 2002. *Democracy's Advocate: The Story of Freedom House*. New York: Freedom House.
- . 2009. Freedom House. In *Encyclopedia of Human Rights*, Vol. 1, ed. David P. Forsythe. New York: Oxford University Press.
- Tannenwald, Nina. 1999. "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use." *International Organization* 53(3): 433–68.
- U.S. State Department. 2008. "Master List of Standard Indicators." Available at <http://2001-2009.state.gov/documents/organization/115255.pdf>, accessed April 3, 2017.
- . 2013. "Master Indicator List." Available at <http://www.state.gov/documents/organization/207793.pdf>, accessed April 3, 2017.
- Zürn, Michael, Martin Binder, and Matthias Ecker-Ehrhardt. 2012. "International Authority and Its Politicization." *International Theory* 4(1): 69–106.