

Dear friends and colleagues,

Thank you very much for reading the attached manuscript, “The Language of Compromise in International Agreements.” In this article, Tom Pegram and I investigate why National Human Rights Institutions, independent agencies designed to promote human rights, spread to over 100 countries, and explore the role of the United Nations in promoting a very idiosyncratic set of institutional design safeguards. This paper was recently accepted in *International Organization*, the leading international relations journal, but Tom and I plan to do much further work, to investigate how NHRIs function in practice. So I would benefit greatly both from comments on the paper itself, and from thoughts on what makes independent agencies effective, especially in adverse political environments.

Warm regards,

Katerina

The Language of Compromise in International Agreements

To reach agreement, international negotiators often compromise by introducing flexibility in language: they make controversial provisions vague, or add options and caveats. Does flexibility in agreement language influence subsequent state behavior? If so, do states follow both firm and flexible language somewhat, as negotiators hope? Or do governments respond strategically, increasing their energies on firmly specified tasks, and reducing their efforts on flexibly specified ones? Testing theories about agreement language is difficult, because states often reserve flexible language for controversial provisions. To make causal claims, we study an unusually drafted agreement, in which states had almost no opportunity to dilute agreement language. We examine the influence of the 1991 Paris Principles on the Design of National Human Rights Institutions (NHRIs), using an original dataset of 22 institutional safeguards of NHRIs in 107 countries, and case studies. We find that variations in agreement language can have large effects on state behavior, even when the entire agreement is non-binding. Both democracies and authoritarian states followed the Principles' firm terms closely. However, authoritarian states either ignored or reduced their efforts on flexibly specified tasks. If flexibly specifying a task is no different from omitting it altogether, as our data suggest, the costs of compromise are much greater than previously believed.

Katerina Linos, J.D./Ph.D.,
Professor
UC Berkeley Law School
klinos@law.berkeley.edu

Tom Pegram, Ph.D.
Lecturer in Global Governance
University College, London
t.pegram@ucl.ac.uk

Forthcoming, *International Organization*

The Language of Compromise in International Agreements¹

Drafting an international agreement is hard. Negotiations typically take many years and include representatives from hundreds of states. If an agreement is reached, it often involves major compromises. Provisions that experts consider important are watered down, or even omitted altogether. To reach a compromise, negotiators often replace firm language with flexible language: they make controversial terms vague or optional, and introduce caveats and conditions.²

How much do these compromises in agreement language affect the chances that the agreement will produce the results negotiators desire? Some theories imply that variation in agreement language will matter little, because international agreements generally have small effects on states' choices.³ These theories predict especially small effects in the human rights field, because mechanisms that could drive compliance in other areas – such as reciprocity, reputation, and retaliation – are unlikely to influence a state's conduct towards its own residents.⁴

Other scholars, and many international negotiators, expect all agreement provisions – even flexible ones – to have some effects on states' behavior.⁵ In these accounts, precise language without any options or caveats is most likely to influence states, because it can prevent self-interested interpretations. Imprecise language, with caveats or options, functions as a second-best solution that triggers smaller but still positive changes in states' behavior.⁶

We introduce a third theoretical possibility, and argue that variation in agreement language can have large and unintended consequences. When an agreement specifies some tasks firmly, some flexibly, and does not mention others, governments may respond strategically, redirecting efforts towards firmly specified tasks, and away from flexibly specified or omitted tasks. When tasks are specified using precise and unconditional formulations, there is less room for dispute about the adequacy of implementation efforts. If a government ignores a firmly specified task, foreign and domestic monitors can easily spot the discrepancy and withhold agreement benefits. Governments face the strongest incentives to focus on firmly specified tasks when they face tight constraints in an issue area, but want to signal that they are performing at the international standard. Flexible language can therefore induce some states to move in the opposite direction from that intended by an agreement's drafters.

However, these theories, and theories about agreement form more generally, are difficult to test empirically. To assess the impact of agreement language on states' choices, researchers must hold constant the substantive importance of agreement provisions. Yet, if language is a tool for reaching compromise, firm language will often be associated with provisions of minor substantive importance. If governments comply swiftly with provisions that employ firm language, it is hard to tell if this is because of the firm language, or because these provisions only called for minor substantive change.

We move away from this conundrum by focusing on an agreement that was negotiated in an unusual way, leaving governments almost no opportunity to dilute the language of controversial terms. As a result, provisions of comparable substantive importance *ex ante* were

¹ For replication data, please see Appendix A.

² E.g., Abbott et al. 2000; Abbott and Snidal 2000; Guzman 2008, 154–61.

³ E.g., Downs, Rocke and Barsoom 1996.

⁴ E.g., Guzman 2008.

⁵ E.g., Abbott et al. 2000, 412–5; Kahler 2000, 673.

⁶ E.g., Abbott and Snidal 2000.

sometimes specified in firm language and sometimes expressed in flexible language. Moreover, since the agreement we study is a non-binding UN General Assembly resolution, it would be surprising to find it had any effects, much less effects traceable to linguistic variations in individual provisions.

We examine the influence of the Paris Principles, an agreement that calls on all countries to establish a National Human Rights Institution (NHRI), and specifies how these institutions should be designed.⁷ NHRIs are independent regulatory agencies created to protect human rights.⁸ Most NHRIs can receive individual complaints by harmed parties, and can either help steer them through the state's administrative mechanisms, or, in some cases, review them in a quasi-judicial manner. Additionally, many NHRIs also play an active part in the lawmaking process by reviewing legislation and submitting recommendations.

Because NHRIs place important limits on government conduct, adopting an NHRI can impose significant costs on governments, especially on authoritarian governments that otherwise operate without much domestic monitoring. Case studies from diverse countries show that NHRIs have assisted torture victims in seeking redress before domestic and international courts, stewarded truth and reconciliation processes, improved legislation protecting vulnerable groups, mediated social conflicts, and mobilized public opinion on environmental rights.⁹ That said, not all NHRIs are effective. Case studies indicate that NHRIs without institutional safeguards of independence and grants of power are less likely to make a mark.¹⁰ NHRIs established in constitutions or national laws rather than through executive orders, or whose officials can be removed only for cause, are less vulnerable to political interference.¹¹ In turn, NHRIs with explicit powers to launch public inquiries, compel testimony or other evidence, and bring complaints to courts can pursue their mandate more effectively. As a leading commentator noted, when NHRIs without strong institutional safeguards were effective, “this was *despite*, not *because*, of their absence.”¹²

We compiled an original dataset of 22 institutional design safeguards in NHRIs in 107 countries. The Paris Principles prescribe some of these safeguards in firm language, and we characterize these as “strongly recommended.” For other safeguards, the Paris Principles use flexible language; we label these “weakly recommended.” Finally, our dataset also includes features that were omitted from the Paris Principles, although they existed in NHRIs that predate the agreement.

We seek to examine how small differences in the language of international agreements affect subsequent state behavior. To develop causal claims, we must reject the alternative hypothesis that states assigned strong recommendation status to provisions they expected to implement easily. We researched the origins of the Paris Principles, and show that the drafters of the Principles lacked the information, incentives, and capabilities to assign strong recommendation status to easy-to-implement features. Drafts of the Principles submitted by negotiators, the minutes of the formal meeting, and interviews with key participants in an informal working group show that a handful of participants – who did not represent states –

⁷ United Nations General Assembly 1993.

⁸ Creamer and Simmons 2013; Hafner-Burton 2013; Kim 2013; Cardenas 2014.

⁹ Okafor 2002; Domingo 2006; Finkel 2012.

¹⁰ Pegram 2012.

¹¹ Carver 2000, 20.

¹² Carver 2005, 7.

made decisions about what to include and exclude in three days and with very limited information. In addition, we show that strong recommendations, weak recommendations, and non-recommendations were similar in key respects, including how prevalent recommended safeguards were in NHRIs established before Paris, how controversial they were at the Paris meeting, and how they were rated by experts.

Our quantitative analyses suggest that whereas both authoritarian and democratic states followed the Paris Principles' strong recommendations, authoritarian regimes likely ignored and may have even reduced their efforts on tasks that were weakly recommended. We first examine only NHRIs set up after Paris and find that strongly recommended features were much more likely to be included in these new institutions than either weakly recommended or omitted features; we see no differences between weak recommendations and omissions. Next, we compare NHRIs set up before Paris to NHRIs set up after Paris. These additional findings are more tentative because some country differences might remain. Nevertheless, the patterns are suggestive: strongly recommended features became more common among both democracies and authoritarian states, whereas weakly recommended and omitted features became rarer among authoritarian states. Finally, we analyze each of the 22 institutional safeguards separately, and show that nearly every strongly recommended feature that could increase in prevalence (because it had not already reached widespread prevalence before Paris) did increase, whereas this almost never happened for other features.

Our qualitative inquiry explores how policymakers and human rights advocates referred to the Paris Principles as they designed national agencies. We focus on hard cases: cases in which there were competing templates for NHRI design. We show how interest groups pointed to firm language in the Principles to pressure politicians to set up NHRIs with all the strongly recommended features. However, we also show how politicians pointed to flexible language in the Principles to exclude institutional safeguards common to NHRIs in their region.

Our findings suggest that the language of international agreements can have large consequences for subsequent state behavior, even when the entire agreement is non-binding. When negotiators cannot agree on strong recommendations, they hope that weak recommendations are a good second-best option, and that a partial agreement will achieve some desired results.¹³ Our data imply that weak recommendations may have no net effect on average, and can even be used to undermine the agreement's goals in certain countries.

Our project contributes to human rights debates by shedding light on the origins and design of new monitoring bodies. We also add to the literature on compliance by studying state responses to individual agreement provisions rather than to entire agreements. Our findings suggest that small net effects on state behavior reported in prior work might not stem from governments' indifference to international agreements, but from large strategic responses that push in opposite directions. We also add clear evidence to theories on norm diffusion and constructivism by identifying the precise time and place at which an international norm changed and how it spread around the world.

I. How the Language of International Agreements Shapes State Behavior

Prominent theories distinguish between three dimensions of the form of international agreements.¹⁴ One dimension, legality, concerns states' formal obligation under international law to comply with a set of rules. Another dimension, delegation, describes whether an independent

¹³ Brewster 2010.

¹⁴ Abbott et al. 2000.

third party is entrusted with monitoring and enforcement. A third dimension, described in the legalization literature as precision, distinguishes between rules that unambiguously specify the content of the commitment, and rules that are vague.

We build on this third dimension of the legalization literature and label the concept we study “flexibility in agreement language.” Flexibility in agreement language includes three related linguistic techniques: vagueness, options, and caveats. To reach agreement, international negotiators can make controversial provisions vague, and allow them to admit multiple interpretations, as the legalization literature posits.¹⁵ Moreover, negotiators can make controversial provisions optional. Finally, negotiators can qualify states’ commitment to comply, and require compliance only if certain political, economic, and social conditions occur. All three techniques are commonly found in international agreements and are discussed in older international law literatures, but are not foregrounded in current international relations debates.¹⁶ The linguistic formulations we emphasize are not the only tools that enhance state flexibility; other scholars use the term flexibility to refer to derogations, reservations, withdrawal, and opt-out clauses. The key difference is that each of these devices requires individual states to invoke them explicitly in specific times and circumstances, whereas the linguistic techniques we study are broadly available to states at many points in time. In short, we label agreement language as firm if it is precise, without options and conditions; we label agreement language as flexible if it is imprecise, or optional, or conditional; and recognize that additional techniques can enhance an agreement’s flexibility.

How does variation in agreement language shape subsequent state behavior? The sections that follow develop two theoretical arguments drawn from the existing literatures in international law and international relations, and introduce a third, novel, theoretical possibility. We also explain why all theories about the form of international agreements present empiricists with big challenges.

A. Why Variation in Agreement Language May Produce No Effects

Diverse scholars predict that international law rarely influences states’ choices, and that non-binding human rights agreements are especially likely to be ignored. Some suggest that international law seldom matters because states themselves choose the commitments they will be held to. In the course of drafting international agreements, states only promise to complete tasks they were planning to complete anyway.¹⁷

Related literatures emphasize the weakness of global institutions. Absent a global police force, states comply with international agreements only when they expect concessions to be reciprocated and rewarded with future cooperation. States’ interests in reciprocity, their concerns about their reputation, and their fears of retaliation can explain compliance in fields such as trade or arms control.¹⁸ However, these mechanisms are less effective in fields such as human rights. Because human rights violations primarily harm a country’s own residents, foreign states have limited incentives to identify and punish violations abroad.¹⁹

Finally, non-binding agreements are considered even less powerful than binding ones. International, regional, and domestic courts give limited weight to non-binding sources. In

¹⁵ Id.

¹⁶ Gold 1983; Gamble 1985.

¹⁷ E.g., Downs, Rocke and Barsoom 1996.

¹⁸ E.g., Guzman 2008; Goldsmith and Posner 2006.

¹⁹ E.g., Guzman 2008.

addition, states can more easily discount non-binding agreements as aspirational, without suffering reputational harms.²⁰

For the above reasons, many scholars would be surprised to see a non-binding human rights agreement produce effects. If the entire agreement has limited influence, it would be even more surprising to find effects traceable to variation in the language of individual provisions.

B. Why Variation in Agreement Language May Produce Moderate Effects, and Lead to a Continuum of Influence

Other international relations theorists, and most international law scholars and practitioners, expect that variation in the form of an international agreement in general, and flexibility in agreement language in particular, shapes compliance. They expect that that firm language will induce the biggest changes in state behavior, flexible language will induce smaller changes, while omissions are unlikely to influence states.²¹ For example, Kenneth Abbott, Robert Keohane, Andrew Moravcsik, Anne-Marie Slaughter and Duncan Snidal suggest that there exists “a multidimensional continuum.”²² When international agreements score highly on each of the three dimensions of legality, monitoring, and precision, states are most likely to comply; weakening the agreement on any dimension lowers the likelihood of compliance.²³ Similarly, Abram and Antonia Chayes argue that that precise language is more likely to elicit compliance than ambiguous language.²⁴ Prior work is not crystal clear on why this is the case, so we develop this hypothesis.

Why might variation in agreement language shape state behavior? We hypothesize that when agreement language is firm – i.e., precise, without options and qualifications – governments, international audiences, and domestic audiences, can more easily agree on whether particular conduct meets the agreement standard. In contrast, vague language, as well as language introducing options and conditions, makes it harder for states and third-party monitors to concur on whether particular behavior conforms to the agreement. For example, while an obligation to destroy 1000 tons of a dangerous chemical is precise, the obligation to destroy “substantial quantities” of the dangerous chemical is vague. One could debate whether a state that destroys 500 tons of the chemical is in compliance or not with the “substantial quantities” requirement. Optional provisions can similarly elicit debate. Optional provisions open with verbs like “may” or “shall make efforts to,” instead of “shall” or “must.” Again, it is hard to assess whether a state that destroyed 500 tons of a chemical has complied with an agreement indicating that the state shall make efforts to destroy 1000 tons. Language that qualifies states’ obligations to comply by conditioning compliance on certain criteria can similarly lead to disagreements. For example, an agreement may indicate that the state shall destroy 1000 tons of the dangerous chemical as long as this is not inconsistent with its national security needs. Again, debate could follow about whether the destruction of 500 tons was appropriate, given particular security threats. International lawyers have written extensively about how vagueness, as well as conditions and caveats, have generated significant interpretation debates around prominent international agreements.²⁵

²⁰ Guzman & Meyer 2010.

²¹ E.g., Abbott et al. 2000, 414.

²² Abbott et al. 2000, 401–02.

²³ Kahler 2000, 673.

²⁴ 1993, 188–92.

²⁵ Gold 1983; Gamble 1985.

In short, firm language makes it easier to separate out the wheat from the chaff: firm language makes it harder for governments performing below the international standard to hide, and easier for governments performing at or above the international standard to shine. Other states, international organizations, and domestic constituencies can then more easily reward compliance and punish non-compliance. While reciprocity and retaliation are rarely used in the human rights arena, other rewards and punishments are employed. Some types of international lending decline when UN bodies condemn a country's human rights practices.²⁶ Similarly, preferential trade access is sometimes linked to compliance with human rights agreements.²⁷ In addition, domestic advocacy groups can be especially convincing when they highlight clear discrepancies between international agreements and government practices.²⁸

For these reasons, and consistently with prior writings, we expect that firm language is more likely than flexible language to elicit compliance with an agreement's goals. However, existing theories of international agreements assume that any changes in the behavior of states will be in the direction the agreement's drafters intended. They expect that at best, flexible language will generate some limited compliance pull, while at worst, flexible language will have no effect. Our next section explores a more troubling possibility.

C. Why Variation in Agreement Language May Produce Large Effects, Including Unintended Effects

We also develop a novel theoretical possibility: that flexible agreement language can sometimes lead states to behave in ways an agreement's drafters did not intend. An agreement that has both firm and flexible terms creates incentives for a government to focus its energies on tasks that are specified firmly. If a government ignores a firmly specified task, international and domestic monitors can easily identify the discrepancy, and withhold some benefits. In contrast, if a government ignores a flexibly specified task, it could argue that it is still substantially in compliance with the agreement, and should enjoy the promised benefits.

A government might redirect resources towards tasks that are specified firmly, and away from tasks that are specified flexibly (or entirely omitted), whenever the government finds it difficult to make progress across the board. Governments can redirect resources in many ways: for example, if an international agreement prioritizes human rights, states could cut back on environmental protection. While many trade-offs are possible, we examine whether substitution happens within the issue area of the agreement, among items that could plausibly be included in the agreement. This narrower focus helps us study the practical questions negotiators face: whether to conclude an imperfect agreement, or keep negotiating.

We expect any substitution effects to be concentrated among states that want to appear to conform to an agreement, but find it costly to do so. For example, in the case of an arms control agreement, we expect states with many hostile neighbors to have the strongest incentives to focus on firmly-worded agreement provisions, and perhaps reduce their disarmament efforts in other ways.

In the case of an agreement concerning human rights monitors, we expect any substitution effects to be concentrated among authoritarian states. This is because authoritarian states want to impress domestic and international audiences with their human rights progress, but fear monitoring bodies that are too powerful. Authoritarian states have established a variety of

²⁶ Lebovic and Voeten 2009.

²⁷ Hafner-Burton 2005; Böhmelt and Spilker 2013.

²⁸ Keck and Sikkink 1998; Simmons 2009; Linos 2013.

partially independent institutions – notably courts – to satisfy foreign critics and build support among domestic constituencies.²⁹ While setting up judiciaries with some independence, authoritarian states prioritize regime stability, and thus limit court jurisdiction on sensitive issues, reduce claimants’ access through tough standing rules, and in extreme cases, dissolve courts that displease leaders.³⁰

In designing National Human Rights Institutions, authoritarian leaders could also behave strategically, and offer only those institutional safeguards strongly recommended in the Paris Principles, while ignoring safeguards that were weakly recommended or omitted. This would be problematic because the Paris Principles, like many international agreements, call on states to make progress on multiple tasks, but do so imperfectly. Even the Principles’ drafters never believed they had identified the exact institutional safeguards all NHRIs should have, but simply listed some basic features, and encouraged countries to set up NHRIs in accordance with national conditions. The Paris Principles might thus induce unintended substitution effects that would be avoided if the agreement emphasized all important safeguards equally, and strongly recommended or weakly recommended or entirely omitted them all.

In short, this third theoretical possibility suggests that flexible agreement language could have different effects on different types of states. States that find it difficult to perform well in a policy area, because of tight domestic policy constraints, are especially likely to make trade-offs, and redirect their efforts on tasks that are inflexibly specified, and away from tasks that are flexibly specified or omitted.³¹ In contrast, states that find it easy to perform well in an issue area will have weaker incentives to substitute tasks in this way. Indeed, in such states, even flexible agreement provisions may lead to policy change in the direction intended by the agreement’s drafters, because the international agreement may put a new item on their national agendas.

D. Empirical Challenges

Theories about agreement form present empiricists with huge challenges. To assess the impact of agreement form on state behavior, a researcher should hold constant the substance of the agreement – i.e., whether the agreement calls for major or minor deviations from what states would otherwise do. However, states typically trade off form and function when they negotiate. When an agreement requires major substantive change, states often weaken its form, and make it non-binding, flexible, and lacking in monitoring.³²

Such trade-offs between form and substance complicate identification strategies. In a large cross-section of international treaties, we would expect a positive correlation between firm language and high compliance. However, we would not be able to determine whether firm language causes high levels of compliance, or whether firm language is reserved for modest substantive commitments. Scholars in the legalization literature recognize that “there is

²⁹ Ginsburg and Mustafa 2008.

³⁰ Solomon 2007, 125.

³¹ Hafner-Burton and Ron (2009), and Conrad, DeMeritt and Fariss (2015) also warn that agreement omissions can be costly, as agreements may induce authoritarian states to replace prohibited violations with other egregious actions, by substituting the practice of torture for enforced disappearances, for instance. In contrast, Fariss and Schackenberg (2014) report that complementarities may be more common than substitution effects. As an anonymous reviewer helpfully highlighted, existing datasets, aggregated at the country-year level, make it very difficult to uncover substitution effects.

³² Koremenos 2005, 563.

considerable difficulty in identifying the causal effects” of agreement form on state behavior.³³ While they compare human rights agreements to agreements governing other issues, such as trade and arms control, they acknowledge that it is difficult to hold all else constant, and that “these examples do not provide a true empirical test of our theory.”³⁴

A unique opportunity to identify whether states around the world respond to variation in agreement form arises from the Paris Principles. When outlining features for effective NHRIs, the Principles provide strong recommendations in some cases – expressed in precise, non-optional, and unqualified language – and weak recommendations in others. Other features experts consider important were entirely omitted from the Principles. Yet, due to an extraordinary set of events surrounding their negotiation, the Principles assign strong or weak recommendation status to various features in ways that are unlikely to influence the likelihood that states will later adopt these features. By focusing on the language used in different provisions of the same international agreement, we are able to hold constant other elements of an agreement’s form that also influence compliance.

States often choose not only the terms of an international agreement, but also whether or not to join the agreement. However, for many non-binding agreements, there is no step similar to accession through ratification; UN General Assembly Resolutions such as the one we study are addressed to all states. Thus, we never compare countries that ratify to countries that do not, and thus sidestep a selection issue that complicates other work.³⁵ Instead, we compare a country’s behavior to itself, using fixed-effects models, by studying twenty-two separate NHRI safeguards, only some of which were prioritized in the agreement. We also compare countries that adopted NHRIs before and after the international agreement; these specifications are more tentative.

We can only test our hypothesis that states prioritized firmly worded provisions, and treated flexibly worded provisions no differently from omissions, among countries that chose to adopt NHRIs. However, we expect that the patterns we report would be even sharper if more countries had adopted NHRIs. This is because our theory and available data suggest that tradeoffs between firm and flexible provisions are sharpest for countries that find it costly to adopt NHRIs. It is likely that countries that find it costly to adopt NHRIs are overrepresented among non-adopters, relative to adopters. We explore these patterns further in the sections that follow, and in Appendix G, which studies non-adopters.

II. The Paris Negotiations and the Paris Principles

In this section, we introduce the Paris Principles, and describe how this unusually negotiated agreement allows us to make significant progress in disentangling the causal effect of agreement language.

A. Negotiating the Paris Principles

In a typical international negotiation, states have the information, capacities, and incentives to assign flexible language to provisions they expect to be difficult to implement. In contrast, negotiators in Paris had little information, only three days, and few incentives to dilute the language of difficult-to-implement provisions. This original historical narrative helps us reject the alternative hypothesis that features assigned strong recommendation status were given this status because they were especially easy to implement.

³³ Abbott et al. 2000, 419.

³⁴ Abbott and Snidal 2000, 424. But Wallace 2013 provides experimental evidence that precise language can influence US public opinion.

³⁵ Böhmelt and Spilker 2013; Lupu 2013a; Lupu 2013b.

Prior to the 1991 meeting in Paris “there were virtually no limitations on the definition of a national human rights institution.”³⁶ Indeed, a leading expert recalls that the term “NHRI” was simply not in circulation then.³⁷ Institutions that would later be labeled NHRIs varied dramatically in form.

With the benefit of hindsight, we now see two loosely defined NHRI archetypes in 1991: the commission and the ombudsman. A human rights commission has a broad mandate, is composed of multiple representatives with human rights expertise, including civil society representatives, and typically researches entire situations and advises on legislation. In contrast, the ombudsman model involves a single appointee, empowered to investigate individual citizens’ grievances against the administration.

Before 1991, UN bodies had sought to harmonize and promote the NHRI concept, to close big gaps between international human rights treaties and human rights practices on the ground. However, Cold War politics made this difficult. In 1978, the UN Human Rights Commission convened a meeting on NHRIs that resulted in “draft guidelines.” But there was an impasse, as the US civil rights commission model was very different from the procurador system found in socialist countries. So the document languished, as neither the Human Rights Commission nor the General Assembly endorsed these “draft guidelines.”³⁸

The end of the Cold War elevated human rights in the international agenda and prompted the UN to revisit the question of NHRIs. In a typical human rights treaty negotiation, representatives from hundreds of states take almost a decade to hammer out compromises.³⁹ Fully aware that some international obligations may involve costly domestic reforms, diplomats submit carefully crafted proposals that preserve leeway for their governments on matters of national importance.

When the UN Human Rights Committee convened a workshop on NHRIs in 1991, its plans were modest. The objective was to “encourage existing National Institutions to step up their action” and to enhance cooperation among them.⁴⁰ Few expected a concrete outcome document to result from the workshop, and thus many governments chose not to participate.⁴¹ The meeting took place over three days in Paris. Fifty individuals from 35 countries attended, representing various national institutions, NGOs, UN agencies, regional human rights bodies, and a small number of governments.⁴² There was no precedent for global NHRI peer-to-peer meetings and most participants were encountering colleagues from other regions for the first time.⁴³

³⁶ Lindsnaes, Lindholt, and Yigen 2001, 8.

³⁷ We thank Richard Carver for this.

³⁸ United Nations General Assembly 1978.

³⁹ Knight and Versteeg 2014.

⁴⁰ United Nations Commission on Human Rights 1990, para. 4.

⁴¹ Tom Pegram’s interview with Maxwell Yalden, former Chief Commissioner of the Canadian Human Rights Commission (1987-1996), April 2012.

⁴² See United Nations Commission on Human Rights (3) 1992.

⁴³ Tom Pegram’s interview with David Mason, former assistant to Brian Burdekin, Australian Federal Human Rights Commissioner, August 2012.

“Everyone was surprised that the Paris Principles came out of the [Paris] meeting,” Chris Sidoti, a leading NHRI practitioner, recalls.⁴⁴ Multiple factors, including the meeting’s stated goals, the diversity of the participants, and the fact that these actors were meeting for the first time for only three days, posed formidable challenges to achieving any consensus on a defined set of NHRI characteristics.⁴⁵

Among the recognizable NHRIs in attendance, two commission models were especially well represented in Paris. First, since the meeting took place in Paris, the French delegation was the largest in size, set the agenda, and ensured crucial negotiations took place in French. Unsurprisingly, features of the French NHRI, an advisory commission, gained heightened prominence in the final text. Also, the Australian NHRI Commissioner had come prepared with a draft document, based on the recently established Australian commission. This happened to be a quasi-judicial human rights commission, and its features gained special consideration. In fact, while individuals from 35 countries attended the plenary session, interviews with key participants indicate that a Working Group of only four NHRI representatives (from Australia, France, Mexico and the Philippines) drafted a document behind the scenes. Minutes of the meeting and interviews suggest that the first two days of the meeting were devoted to statements about how existing NHRIs worked. Then, the draft Paris Principles prepared by the Working Group suddenly appeared on the final day, and were unanimously adopted in plenary without debate.

The Paris Principles’ genesis shaped their content. The absence of a formal diplomatic process with the requisite support and expertise, combined with the extreme time pressure, ensured these negotiators made a number of mistakes. Some are simple translation errors, but others are fundamental. Significant mistakes include the relegation of features typical of Latin-American ombudsmen to weak recommendations and to omissions. In Paris, the experts promoted the advisory and quasi-judicial commission models rather than the ombudsman model, because human rights ombudsmen from Latin America were severely under-represented at the Paris meeting, and other participants did not know enough about human rights ombudsmen to promote their features. As someone close to the negotiation process highlights: “it was a different time...we didn’t have Google, we literally did not know that the Spanish *Defensor del Pueblo* existed.”⁴⁶ Moreover, European “classical” ombudsmen, with a more limited human rights mandate than Latin American “human rights” ombudsmen, made statements at the meeting which led some participants to believe that ombudsmen in general were not well suited to address human rights concerns.⁴⁷

Few, if any, of the Paris meeting participants would have predicted that a year later the Human Rights Commission would endorse the principles, and two years later, in December 1993,

⁴⁴ Tom Pegram’s correspondence with Chris Sidoti, former Australian Federal Human Rights Commissioner, January 2011 (on file with author).

⁴⁵ Tom Pegram’s interview with Brian Burdekin, former Australian Federal Human Rights Commissioner (1986-1994), March 2012.

⁴⁶ Tom Pegram’s interview with David Mason, former research assistant to Brian Burdekin, Australian Federal Human Rights Commissioner, August 2012.

⁴⁷ See, e.g. statement of Mr. Jacob Söderman (Parliamentary Ombudsman, Finland). Report of the International Workshop on National Institutions for the Promotion and Protection of Human Rights 1991, 239–42.

the UN General Assembly would do the same, without modification.⁴⁸ Observers speculate that many delegations were simply not aware of what they were endorsing. If the resolution had been subject to debate and a vote it may have been blocked. As it was, the resolution – like many General Assembly resolutions – passed without a vote.

In sum, in a normal treaty drafting process, with hundreds of countries negotiating over many years, we would expect the final document to reflect key compromises among powerful states. But the significant time pressures and unusual composition of the Paris meeting generated an atypical document. Since the drafters of the Principles lacked the incentives, information, capabilities to assign easy-to-implement features to the strongly recommended category, we are less worried that selection effects drive our subsequent results.⁴⁹

B. The Content of the Paris Principles: NHRI Institutional Design Safeguards

Next, we analyze the Principles themselves. We show that although individual safeguards vary in substantive importance and in ease of implementation, on average, strongly recommended features are no less important or more difficult to implement than features that were weakly recommended or omitted. We thus provide further evidence disconfirming the alternative theory of selection effects.

Our dataset of NHRIs in 107 countries codes NHRI institutional structures at the time each NHRI was established. NHRI institutional design is quite stable over time; we examine some changes through qualitative case studies. The 22 features were selected based on the minutes of the Paris meeting, the text of the Paris Principles, an extensive review of NHRI scholarly and practitioner literature, and consultation with leading NHRI practitioners. Prior research examines NHRI adoption as a binary variable;⁵⁰ our dataset is thus significantly more fine-grained. Data was coded by law students who were provided with a codebook and detailed guidance to ensure NHRI charters were coded in conformity with the coding rules. The data was drawn from a variety of sources, principally NHRI compendia and institutional charters contained on official websites and state outlets. Where necessary, historical records were used.

Because the Paris Principles call on all states to establish NHRIs, and emphasize particular institutional safeguards, compliance with this agreement involves the establishment of an agency with these safeguards, rather than a reduction in arbitrary detention, an expansion of religious freedoms, or the improvement of living conditions for minority communities. That said, we focused on institutional safeguards considered consequential for NHRI effectiveness on the ground. Many features we study protect NHRI independence by limiting the power of the executive to disestablish the institution, fire its members, or pack it with pro-executive appointees. Other safeguards ensure that the agency has the formal power to take monitoring actions that governments could have otherwise blocked. Table 1 below lists the 22 safeguards we study, and briefly describes how each could contribute to NHRI effectiveness; Appendix B provides more details.

⁴⁸ United Nations Commission on Human Rights (2) 1992; United Nations General Assembly 1993.

⁴⁹ Dunning 2012, 236–39.

⁵⁰ E.g., Kim 2013; Cole and Ramirez 2013.

Table 1a: The Paris Principles – Strongly Recommended Features

Feature	Rationale	% of NHRIs set up before Paris with Feature	Mean Importance	Mean Difficulty
Broad Rights Mandate	Protects human rights broadly, including social, economic and cultural rights	92	4.7 (0.5)	2.8 (1.5)
Constitution or Legislation	Establishment by constitution or legislation makes NHRI charter harder to amend, and NHRI more stable	81	4.8 (0.5)	2.4 (1.3)
Power to Investigate	When NHRI can investigate on its own initiative, it can have proactive role, in contrast to reactive role of judiciary	94	5.0 (0.2)	2.9 (1.2)
Civil Society Representation	Civil society representatives facilitate contact with diverse societal groups	26	3.9 (1.1)	2.4 (1.3)
Harmonize IHR	Allows NHRI to help harmonize domestic law with international human rights standards	60	4.8 (0.6)	2.4 (1.2)
Education and Promotion	Promotes human rights among government agencies, schools, and civil society	57	4.3 (0.9)	2.3 (1.3)
Advise on Legislation	Helps make domestic legislation consistent with human rights standards	63	4.6 (0.8)	2.4 (1.1)
Engage with IOs	Helps connect NHRI to international organizations	31	4.2 (0.9)	2.7 (1.2)

Table 1b: The Paris Principles – Weakly Recommended Features

Feature	Rationale	% of NHRIs set up before Paris with Feature	Mean Importance	Mean Difficulty
No Government Representation	Government representatives may compromise NHRI autonomy	54	4.3 (0.9)	2.2 (1.3)
Not Designated by Executive	NHRI officials appointed by the executive may have limited independence	53	4.1 (1.1)	2.4 (1.4)
Long Mandate	A very short mandate can impede organizational stability	74	3.9 (1.0)	1.9 (1.0)
Possibility of Reappointment	Possibility of reappointment facilitates continuity of leadership	71	3.0 (1.2)	1.7 (1.2)
Individuals' Complaints	Power to hear individual complaints offers individuals direct access to NHRI	94	4.1 (1.0)	3.0 (1.4)
Enforcement Powers	Enforceable remedies expedite implementation of NHRI decisions	11	3.1 (1.4)	3.9 (1.2)
Can Refer Complaints	Facilitates access of vulnerable groups to courts	77	4.2 (0.8)	2.6 (1.1)
Can Compel Evidence / Testimony	Strengthens investigation and complaint-handling powers	60	4.6 (0.6)	3.2 (1.3)
Annual Report	Helps focus public opinion on country's human rights situation	81	4.4 (0.8)	1.9 (1.0)

Table 1c: The Paris Principles – Features not Mentioned

Feature	Rationale	% of NHRIs set up before Paris with Feature	Mean Importance	Mean Difficulty
Immunity	Immunity from prosecution helps safeguard NHRI leaders' independence	54	4.3 (0.9)	2.2 (1.2)
No Dismissal without Cause	Dismissal only for cause helps safeguard NHRI leaders' independence	49	4.7 (0.6)	1.9 (1.1)
Amicus Curiae Powers	The power to provide the courts with amicus curiae briefs is a supplementary tool	0	4.3 (0.8)	2.3 (0.8)
Security Facilities	Power to oversee prisons allows NHRI to monitor a site of potentially grave human rights violations	77	4.4 (1.0)	2.8 (1.3)
Single Head	Allows NHRIs to have a recognizable public representative	58	2.7 (1.3)	1.9 (1.1)

Notes: The last two columns in Tables 1a-c summarize experts' views of the importance (5=very important) and the difficulty to implement each safeguard (5=very difficult). Standard deviations are in parentheses.

The goal of our study is not to assess whether NHRIs with particular safeguards are especially able to improve human rights practices on the ground; we are devoting a separate paper to this critical question. However, our theory assumes that NHRIs have some impact – more specifically, our theory assumes that adopting an NHRI with many institutional safeguards is somewhat costly for governments, especially governments of authoritarian states. Qualitative work in Appendix B substantiates this assumption. In addition, to better assess the performance of NHRIs on the ground, we conducted an expert survey among NHRI heads and others with extensive knowledge of NHRIs, further detailed in Appendix C. Thirty-six of the sixty experts we contacted (60%), drawn from all five continents, completed the survey. Respondents rated the effectiveness of various NHRIs in different countries. We found that NHRIs that incorporated more of the 22 safeguards in their design were rated more effective. This positive correlation holds when we limit our sample to authoritarian states and when we break down our data by region. This suggests that institutional design is positively linked to effectiveness on the ground, but we plan to investigate this question in more detail in further work.

We characterize a feature as strongly recommended if the relevant language in the Paris Principles is precise, non-optional, and unqualified. Accordingly, we code a feature as weakly recommended when it comes with language that is either vague, or optional, or imprecise (or a combination). Features that were on the agenda in Paris meeting, but not included in the final document due to time pressures, are coded as omitted. Classifying individual provisions is relatively straightforward, because the Principles mainly use the verb “shall” to preface strong recommendations and the verb “may” to preface weak recommendations. Nevertheless, there is some subjectivity in the classification so we asked another international law scholar to recode the provisions. 21 of the 22 provisions were placed in the same categories, giving us very high inter-coder reliability (Krippendorff’s $\alpha = 0.93$).

While each of the safeguards in Table 1 can help an NHRI perform its mission effectively, not all are equally important or easy to implement. If relatively unimportant and easy to implement features were especially likely to be strongly recommended, this would give credence to the alternative theory of selection effects. Below, we assess this alternative hypothesis.

Features that were prevalent in NHRIs established before Paris may have been especially important to effective NHRI function, or especially easy for governments to adopt. Yet both very common and very rare features ended up strongly recommended, weakly recommended and omitted from the Paris Principles. On average, NHRIs established before Paris had 63% of features that ended up strongly recommended, 64% of features that ended up weakly recommended, and 48% of features that were omitted.⁵¹

Institutional features considered important or hard to implement might have attracted increased attention or controversy during negotiations in Paris. To identify these features, we examined the minutes of the plenary; coding details are in Appendix D. Overall, strongly recommended features were no less controversial than weakly recommended features. Strongly recommended features had an average controversy score of 0.63, weakly recommended features

⁵¹ While none of the differences in means are statistically significant, equivalence tests give us confidence that strong and weak recommendations likely came from the same distribution, but we have too little power to tell whether this is also the case for features not mentioned.

had a score of 0.67; the difference in means is small and not statistically significant. Features in the “not mentioned” category were not discussed due to time constraints.⁵²

Pre-1990 efforts to create a global NHRI template can also help determine whether some institutional features were understood as critical, and thus received “strong recommendation” status in the Principles. In particular, the discarded 1978 UN template provides a set of features that had attracted states’ attention before the Paris meeting. Yet, we find no association between the two documents: Included in the 1978 UN template were 75% of the strongly recommended features in Paris, 56% of the weakly recommended features, and 80% of the omitted features.⁵³ This suggests that the Paris Principles did not reflect a clear pre-1990 consensus on what safeguards NHRIs should include.

We also took advantage of the expert survey discussed above to understand how institutional safeguards vary.⁵⁴ According to the average expert ratings, features strongly recommended are no less important than weak recommendations or features not mentioned, and no easier to implement. More specifically, experts consider strongly recommended features to be slightly more important than weakly recommended ones. On a 5 point scale, where 5 is very important, strongly recommended features receive a score of 4.5, weakly recommended features a score of 4.0, and features not mentioned a score of 4.1. The difference between the mean importance of strongly and weakly recommended features is significant at the 0.05 level. In addition, on a 5 point scale where 5 indicates that a feature is very difficult to implement, our experts gave strongly recommended features a score of 2.54, weakly recommended features a score of 2.53, and features not mentioned a score of 2.22. The difference between strongly recommended features and features not mentioned is significant at the 0.10 level. Nevertheless, these differences are not consistent with an alternative theory based on selection effects; that is, the alternative theory would predict that strongly recommended features would be *less* important and *easier* to implement than other features. Moreover, because our survey was fielded after the Paris Principles were adopted, and could thus be influenced by the Principles, we give less weight to this measure. With these caveats, we include our expert survey data in Table 1 below, to offer a clear picture of how experts view each safeguard.

In short, we could not find evidence to support the alternative theory of selection effects. We find that a global template available before Paris, and features common to institutions established before Paris do not correlate with the assignment choices made in Paris. We also see that strongly recommended features were no less controversial at the Paris meeting than other features. In addition, our expert survey did not suggest that strongly recommended features were especially easy to implement or unimportant. Each of these tests has important limitations, but put together, they increase our confidence that strongly recommended features were not systematically different from weak recommendations in ways that could speed up their later adoption by diverse governments.

Instead, accidents of the negotiating process seemed to determine the category to which particular provisions were assigned. To the extent that we can identify any pattern, we note that features rarely found in commissions, but common in ombudsmen, were more likely to be

⁵² Tom Pegram’s interview with Brian Burdekin, former Australian Federal Human Rights Commissioner (1986-1994), March 2013.

⁵³ None of these differences between these means are statistically significant, but the absolute differences are large, and our confidence intervals are wide.

⁵⁴ See Appendix C for survey details.

weakly recommended or omitted. But this did not reflect a consensus view that commissions or ombudsmen were better at protecting human rights, or more costly or challenging for governments to introduce. Instead, this assignment simply reflects the fact that participants at the Paris meeting had limited knowledge of the ombudsman model.

III. The Impact of the Paris Principles on State Behavior: Quantitative Analysis

This section assesses the influence of the Paris Principles on the design of NHRIs using an original dataset of 22 features of 107 NHRIs. First, we study whether countries that adopted NHRIs after the Paris meeting followed the Principles' recommendations in establishing new institutions. Next, we compare the design of NHRIs established before and after Paris. In Appendix E, we examine each of the 22 features separately, and conduct robustness checks that exclude particular features, to check whether a few unusual features drive our aggregate results. Each of these analyses indicates that strong recommendations influenced countries significantly more than weak recommendations or omitted features, that weak recommendations were no different from omissions, and that distinctions between strong and weak recommendations were especially sharp for authoritarian states. In addition, the last two analyses suggest that, compared to authoritarian states adopting NHRIs before Paris, authoritarian states adopting NHRIs after Paris included many more strongly recommended features, and many fewer weakly recommended and omitted features. However, this last inference is more tentative, because it is based on comparisons of NHRIs in different countries, and we may not be able to fully control for every relevant country characteristic.

First, we study only countries that adopted NHRIs after Paris. The dependent variable is whether a country adopted a safeguard (1) or not (0). Thus, we present pooled logit models; OLS models yield very similar results. We report robust standard errors, clustered by country. Models in Table 2 estimate whether safeguards that were strongly or weakly recommended in the Principles were more likely to be adopted compared to safeguards not mentioned in the Principles (the omitted category). Estimates of the relative effects of strong recommendations are very similar when we use weak recommendations as a baseline (and exclude omitted features). Model 1 includes country fixed effects, as well as three feature-level variables – whether a feature was included in a 1978 template, how controversial it was at Paris, and how prevalent it was in NHRIs set up before Paris. Model 2 also includes two feature-level variables that were measured after the Paris Principles - how important and difficult to implement experts consider particular features in 2013. Model 3 removes the country-level fixed effects, to explore what types of countries were likely to adopt NHRIs with more of the features we study.

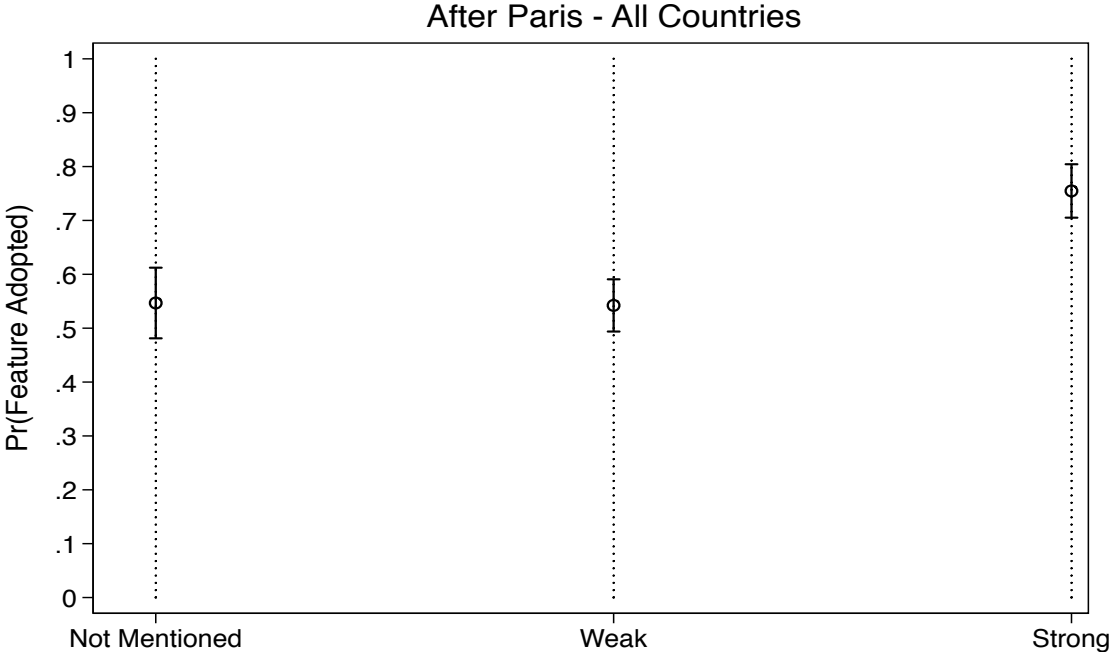
Across specifications, we find that strong recommendations had statistically and substantively significant effects, whereas weak recommendations made no difference. For example, Model 3 shows that whereas the probability that a country adopts a feature omitted from the Paris Principles is 0.55, the probability increases to 0.74 if the feature is strongly recommended, holding all other variables at their mean. Figure 1 graphs these predicted probabilities. Similarly each model indicates that features that were common in institutions established before Paris were common in institutions set up after Paris. NHRIs set up after Paris, Model 2 tells us, are also slightly more likely to include features experts consider important, but this result is tentative, because we conducted our expert survey after the Paris Principles were adopted. In Model 3, we include country and region level controls, and discuss these in Appendix D.

Table 2: NHRI Feature Adoption – Only Countries Adopting NHRIs After Paris

	Model 1 All Countries	Model 2 All Countries	Model 3 All Countries	Model 4 Democracies	Model 5 Authoritarian States
Feature					
Characteristics					
Strongly Recommended	1.14*** (0.18)	1.08*** (0.18)	1.22*** (0.27)	0.80** (0.35)	2.09*** (0.46)
Weakly Recommended	-0.03 (0.17)	0.03 (0.18)	-0.02 (0.16)	0.09 (0.22)	-0.16 (0.24)
1978 Template	-0.09 (0.15)	-0.07 (0.17)	-0.13 (0.17)	-0.20 (0.23)	0.08 (0.29)
Controversy	0.11 (0.10)	0.11 (0.11)	0.12 (0.09)	0.12 (0.11)	0.14 (0.19)
Prevalence before Paris	3.87*** (0.30)	3.66*** (0.32)	3.82*** (0.38)	4.32*** (0.55)	3.13*** (0.52)
Importance		0.24** (0.12)	0.18 (0.11)	0.50*** (0.13)	-0.42** (0.18)
Difficulty		-0.09 (0.16)	-0.07 (0.19)	-0.18 (0.26)	0.09 (0.34)
Country					
Characteristics					
Democracy (Polity IV)			0.04** (0.02)	0.08* (0.05)	0.00 (0.10)
Human Rights (Physical Integrity)			-0.16*** (0.04)	-0.24*** (0.06)	0.15* (0.09)
Ethnic Fractionalization			0.14 (0.46)	0.41 (0.71)	-1.00 (0.90)
GDP per capita			0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
Legal Origins (Common Law)			0.33 (0.25)	0.00 (0.30)	1.36*** (0.40)
Constant			-2.97*** (0.74)	-4.14*** (1.11)	0.63 (1.30)
Regional Dummies	No	No	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	No	No	No
Observations	1,539	1,539	1,351	896	455
Countries	71	71	62	41	21

Notes: Cell entries are logit coefficients, followed by robust standard errors clustered by country in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Features Adopted After Paris – All Countries



Note: Figure shows that NHRIs established after Paris included significantly more strongly recommended features (as compared to weakly recommended and omitted features).

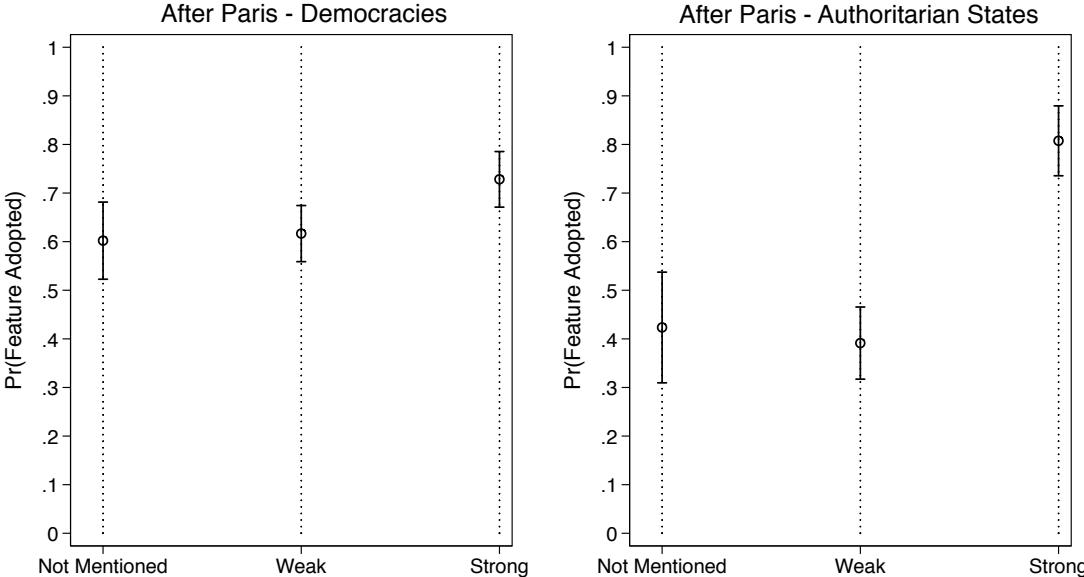
While both democracies and authoritarian states responded to the Principles' strong recommendations, and ignored the Principles' weak recommendations, authoritarian governments responded much more than democracies. Specifically, Model 4, which limits our sample to democracies, shows that among democracies, the probability of adoption of an omitted feature is 0.60, and increases to 0.73 for a strongly recommended feature. Model 5 shows that among authoritarian states, the probability of adoption of omitted features is 0.42, and it increases to 0.81 for strongly recommended features. Figure 2 presents these predicted probabilities. Authoritarian states may desire weak monitoring bodies, but may find themselves exposed to strong criticism of their practices. To balance these two competing pressures they could design institutions that have almost every feature that is strongly recommended by the Paris Principles, but few other safeguards.

We also study the influence of the Paris Principles by comparing countries that adopted NHRIs before the Paris Principles to countries that adopted them afterwards. This comparison requires us to carefully consider how differences in the characteristics of countries in the before and after group could influence our results. If countries that are richer, more democratic, or more protective of human rights are over-represented among the early adopters of NHRIs, and wealth, democracy, and human rights levels lead countries to adopt strong NHRIs with more safeguards, we should see a decrease in the prevalence of *all 22* NHRI features. While this would influence our results, it would lead to a change across the board, rather than to an effect concentrated among strongly recommended features. Thus, most country-level differences would not lead us to conclude that agreement language matters, and would not lead us to find that strong recommendations have bigger effects than weak recommendations, or that weak recommendations have bigger effects than no recommendations.

It turns out that countries that established NHRIs before Paris are very similar in their levels of wealth, democracy, and human rights protection to countries that adopted NHRIs after Paris, as Appendix D shows. This makes the before/after comparison more plausible. In Table 3, Model 1 only includes the main theoretical variables of interest to us: the interactions between recommendation strength and adoption of an NHRI after the Paris meetings, and the component terms of these interactions. Model 2 also includes controls for democracy, human rights (physical integrity index), ethnic fractionalization, GDP per capita, a common law legal system, plus regional dummy variables. Models 3 and 4 repeat Model 2 results separately for democracies and authoritarian regimes. Our results are consistent across these and additional specifications.

Figure 3 presents the change in prevalence of different safeguards, based on our main specification, Model 2. Features that were strongly recommended in Paris increased in prevalence; on average, and with all control variables held at their means, the probability that an NHRI established before Paris would have a strongly recommended feature was 0.65, and the corresponding probability for NHRIs established after Paris was 0.78, a highly significant increase of 0.13 ($p < 0.01$). In contrast, the probability that an NHRI established before Paris would have a weakly recommended feature was also 0.65, but fell to 0.58 for NHRIs established after Paris, a decrease of 0.07 that is marginally statistically significant ($p < 0.10$). Features that were not mentioned were just as prevalent before and after Paris.

Figure 2: Features Adopted by Democracies and Authoritarian States After Paris



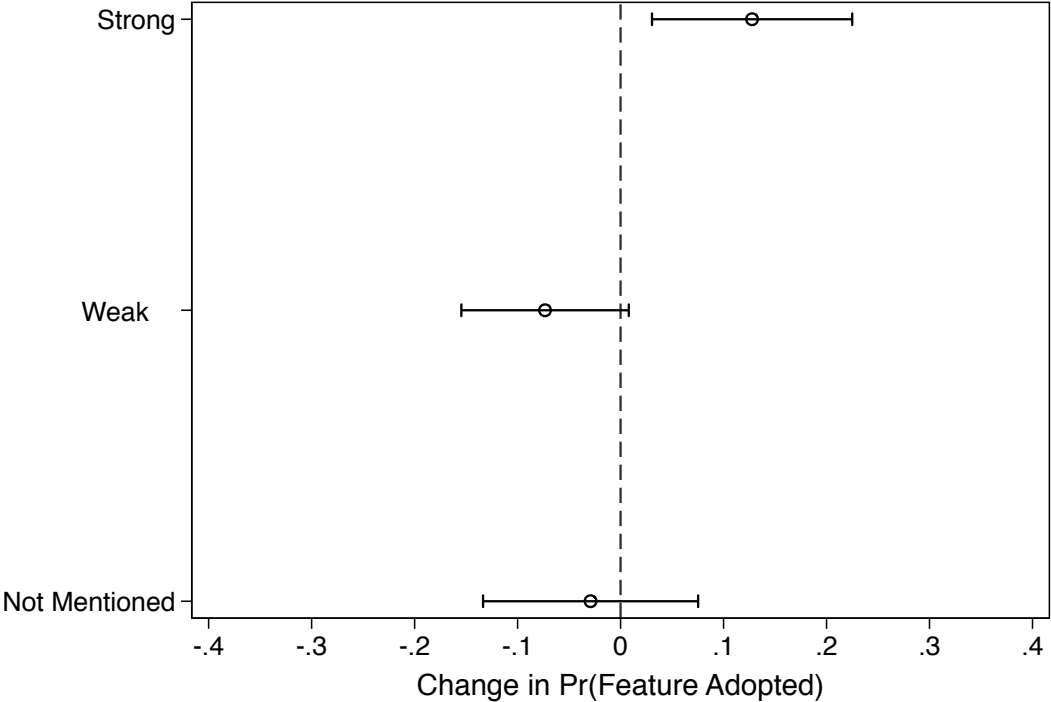
Note: Figure shows that democracies and especially autocracies included many more strongly recommended features (compared to weakly recommended and omitted features) in NHRIs they set up after Paris.

Table 3: Adoption of NHRI Features - Countries Adopting NHRIs Before and After Paris

	Model 1	Model 2	Model 3	Model 4
	All Countries	All Countries	Democracies	Authoritarian States
After Paris	-0.08 (0.21)	-0.12 (0.23)	0.05 (0.29)	-0.58 (0.37)
Weak Recommendation	0.67*** (0.13)	0.76*** (0.16)	0.79*** (0.18)	0.69** (0.33)
Strong Recommendation	0.62** (0.25)	0.76** (0.30)	0.77** (0.33)	0.74 (0.64)
After Paris * Weak Rec.	-0.15 (0.16)	-0.20 (0.19)	-0.12 (0.23)	-0.31 (0.37)
After Paris * Strong Rec.	0.73** (0.32)	0.78** (0.36)	0.48 (0.41)	1.47** (0.73)
Democracy		0.01 (0.01)	0.08** (0.03)	-0.04 (0.05)
Human Rights (Physical Integrity)		-0.08** (0.03)	-0.13*** (0.04)	0.02 (0.10)
Ethnic Fractionalization		-0.01 (0.31)	0.08 (0.45)	0.52 (0.72)
GDP per capita		-0.00 (0.00)	-0.00 (0.00)	-0.00* (0.00)
Legal Origins (Common Law)		0.04 (0.22)	-0.13 (0.23)	0.54 (0.34)
Constant	-0.09 (0.16)	-0.02 (0.37)	-0.33 (0.55)	-0.12 (0.84)
Regional Dummies	No	Yes	Yes	Yes
Observations	2,313	1,949	1,314	635
Countries	107	90	60	30

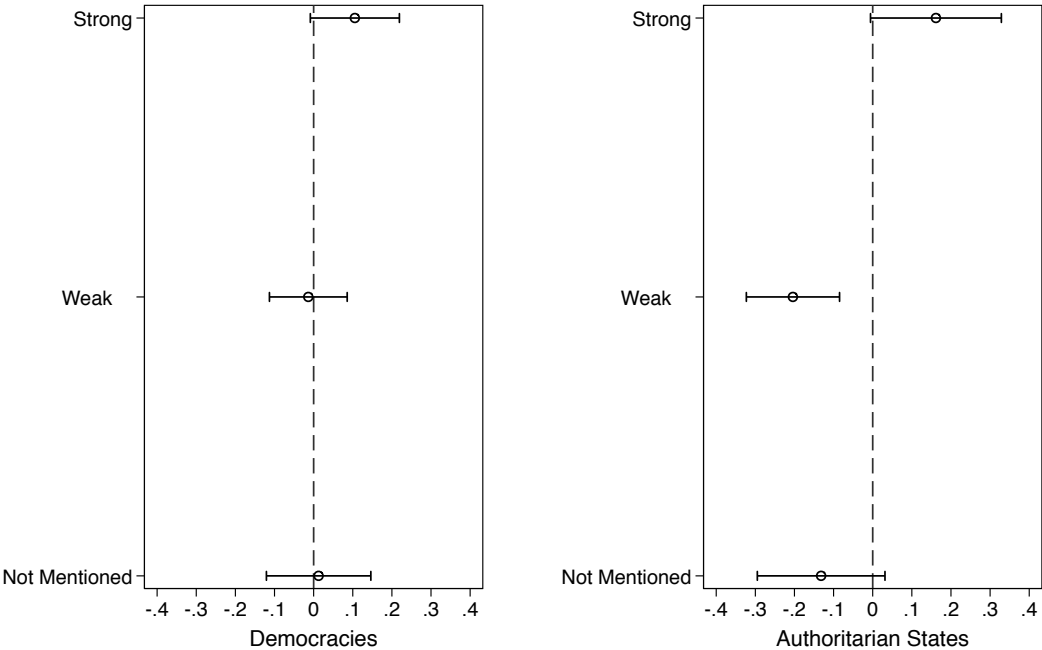
Notes: Cell entries are logit coefficients, followed by robust standard errors clustered by country in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3: Differences in Features Adopted (Comparing Before and After – All Countries)



Note: Figure shows the difference in the probability that an NHRI established after Paris would include a feature, as compared to an NHRI established before Paris. Across all countries, we only see a significant increase for strongly recommended features.

Figure 4: Differences in Features Adopted (Comparing Before and After – Democracies and Authoritarian States)



Note: Figure shows the difference in the probability that an NHRI established after Paris would include a feature, compared to an NHRI established before Paris. With democracies and authoritarian states analyzed separately, we see a marginally significant increase for strongly recommended features. For authoritarian states, we see a significant decrease for weakly recommended features, and a marginally significant decrease for features not mentioned.

Figure 4 illustrates a sharp difference between the responses of democracies and authoritarian states. The prevalence of strong recommendations increases somewhat among democracies. Among democracies, the probability that an NHRI established before Paris would have a strongly recommended feature was 0.65, as compared to 0.76 for NHRIs established after Paris, an increase of 0.11 ($p < 0.10$). In contrast, the prevalence of weak recommendations does not change in democracies: For NHRIs established before Paris, the probability of inclusion of a weakly recommended feature was 0.65, compared to 0.64 for NHRIs established after Paris. Similarly, the prevalence of features not mentioned in the Principles also does not change among democracies.

The increase in strongly recommended features is even more pronounced in authoritarian states. Among authoritarian states, the probability of inclusion of a strongly recommended feature increased from 0.65 in the before group, to 0.81 in the after group, an increase of 0.16 ($p < 0.10$). However, the prevalence of weakly recommended features falls sharply among authoritarian states. Among authoritarian states, for NHRIs established before Paris, the probability of including a weak recommended feature was 0.64, whereas it fell to 0.44 for NHRIs established after Paris, a decrease of 0.20 ($p < 0.01$). Similarly, the prevalence of features not mentioned in the Paris Principles declines among authoritarian states, although the difference is not statistically significant ($p < 0.15$). We discuss control variables in Appendix D.

To examine whether the effects we report are driven by a few unusual and unimportant features, we break down our data feature-by-feature in Appendix E, and also re-run the above models while excluding outliers. We find that the effects we report are generally homogenous within feature types. Indeed, almost every strongly recommended feature increases in prevalence, both in authoritarian states and in democracies, whereas this happens for very few weakly recommended features or features not mentioned; this pattern strengthens the inference that strongly recommending a feature matters.

IV. Case Studies

This section provides qualitative evidence to illustrate the mechanisms through which the Paris Principles influenced the design of particular NHRIs. We selected cases unlikely to confirm our theory. “Least-likely” cases offer strong “support for the inference that the theory is even more likely to be valid in most other cases, where contrary winds do not blow as strongly.”⁵⁵ The Paris Principles are least likely to influence states that have a competing template available. Such alternatives to the Paris Principles often come from dominant regional paradigms and from former colonial relationships. We present evidence from Chile below, because the hybrid ombudsman model common to much of Latin America was a plausible alternative. In Appendix F, we present evidence from Ireland and Morocco, as models common to the (British) Commonwealth and to the Francophone world, respectively, were plausible alternatives templates. Our case studies highlight that that within each country, the language of the Principles mattered to progressives and conservatives alike, and strong recommendations often got adopted while weak recommendations often did not.

Almost all Latin American countries now have NHRIs. Many NHRIs were established before the Paris Principles, and follow the Spanish hybrid ombudsman, or *Defensoria del Pueblo*, model. Ombudsman offices are typically led by a single head and their main task is to investigate individual complaints. In addition, in much of Latin America, Ombudsman offices are

⁵⁵ Odell 2001, 165.

constitutionally entrenched, and have no government representation. Absent the Paris Principles, Chile would likely have followed the Ombudsman model common in much of Latin America.

Instead, Chile followed the Commission model promoted in Paris, and incorporated a number of institutional safeguards strongly recommended Paris Principles, but rare in Latin American NHRIs set up before Paris. These include civil society representation, an explicit mandate to engage with international organizations, and authorization to harmonize domestic law with international human rights standards. Indeed, compliance with the Paris Principles was a prominent concern for NHRI adopters in Chile, with the President of the Republic emphasizing that the new law “takes into account comparative models and strictly adheres to the Paris Principles in light of our recent history.”⁵⁶

At the same time, Chile did not have a long and continuous history of democracy, as in the case of many democracies in our dataset. Likely because of this recent history of authoritarianism, Chile’s response to the Paris Principles was similar to that of some authoritarian states that declined to adopt safeguards that were only weakly recommended in Paris. Chile only emerged in 1990 from a protracted period of authoritarian government and the Pinochet regime remains highly controversial.⁵⁷ Chile’s NHRI was only created in 2009, as conservative political actors stalled earlier efforts, worried that an NHRI might revisit the Pinochet era and advocate for repeal of the amnesty laws decreed by Pinochet prior to his departure from office.

The resulting NHRI structure bears the hallmarks of built-in “authoritarian enclaves”, designed to facilitate peaceful transition to democracy from authoritarian rule.⁵⁸ The Chilean NHRIs lack several features typical of Latin American ombudsmen, and did not fully satisfy human rights advocates in Chile. For example, the Human Rights Centre of the University Diego Portales pointedly asked in its 2008 Annual Human Rights Report, “is the National Institute for Human Rights an authentic national human rights institution?”⁵⁹ A question of particular importance in NHRI design is whether they can receive individual complaints. Individual complaint-handling encourages NHRIs to directly engage with victims, and provides marginalized groups with a rare point of access to official redress. Also, the accumulation of individual complaints may help in exposing systematic abuses. The Chilean Institute lacks this faculty, as well as other investigative prerogatives such as inspection powers.⁶⁰

Government leaders used the Paris Principles’ weak recommendations to justify this omission. For example, the prominent socialist politician Antonio Viera-Gallo, serving at the time as the Minister Secretary General of the Presidency, argued that because the Paris Principles only weakly recommend individual complaint-handling, this tool was not necessary:

According to the Paris Principles, it is not indispensable that institutions of this kind have the faculty to receive complaints—there exist various [NHRIs] that do not have this faculty—especially given that in our penal system the prosecutor’s office holds a quasi-monopoly over

⁵⁶ Biblioteca del Congreso Nacional de Chile 2009, 7. See Biblioteca del Congreso Nacional de Chile 2009 for extensive discussion of Paris Principles compliance by diverse government officials and legislators.

⁵⁷ Sikkink and Booth Walling 2007, 434.

⁵⁸ Siavelis 2000.

⁵⁹ Vásquez 2008, 565.

⁶⁰ National Congress of Chile 2009, Art. 3(5).

such matters. What no one denies is that the Institute is entitled to denounce violations just like any other individual or entity, although, obviously, its denunciations would carry enormous political and juridical weight...⁶¹

Indeed, civil society advocates and even UN representatives took the unusual position of arguing against the strict application of the international template, because the regional template provided for more features than the international Principles. Robert Garretón, the UN Representative of the High Commissioner for Human Rights in Latin America argued that Chile would be the only Latin American country whose NHRI would not be able to receive individual complaints:

...a National Institute for Human Rights with no mandate to receive and investigate complaints is not exactly what is needed for the citizen whose rights have been violated. In concrete terms, the project does not grant the Institute the ability to receive, process, investigate and adopt recommendations in respect to individual complaints. It would be the only NHRI in Latin America to lack these essential attributes.⁶²

Similarly, civil society advocates in Chile, such as Cecilia Medina, the Director for the Center of Human Rights at the University of Chile, advocated that a margin of appreciation should be applied in Chile's interpretation instead of holding firmly to the integrity of the Paris Principles as design obligations:

The Paris Principles related to the status and functioning of national institutions for the promotion and protection of human rights [...] should not be understood necessarily as meaning that all of these attributes are to be found within one institution. The situation of each country will determine how the Principles are implemented. . . .⁶³

Chile also disregarded other weak recommendations in the Paris Principles. For example, the Paris Principles only weakly recommend “no government representation,” and the Chilean NHRI is highly unusual in Latin America due to the fact that the President appoints two of its seven Council Members.⁶⁴

In the end, the Chilean NHRI generally follows the Paris Principles closely, on items that are strongly recommended, and top Chilean leaders take pride in noting that their institute is Paris Principles compliant. However, the Chilean case also illustrates that weak recommendations in the Paris Principles can empower opponents of strong NHRIs to argue that other safeguards are not essential. We see very similar patterns in Uruguay whose NHRI, also created in 2009, conforms closely to the letter of the Paris Principles and diverges from the more traditional Latin American ombudsman model (see Appendix F).

V. Scope Conditions and Generalizability

Since our empirical strategy focuses on a single international agreement with an unusual negotiation process, it is important to explore how our results might generalize to agreements

⁶¹ Biblioteca del Congreso Nacional de Chile 2009, 553.

⁶² Biblioteca del Congreso Nacional de Chile 2009, 64.

⁶³ Biblioteca del Congreso Nacional de Chile 2009, 57.

⁶⁴ National Congress of Chile 2009, Art. 6.

with other features. Whereas the Paris Principles are non-binding, many international agreements are legally binding, and states may see compliance with any resulting obligations as especially important, or especially likely to be given heightened consideration by international and domestic courts. Moreover, we expect the distinction between may and shall, and firm and flexible language more generally, to matter more in a legally binding agreement, because one can debate whether flexibly phrased provisions are legally binding. Also, the Paris Principles concern human rights; in other issue areas, the mechanism of reciprocity provides additional incentives for compliance. For example, a government can credibly propose that it will comply with a truce only as long as the enemy also complies, but cannot easily signal that it will protect its own residents' human rights only as long as another state reciprocates. These arguments each suggest that the Paris Principles were a hard test case for theories of compliance, and that we should find effects that are at least as large in other fields.

Another feature of the Paris Principles – their unusual negotiation process – could also influence generalizability. In many settings, such as negotiations between individuals, we might expect a close link between negotiation processes and compliance, as bitter negotiations could sour subsequent implementation efforts. But in international negotiations, different branches of the state are often tasked with negotiation and implementation; for example, foreign ministry diplomats might negotiate an agreement, and environmental ministry staffers might be called on to implement it years later, as part of a different administration. Moreover, many small states often lack the capacity to participate meaningfully in negotiations; agreement terms can be treated as exogenous for such states. If negotiation and implementation are separable, an agreement that has the same terms should elicit the same type of compliance response, whether it was carefully negotiated or randomly designed. This (admittedly strong) assumption underlies existing empirical work on compliance, as prior empirical work controls for the content of the agreement (as do we), but not for the negotiation process.

Nevertheless, it remains theoretically plausible that the process of agreement negotiation influences compliance. Most prominently, Thomas Franck argues that problematic negotiation processes reduce international agreements' compliance pull, because the absence of ritual, pedigree and symbolic validation can create the perception that an agreement is unfair and illegitimate.⁶⁵ If negotiation and compliance are closely interlinked, and governments called on to implement the Paris Principles were aware of and troubled by their limited role in the Principles' drafting, then we should see even larger effects in agreements negotiated with more buy-in from states. That is, our finding that strong recommendations shape state behavior should be even more pronounced in other contexts. In contrast, our more tentative finding that weak recommendations are no different from omissions might not travel as well. Further research on how international agreements are negotiated, and on the relationship between negotiation and compliance is much needed, especially since a small but growing number of agreements, especially non-binding ones, are now drafted primarily by non-governmental actors. In short, concerns about generalizability are somewhat mitigated by the fact that the Paris Principles constitute a hard test case for our theory.

VI. Conclusions and Implications

This study documents a remarkable transformation in the world of human rights monitoring. In the early 1990s, the UN General Assembly proposed that all countries establish national monitoring bodies, to help close the gap between human rights ambitions and uneven

⁶⁵ 1990, 91-110.

practices on the ground. NHRIs spread around the world – from approximately twenty before 1990 to over one hundred today. Not only is the spread of NHRIs surprising; so is the growing homogeneity in their design. Before 1990, NHRIs took varied forms. Prominent examples included the commission model, which focused on shaping legislation, and the ombudsman model, which focused on processing individual complaints through the administrative state. For idiosyncratic reasons, the UN favored the commission model over the ombudsman model, and thus the commission model spread widely, even to parts of the world where ombudsmen had traditionally been powerful.

We find that negotiators’ choices to strongly recommend particular NHRI safeguards were rushed and unexpected. Moreover, these recommendations were placed in a non-binding General Assembly resolution, rather than in a legally binding treaty with strong implementation mechanisms. Nevertheless, both democratic and authoritarian governments in very different parts of the world, including in regions with prominent alternative templates for NHRI design, followed the Paris Principles’ strong recommendations.

In contrast, negotiators’ weak recommendations failed to produce the hoped-for results. Governments may have simply ignored weak recommendations. Our data are also consistent with a more disconcerting theory. Governments most likely to be constrained by the establishment of powerful NHRIs – authoritarian governments – may have redirected their efforts because the Paris Principles gave different weight to different institutional safeguards. Authoritarian governments may have reduced their efforts on safeguards that were weakly recommended or omitted from their Principles, to offset some of their increased efforts on safeguards that were strongly recommended.

Our findings add much-needed empirical data to important debates on norm diffusion and human rights. It is often difficult to test important theories about norm diffusion, because norms develop slowly and in similar ways in many parts of the world. However, we pinpoint the moment when a particular set of norms was born, and trace their influence. As we do not use a binary dependent variable, but study 22 dimensions of institutional design, we build on constructivist work on isomorphism - or how institutions with similar forms spread.⁶⁶ We identify a top-down international mechanism that displaced alternative regional transmission mechanisms, and led to significant homogeneity around the world.

We suspect that strongly recommended features spread quickly in part because a peer-network gave “A” grades to NHRIs that included all of the Principles’ strong recommendations. These letter grades did not have major material consequences for states. For example, no foreign lending or aid seems to have been directly conditioned on a good grade, only status and speaking rights in certain UN bodies. Nevertheless, letter grades may have helped advocacy groups draw attention to under-performing NHRIs, because a simple summary grade from an external monitor helps focus diverse audiences’ attention. We thus add to an emerging literature on indicators that suggests that advocacy may be especially effective when information is summarized in a number, a letter grade, or a ranking.⁶⁷ That said, the more powerful non-governmental standard-setting bodies become, the more we should worry about their accountability and grading methods. In particular, important recent work suggests that, over time, human rights monitors like Amnesty International look harder for abuse, and classify more acts as abuse, thus masking some

⁶⁶ Finnemore 1993.

⁶⁷ Linos 2011, 2013; Kelley and Simmons (forthcoming).

progress.⁶⁸ Our study adds to this literature by offering firm language as a tool that may not only properly incentivize governments, but could also facilitate consistent grading.

We also contribute to human rights debates. While there is ample data connecting human rights agreements to human rights practices on the ground, not all of it points in the same direction.⁶⁹ By focusing our attention on NHRIs as a key intervening institution connecting international pronouncements with the protections of rights at the national level, and improving on causal identification, we can understand states' decisions to respond to international agreements step by step. We show that states responded strongly to the international agreement we studied, though they may have also responded strategically. By promoting a template that enshrined particular institutional design features as essential, the Paris Principles provided a baseline for the institutional design of NHRIs. Although this baseline was idiosyncratic, once a national structure is in place it can be more difficult for governments to violate human rights undetected and to resist human rights norms as illegitimate, as a body of case study work, and some preliminary survey work we include above, documents. Nevertheless, further research is needed to explain exactly how particular institutional features safeguards contribute (or fail to contribute) to NHRI effectiveness.

With some additional assumptions, outlined above in the section on scope conditions and generalizability, we can also shed light on debates on treaty negotiation and compliance with international agreements more generally. Scholars have long debated whether international agreements, and international norms more generally, influence governments' decisions. We offer evidence that a non-binding and hastily negotiated agreement can have large effects on state behavior. Although our evidence is drawn from our single agreement, we believe these findings could extend to other agreements both because negotiation and compliance are often separable, and because the Paris Principles constitute a hard test case for our claim. Effects in this context are surprising, as several theorists expect compliance to be greatest when agreements are legally binding, and the negotiation process is thorough. They are also surprising given prior work suggesting that compliance is especially challenging in the area of human rights. Given these caveats, perhaps the small net effects of treaty ratification on national human rights practices are not due to government indifference to international law, as some prior scholarship assumes. Indeed, rigorous new work suggests that, over time, human rights monitors have been looking harder, and classifying more actions as abusive, perhaps inadvertently concealing progress.⁷⁰ We provide an additional potential explanation for the puzzle of apparently limited treaty effects: small net effects could conceal larger strategic responses to individual agreement provisions that push in opposite directions. Our research implies that studying the effects of individual agreement provisions may yield very different conclusions than studying the net effects of an entire agreement.

The final contribution concerns the process of international negotiations. When should negotiators settle on an agreement in which important terms are included in flexible language? Prior work emphasizes a key advantage of flexible agreement structure: namely, that it can encourage a broader range of countries to join in.⁷¹ In the human rights field, widespread participation could lead to gradual progress over time and to norms that derive their legitimacy

⁶⁸ Fariss 2014.

⁶⁹ Simmons 2009; Hafner-Burton and Tsutsui 2005; Hathaway 2002; Fariss 2014.

⁷⁰ Fariss 2014.

⁷¹ Rosendorff 2005.

from their near-universal acceptance. But widespread, formal acceptance of obligations decoupled from any implementation efforts could also undermine the entire regime. If some of the strong recommendations we study had been more flexible, might more countries have created NHRIs earlier on? We explore this question in Appendix G, by studying why a dozen countries we would expect to have active NHRIs do not. Our qualitative data suggest that a different template would not have mattered in most of these cases, as obstacles such as civil war, brutal dictatorships, and general disengagement from the international community would have probably impeded the creation of any human rights agency. Nevertheless, we find some tentative evidence that three important countries, Pakistan, Iran, and China might have set up NHRIs more quickly if the Paris Principles' strong recommendations had been phrased more flexibly.

While flexibility may have broader participation as an important advantage, our data point to an underestimated cost. Many international negotiators and academics believe that even flexible language can help advance agreement goals, by leading some states to improve their behavior gradually, and serving as a complement to other more precise agreement terms. Our data suggest that language intended to introduce some flexibility for states dramatically reduces the impact of those provisions on state behavior, and thus that the tradeoffs between firm and flexible language are starker than previously believed. We can only observe this finding among countries willing to set up an NHRI. But because the data we have indicates that authoritarian regimes are especially likely to distinguish sharply between firm and flexible terms, we suspect that other authoritarian regimes that have not yet adopted NHRIs might also ignore or reduce efforts when tasks are specified flexibly. We thus end on a cautionary note, and suggest that we focus more attention and research on potential downsides of compromises struck in the course of international negotiations.

References

- Abbott, Kenneth W., Robert O. Keohane, Andrew Moravcsik, Anne-Marie Slaughter, and Duncan Snidal. 2000. The Concept of Legalization. *International Organization* 54 (3): 401–19.
- Abbott, Kenneth W. and Duncan Snidal. 2000. Hard and Soft Law in International Governance. *International Organization* 54 (3): 421–56.
- Biblioteca del Congreso Nacional de Chile. 2009. *Del Instituto Nacional de Derechos Humanos. Historia de la Ley No. 20.405, 10 Dec.*
- Böhmelt, Tobias and Gabriele Spilker. 2013. The Impact of Preferential Trade Agreements on Governmental Repression Revisited. *Review of International Organizations* 8 (3): 343-361.
- Brewster, Rachel. 2010. Stepping Stone or Stumbling Block. *Yale Law and Policy Review* 28 (2): 245–312.
- Cardenas, Sonia. 2014. *Chains of Justice: The Global Rise of State Institutions for Human Rights*. Philadelphia: University of Pennsylvania Press.
- Carver, Richard. 2000. *Performance and Legitimacy: National Human Rights Institutions*. Versoix: International Council on Human Rights Policy.
- Carver, Richard. 2005. *Assessing the Effectiveness of National Human Rights Institutions*. Geneva: International Council on Human Rights.
- Chayes, Abram, and Antonia Handler Chayes. 1993. On Compliance. *International Organization* 47 (2): 175–205.
- Cole, Wade M., and Francisco O. Ramirez. 2013. Conditional Decoupling: Assessing the Impact of National Human Rights Institutions, 1981 to 2004. *American Sociological Review* 78(4): 702–725.
- Courtenay Conrad, Jacqueline DeMeritt, and Christopher Fariss. 2015. Why International Human Rights Advocacy Doesn't Always Decrease Government Repression. (Working Paper)
- Domingo, Pilar. 2006. Weak Courts, Rights and Legal Mobilization in Bolivia. In *Courts and social transformation in new democracies: an institutional voice for the poor?*, edited by Roberto Gargarella, Pilar Domingo and Theunis Roux, 233–54. Burlington: Ashgate.
- Downs, George W., David M. Rocke and Peter N. Barsoom. 1996. Is the good news about compliance good news about cooperation? *International Organization* 50 (3): 379–406.
- Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
- Fariss, Christopher J. 2014. Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability. *American Political Science Review* 108(2): 297-318.
- Fariss, Christopher J. and Keith E. Schnakenberg. 2014. Measuring Mutual Dependence Between State Repressive Actions. *Journal of Conflict Resolution* 58(6): 1003-1032.

- Finkel, Evgeny. 2012. The Authoritarian Advantage of Horizontal Accountability. *Comparative Politics* 44 (3): 291–310.
- Finnemore, Martha. 1993. International Organizations as Teachers of Norms: the United Nations Educational, Scientific and Cultural Organization and Science Policy. *International Organization* 47 (4): 565–97.
- Franck, Thomas M. 1990. *The Power of Legitimacy Among Nations*. Oxford: Oxford University Press.
- Gamble, John King Jr. 1985. The 1982 United Nations Convention on the Law of the Sea as Soft Law. *Houston Journal of International Law* 8 (1): 37–47.
- Ginsburg, Tom and Tamir Moustafa. 2008. *Rule by Law: The Politics of Courts in Authoritarian Regimes*. Cambridge: Cambridge University Press.
- Gold, Joseph. 1983. Strengthening the Soft International Law of Exchange Arrangements. *American Journal of International Law* 77 (3): 443–89.
- Goldsmith, Jack, and Eric Posner. 2006. *The Limits of International Law*. Oxford: Oxford University Press.
- Guzman, Andrew T. 2008. *How International Law Works*. Oxford: Oxford University Press.
- Guzman, Andrew T. and Timothy L. Meyer. 2010. International Soft Law. *Journal of Legal Analysis* 2 (1): 171–225.
- Hafner-Burton, Emilie. 2005. Trading Human Rights: How Preferential Trade Agreements Influence Government Repression. *International Organization* 59 (3): 593–629.
- Hafner-Burton, Emilie and Kiyoteru Tsutsui, 2007. Justice Lost! The Failure of International Human Rights Law To Matter Where Needed Most. *Journal of Peace Research* 44 (4): 407–25.
- Hafner-Burton, Emilie M. and James Ron. 2009. Seeing Double: Human Rights Impact through Qualitative and Quantitative Eyes. *World Politics* 61 (2): 360–401.
- Hafner-Burton, Emilie M., Laurence R. Helfer, and Christopher Fariss. 2011. Emergency and Escape: Explaining Derogations from Human Rights Treaties. *International Organization* 65(4): 673-707.
- Hafner-Burton, Emilie. 2013. *Making Human Rights a Reality*. Princeton: Princeton University Press.
- Hathaway, Oona. 2002. Do Human Rights Treaties Make a Difference? *Yale Law Journal* 111 (8): 1935–2042.
- Institución Nacional de Derechos Humanos. 2009. *Law no. 18.446*, 27 January.
- Kahler, Miles. 2000. Conclusion: The Causes and Consequences of Legalization. *International Organization* 54 (3): 661–83.
- Keck, Margaret E. and Kathryn Sikkink. 1998. *Activists Beyond Borders*. Ithaca: Cornell University Press.

- Kelley, Judith G. and Beth A. Simmons. Forthcoming. Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science*.
- Kim, Dongwook. 2013. International Nongovernmental Organizations and the Global Diffusion of National Human Rights Institutions. *International Organization* 67 (3): 505–39.
- Knight, Alec and Mila Versteeg. 2014. Small States in International Human Rights Law. Working paper.
- Koremenos, Barbara. 2005. Contracting around International Uncertainty. *American Political Science Review* 99 (4): 549–65.
- Lebovic, James and Erik Voeten. 2009. The Cost of Shame: International Organizations, Foreign Aid, and Human Rights Norms Enforcement. *Journal of Peace Research* 46 (1): 79–97.
- Lindsnaes, Birgit B., Lone Lindholt, and Kristine Yigen, eds. 2001. Establishment and development of the functions of national human rights institutions. Working paper. Copenhagen: The Danish Centre for Human Rights.
- Linos, Katerina. 2011. Diffusion through Democracy. *American Journal of Political Science* 55 (3): 678-95.
- Linos, Katerina. 2013. *Democratic Foundations of Policy Diffusion: How Health, Family and Employment Laws Spread Across Countries*. Oxford: Oxford University Press.
- Lupu, Yonatan. 2013a. Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements. *International Organization* 67(3): 469-503.
- Lupu, Yonatan. 2013b. "The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects." *American Journal of Political Science* 57(4): 912-925.
- National Congress of Chile. 2009. *Law No. 20405: Del Instituto Nacional de Derechos Humanos*, 10 December.
- Odell, John S. 2001. Case Study Methods in International Political Economy. *International Studies Perspectives* 2 (2): 161–176.
- Okafor, Obiora. 2002. On Legalism, Popular Agency and ‘Voices of Suffering’: The Nigerian Human Rights Commission in Context. *Human Rights Quarterly* 24 (3), 662–720.
- Pegram, Thomas. 2008. Accountability in Hostile Times: the Case of the Peruvian Human Rights Ombudsman 1996–2001. *Journal of Latin American Studies* 40 (1): 51–82.
- Pegram, Thomas. 2012. National human rights institutions in Latin America: Politics and institutionalization. In *Human Rights, State Compliance, and Social Change: Assessing National Human Rights Institutions*, edited by Ryan Goodman and Thomas Pegram, 210–40. New York: Cambridge University Press.
- Report of the International Workshop on National Institutions for the Promotion and Protection of Human Rights, Paris 7-9 October 1991. 1991. UN Doc. No. E/CN.4/1992/43, 16 December.
- Rosendorff, Peter. 2005. Stability and Rigidity: Politics and the Design of the WTO’s Dispute Resolution Procedure. *American Political Science Review*, 99(3): 389-400.

- Siavelis, Peter. 2000. *The President and Congress in Postauthoritarian Chile: Institutional Constraints to Democratic Consolidation*. University Park: Pennsylvania University Press.
- Sikkink, Kathryn and Carrie Booth Walling. 2007. The Impact of Human Rights Trials in Latin America. *Journal of Peace Research* 44 (4): 427–45.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights*. Cambridge: Cambridge University Press.
- Solomon, Peter H. 2007. Courts and Judges in Authoritarian Regimes. *World Politics*: 60 (1): 122-45.
- Vásquez, Marcelo Rojas. 2008. Instituciones Nacionales de Derechos Humanos. In *Informe Anual Sobre Derechos Humanos en Chile 2008*, edited by Marcelo Rojas Vásquez, 539–72. Santiago: Centro de Derechos Humanos, Universidad Diego Portales.
- United Nations. 1978. United Nations Document ST/HR/SR.A/2.
- United Nations Commission on Human Rights. 1990. RES/1990/73, 7 March.
- United Nations Commission on Human Rights (1). 1992. *Principles Relating to the Status and Functioning of National Institutions for the Promotion and Protection of Human Rights*. RES/1992/54, 3 March.
- United Nations Commission on Human Rights (2). 1992. Document E/CN.4/1992/SR.52, 27 April.
- United Nations Commission on Human Rights (3). 1992. Res. E/CN.4/1992/43/Add.1, 23 January.
- United Nations General Assembly. 1978. *National institutions for the promotion and protection of human rights*. RES/33/46, 14 December.
- United Nations General Assembly. 1993. *National institutions for the promotion and protection of human rights*. A/RES/48/134, 20 December.
- Wallace, Geoffrey. 2013. International Law and Public Attitudes towards Torture: An Experimental Study. *International Organization* 67 (1): 105–40.
- Weeks, Jessica. 2008. Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization* 62 (1): 35–64.