International Monitoring and Domestic Accountability: Evidence from the United Nations Human Rights Council*

Averell Schmidt†

August 6, 2025

Word Count: 9,529 (excluding reference list)

Abstract

This article examines how international monitoring affects legal accountability for human rights violations. International organizations often publicize human rights violations to pressure non-compliant states to change their behavior; however, recent research suggests that publicity can prompt intransigence or backlash, undercutting the advancement of human rights. Drawing on the sociological concept of reactivity, I argue that states respond to monitoring strategically: they adopt some visible accountability measures to signal their intention to comply, while also attempting to reduce the risk that future violations will be uncovered. To test this argument, I leverage a lottery used to determine the schedule of Universal Periodic Review, a human rights monitoring mechanism implemented by the United Nations Human Rights Council. Consistent with my argument, I find that monitoring increases the likelihood that political prisoners are released and that government officials are prosecuted for past violations, but decreases the likelihood that truth commissions are created.

*I thank Geoff Dancy, Rachel Hulvey, Bo Won Kim, Jialu Li, Ruofan Ma, RyuGyung Park, Kathryn Sikkink, and audiences at Oxford University, Brown University's Watson Institute, and the International Studies Association for feedback on this project. Any errors are my own.

†Assistant Professor, Department of Government and Brooks School of Public Policy, Cornell University, aschmidt@cornell.edu

1 Introduction

A core insight of international relations scholarship is that international institutions can foster cooperation by monitoring state behavior and publicizing the violation of international laws. In doing so, international institutions reduce uncertainty among states, increase the expected costs of non-compliance, facilitate enforcement, and socialize states to international norms (Keohane, 1984; Finnemore and Sikkink, 1998). The collection and dissemination of information is thus central to the design and operation of many international institutions (Koremenos, Lipson and Snidal, 2001; Dai, 2002; Koremenos, 2016). However, recent research has suggested a dark side to transparency, indicating that the decision to disclose transgressions is colored by perverse incentives (Nutt and Pauly, 2021) and can even undermine international order (Carnegie and Carson, 2018).

These critiques have been especially acute in the domain of human rights, where "naming and shaming" is typically the sole mechanism available to enforce international laws and norms. Although some research shows that activists, governments, and international organizations can leverage compliance information to pressure recalcitrant states to change their policies (Keck and Sikkink, 1998; Murdie and Davis, 2012), other scholars have found that international pressure can be counterproductive (Hafner-Burton, 2008; Lupu, 2013*a*; Snyder, 2020) and that its application is rife with political bias (Lebovic and Voeten, 2006; Terman and Byun, 2022). Alongside these conflicting empirical claims have been recurrent debates about appropriate methods and standards of evidence (Hafner-Burton and Ron, 2009; Strezhnev, Kelley and Simmons, 2021). In this article, I seek to reconcile these contradictory claims by weaving together new theory and data with design-based causal inference.

Rather than either simply advancing or undercutting human rights, I argue that international monitoring compels states to navigate a policy balancing act. On one hand, monitoring incentivizes states to implement policy reforms that signal their intention to comply with international standards. On the other hand, publicizing human rights violations through international monitoring raises the expected costs of noncompliance, encouraging states to avoid policies that increase transparency or could lead to the exposure of transgressions. Demonstrating compliance while working to avoid future detection is a difficult policy line to walk. It requires policymakers empowered to design and implement policies promptly and independently. The governments best able to strike this balance are those in which the executive is least legally constrained.

Testing this argument requires addressing two longstanding empirical challenges. First, states are often targeted with publicity precisely because they are violating human rights. The relationship between international scrutiny and human rights outcomes is consequently confounded by selection, a problem that biases both cross-national statistical analyses and the choice of country cases for qualitative and survey research. Second, most studies measure outcomes in terms of states' aggregate human rights practices; however, in contexts of limited state capacity and principle-agent problems, decision-makers have a little ability to curtail abuse in the short-run. Instead, decision-makers respond to scrutiny by varying the implementation of policy reforms aimed at addressing past abuse. Identifying the effects of monitoring requires measuring these policies directly.

I address both issues in this article. I leverage a natural experiment occurring at the United Nations to address issues of selection. The UN overhauled its human rights system in 2006, creating the Human Rights Council to replace the Commission on Human Rights, which had been "undermined by the politicization of its sessions and the selectivity of its work" (Annan, 2005). A key reform was the introduction of Universal Periodic Review (UPR), a mechanism in which each UN member state undergoes regular reviews of its human rights practices. The General Assembly created UPR with the twin goals of ensuring "universal coverage and equal treatment of all States" (UNGA, 2006). In pursuit of these goals, the Human Rights Council conducted a lottery to determine when each state would undergo UPR. I combine this lottery with new measures of human rights accountability reforms to assess the impact of international monitoring on state behavior.

I find that UPR causes states to grant more amnesties to political opposition figures and initiate more domestic trials of government officials for human rights violations; however, UPR also causes states to create fewer truth commissions, which is consistent with states working to conceal incriminating evidence of abuse. I show that these effects are conditional on the degree to which executives act independently of judicial constraints. The more independent decision-makers are, the more likely UPR is to cause them to adopt human rights accountability measures. Furthermore, across a wide range of measures, I find no evidence of human rights backlash. UPR does not impact – neither positively nor negatively – state respect for human rights; instead, it shapes the implementation of human rights accountability. These findings are consistent with a straightforward theoretical explanation: when states become the focus of international monitors, they attempt to comply with laws and norms when possible and conceal violations when necessary.

My argument and findings foreground the policymaking process in explaining how international monitoring affects government behavior. Prior research suggests that regime type and domestic legal institutions moderate the effects of advocacy campaigns and international law on states' human rights practices (Powell and Staton, 2009; Hendrix and Wong, 2013). Here, I qualify these insights, showing that when executive-level decisions-makers operate independent of legal constraints, they are more likely to respond to human rights monitoring by adopting reforms that cast their regime in a positive light. These findings indicate that international monitoring can direct government resources toward human rights accountability, possibly mitigating human rights abuse due to limited capacity or misgovernance (Chayes and Chayes, 1993; Slough and Fariss, 2021). At the same time, these findings suggest that states' implement human rights accountability strategically to bolster their international reputations for compliance (Morse and Pratt, 2022).[1]

This article also advances ongoing debates over the effectiveness of UPR in the struggle to promote human rights. Policymakers greeted the creation of UPR with great expectations.

[1] Espeland and Sauder refer to this as "gaming strategies" (2007, 3).

Secretary-General Ban Ki-moon claimed UPR had "the greatest potential to promote and protect human rights in the darkest corners of the world" (Hafner-Burton, 2013, 52). This optimism was short-lived. UPR has failed to ensure states' "equal treatment" (Terman and Voeten, 2018; Carraro, 2019) and, by empowering states to serve as prosecutor, judge, and jury in the application of human rights standards, it has remained marred by politicization (Terman and Byun, 2022). Despite these shortcomings, the evidence presented here suggests UPR has an observable effect on human rights accountability, prompting states to grant amnesties to political opposition figures and to prosecute their own agents for past abuse. Likewise, I find no evidence that UPR causes a backlash by undercutting state human rights practices. These findings indicate that even politicized international institutions can contribute to the advancement of human rights.

2 The Promise and Peril of Human Rights Monitoring

A large literature in international relations and comparative politics seeks to explain the determinants of human rights violations. This research provides two alternative accounts of how international monitoring affects government human rights practices and the implementation of domestic accountability measures.

One account suggests that monitoring can improve human rights outcomes. It can do so via two mechanisms. First, a large body of research indicates that government decision-makers are rational actors seeking to maximize their chances of survival and returns to office (Poe and Tate, 1994; Bueno de Mesquita et al., 2003). If international monitoring increases the costs of non-compliance, then it can alter these decision-makers' "calculus of abuse," making human rights violations less likely to occur (Hafner-Burton, 2013, 19). Second, exposure to international monitoring can diffuse policy ideas, spread best practices, and socialize states to standards of appropriate behavior (Johnston, 2001; Greenhill, 2010). The dissemination of compliance information generally – and naming and shaming human rights violations specifically – can dissuade abuse by directing repressive governments toward appropriate remedies and by coordinating the policies of other actors toward the repressive government (Keck and Sikkink, 1998). The expectations of this account are consistent with prior research showing a positive impact of international advocacy campaigns on respect for human rights (Franklin, 2008; Murdie and Davis, 2012).

Existing research also suggests that governments with robust democratic and domestic legal institutions are most responsive to international monitoring. There is a well-established, positive association between democracy and respect for human rights (Davenport, 2007*a*). Democracies with open political systems are most susceptible to the types of legalized domestic political mobilization empowered by international human rights law (Simmons, 2009), and human rights treaties are more effective when executive authority is checked by other domestic veto players (Lupu, 2015). Legal challenges to state policy and alternative centers of decision-making authority are much less likely to exist in more authoritarian contexts, limiting the channels by which international monitoring can affect government human rights practices.

A second account suggests that international monitoring is more likely to cause human rights intransigence or backlash than normatively positive change (Vinjamuri, 2017, 116-118). Several quantitative studies have found that human rights criticism either has no effect or worsens human rights violations (Hafner-Burton, 2008; Hafner-Burton and Tsutsui, 2005). International shaming can prompt actors to double-down on prior beliefs and resist outgroup pressure (Snyder, 2020), especially if shame is cast by an adversary or rival (Terman, 2023).[2] Velasco (2023) shows how international institutions have fostered backlash against LGBT+ rights, suggesting that international institutions can both facilitate the diffusion of human rights and provide institutional structures that help states counteract them. These findings are consistent with the response of some states to UPR. Equatorial Guinea, for example,

[2] But Guess and Coppock (2020) find evidence inconsistent with backlash in survey experiments.

responded to UPR by increasing oppression and arresting arbitrarily local LGBT+ activists.[3] This evidence suggests international monitoring is at best ineffective – and possibly even counterproductive – to the advancement of human rights.

In this article, I develop an alternative to these two prevailing accounts. Rather than simply advancing or undermining human rights, I argue that monitoring – and UPR specifically – has a *mixed* impact on government human rights policies. Governments time the adoption of accountability measures to coincide with periods of international scrutiny in order to enhance their reputations for compliance with international human rights standards. Some of these measures – such as the initiation of domestic trials or the amnesty of opposition figures – are costly signals that affect human rights outcomes on the ground. Others – like the withholding of truth commissions – are attempts to guard against unfavorable evidence coming to public light through the review process. Monitoring, in short, causes states to strategically adopt legal accountability measures that they otherwise would not.

3 A Theory of Reactivity and Legal Accountability

My argument widens the existing theoretical aperture in two respects. First, I look beyond conventional indicators of government human rights practices to assess how monitoring impacts the adoption of accountability measures used to address past abuse, such as the creation of truth commissions, the issuance of amnesties, or the prosecution of state agents for prior violations. Second, I argue that governments can change their behavior either in anticipation of future monitoring or in response to monitoring when it occurs. My theory is attuned to both possibilities.

I develop my theoretical argument in two parts. I first claim that UPR is a form of monitoring that affects state policy via social pressure. I then explain how the ability of

[3] U.S. Department of State, "Background & Social Media Consideration: Equatorial Guinea: Arbitrary Arrests of LGBTI Citizens Spark Concern; Reaction to UPR?," July 18, 2019, Case No. FL-2021-00526, Doc No. A-00000462397.

decision-makers to implement policies independently shapes the responsiveness of the policy process to monitoring and, thus, the observed effects of UPR.

3.1 Human Rights Monitoring as Social Pressure

An established literature indicates that international monitoring affects state policy. By collecting and disseminating information on whether states are complying with international standards, rational functionalist theories suggest that monitoring reduces uncertainty and incentivizes states to cultivate cooperative reputations (Keohane, 1984; Guzman, 2008). But monitoring does not just affect state policy by altering the costs of non-compliance; it can also act as an engine of socialization by pressuring states to adopt new policies and ideas (Johnston, 2001; Finnemore and Sikkink, 1998, 902).

Building on scholarship on global performance indicators (Kelley and Simmons, 2015; Kelley, 2017; Kelley and Simmons, 2019), which shows how international organizations, governments, and non-state actors wield numeric evaluations and rankings to shape government policy, I argue that the type of human rights monitoring provided by UPR is a form of social pressure. Although UPR does not produce numeric measures of states' human rights policies, it does entail recurrent public monitoring and evaluation in a manner that facilitates the comparison of states with both to their own and others' past performance. Consistent with the GPI literature, I argue that UPR has the capacity to influence state policy through the mechanisms of elite shaming, transnational pressure, and domestic politics (Kelley and Simmons, 2015, 58).

Elite shaming is built into the design of UPR. UPR is a *peer review* mechanism. Each review consist of a 3.5-hour "interactive dialogue" where delegates from the state under review respond to questions, comments, criticisms, and other recommendations from representatives of other UN members states. These recommendations – and whether the state under review accepts them – are recorded in an outcome report, which forms the basis of the state's next review. States take interactive dialogue seriously. States under review send large delegations with cabinet- or ministerial-level representation. Other states participate actively in interactive dialogue. According to data from UPR Info, states issued a total of 57,966 recommendations over the course of 387 reviews during UPR's first two cycles from April 2008 through November 2016, an average of 150 recommendations per review.

Existing research on UPR conceptualizes human rights recommendations as instances of "naming and shaming" in which diplomats identify and publicize their peers' transgression of human rights norms, seeking to increase the social costs (or shame) of noncompliance (Carraro, 2017; Terman and Voeten, 2018). This research often emphasizes how the effects of shaming at UPR operate through states' diplomatic relations. For instance, Terman (2023, 27-94) develops a "relational theory of international shaming" and tests this argument, in part, by demonstrating that states' decisions to issue recommendations at UPR are shaped by their geopolitical relations with the state under review. UPR leverages social and diplomatic relations between states to pressure decision-makers in the state under review to bring their government's human rights practices in line with international laws and norms.

The UPR process is also designed to foster transnational pressure. Reviews are based on three reports submitted at least five months before interactive dialogue: a national report submitted by the state under review and two sets of reports submitted to the UN High Commissioner on Human Rights: first, reports by treaty bodies, UN special procedures, and other official UN documents; second, reports by stakeholders including national human rights institutions and non-governmental organizations.[4] Many of these stakeholders travel to Geneva to lobby UN member state delegations to broach specific issues during interactive dialogue; the NGO UPR Info, for instance, has coordinated "pre-session" meetings between over 1,500 national human rights institutions and NGO representatives and 163 permanent

[4] UN Human Rights Council, "Institution-building of the United Nations Human Rights Council," A/HRC/RES/5/1, 18 June 2007, available at: http://ap.ohchr.org/documents/E/HRC/resolutions/A_HRC_RES_5_1.doc. Stakeholders are defined as "NGOs, national human rights institutions, human rights defenders, academic institutions and research institutes, regional organizations, as well as civil society representatives"; United Nations Human Rights Council, "Information and Guidelines for Relevant Stakeholders on the Universal Periodic Review Mechanism," July 2008, available at: https://www.ohchr.org/sites/default/files/HRBodies/UPR/Documents/TechnicalGuideEN.pdf.

missions.[5] UPR links local and international NGOs with diplomats in other states, which existing research shows is critical to the power of transnational advocacy networks and human rights diplomacy (Keck and Sikkink, 1998; Murdie and Davis, 2012; Risse, Ropp and Sikkink, 2013).

UPR also exercises transnational pressure through the diffusion of policy ideas and the support it provides to states to implement these ideas. The review process helps states identify areas for human rights improvement and exposes policymakers to policy options and best-practices. After review, the Human Rights Council provides financial and technical assistance to help states implement recommendations.[6] Consistent with prior research, UPR acts as an engine of policy reform by fostering policy learning, disseminating international norms, and incentivizing states to institutionalize new policies in domestic practice (Haas, 1992; Finnemore, 1993; Greenhill, 2015).

Finally, UPR provides an exogenous shock to domestic politics, increasing the prioritization of human rights on the policy agenda. The efficacy of international laws and organizations stems, in part, from their ability to mobilize and empower pro-compliance domestic constituencies who advocate for policies in line with international standards (Gurowitz, 1999; Dai, 2007; Simmons, 2009; Ritter and Conrad, 2016). UPR does this by conveying compliance information to domestic audiences, mobilizing domestic NGOs through the UPR reporting process, and offering a focal event – interactive dialogue – that punctuates the normal course of domestic politics. For illustration, I show in appendix C that UPR causes an increase in Google searches of human rights related-topics in the state under review, indicating an increase in the salience of human rights in domestic politics (Dancy and Fariss, 2024). Diplomatic reporting within the State Department suggests that these dynamics apply even in authoritarian settings. In Burma, for instance, domestic newspapers provided extensive

[5] See, UPR Info, 2024, "Pre-sessions," available at: `https://upr-info.org/en/presessions`; UPR Info, 2016, "UPR Info Pre-sessions Empowering human rights voices from the ground," available at: `https://upr-info.org/sites/default/files/documents/2016-12/pre-sessions_web.pdf`.

[6] The Voluntary Trust Fund for Financial and Technical Assistance for the Implementation of the UPR was established by Human Rights Council, Resolution 6/17, "Establishment of funds for the universal periodic review mechanism of the Human Rights Council," 28 September 2007.

and candid coverage the country's periodic review in 2011, despite widespread censorship and oppression.[7] By increasing the domestic salience of human rights, UPR helps to reorient policy attention and resources toward the problem of human rights abuse.

3.2 Conditional Reactivity

I argue that states' responses to the social pressure generated by UPR are characterized by conditional reactivity. The concept of reactivity was developed to explain how subjects modify their behavior when they know they are being monitored or evaluated (McCall, 1984).[8] While initially seen as a threat to valid measurement (Campbell, 1957), sociologists have more recently shown that monitors can harness reactivity to shape organizational behavior. Influential research by Espeland and Sauder (2007), for instance, shows that universities have reallocate resources and adopted "gaming strategies" in response to *U.S. News and World Report* law school rankings.

States respond to monitoring in similar ways. Bisbee et al. (2019) show that the indicators created by the Millennium Development Goals led states to reallocate resources toward educational outcomes measured by the initiative, and Honig (2019, 174) contends that increased monitoring causes international development organizations, like the US Agency for International Development, to prioritize presentation over performance. I argue that reactivity is also a useful concept for explaining the consequences of UPR. However, UPR's effect on state policy is not unconditional and uniform across contexts. Organizational pathologies and legal constraints on executive policymaking shape the types of outcomes affected by UPR and where these effects are likely to occur.

Two organizational pathologies limit the impact of UPR on human rights outcomes. First, organizations that abuse human rights are often rife with principal-agent problems (Mitchell, 2004; Conrad and Moore, 2010). Decision-makers typically delegate violations in a manner

[7] U.S. Department of State, "Burma: Some Signs – though Inconsistent – of a Changing Media Environment," July 1, 2011, Case No. F-2014-22713, Doc No. C05740789.

[8] Reactivity is also known as a "Hawthorne Effect" based on the seminal experiments conducted at the Western Electric Company in Cicero, Illinois from 1924 and 1932 (Adair, 1984).

that increases plausible deniability and reduces legal liability (James, 2002; Carey, Mitchell and Lowe, 2013, 250).[9] Once abuse is delegated, repressive agents work to insulate themselves from oversight and accountability, a process Rejali refers to as "bureaucratic devolution" (2007, 526-530). As a result, human rights violations become routinized, persistent, and difficult to stop (Hafner-Burton and Ron, 2009; Hafner-Burton, 2013, 19, 35). In contexts characterized by principle-agent problems, human rights outcomes would not be affected by international monitoring in the short run, even if policymakers intended to comply with monitors' demands.

Second, many human rights violations are not due to a strategic choice to repress but follow a "logic of misgovernance," where rights violations arise because state agents are failing to do their jobs well (Slough and Fariss, 2021). Poor training could make rights violations by military officers more likely, and judicial backlogs could cause unlawful imprisonment. These same issues are often-cited drivers of the use of torture; as noted by Simmons and Creamer "[t]orture practices are multicausal, tied to deep cultural practices and governing dysfunctions, are often committed by decentralized agents, and are therefore difficult to influence under any circumstances" (2019, 1055). The same logic is reflected in legal research emphasizing how limited state capacity accounts for non-compliance with international laws (Chayes and Chayes, 1993). In the case of misgovernance, abuse would occur even under the leadership of elites intent on respecting human rights.

The mechanisms of social pressure leveraged by UPR – elite shaming, domestic politics, or transnational advocacy – do not address the principal-agent or misgovernance problems that allow persistent human rights violations. UPR is consequentially unlikely to have a direct, short-term impact on aggregate patterns of human rights abuse.

Instead, I contend that UPR affects the probability of states adopting transitional justice policies aimed at addressing past human rights abuse. Elites, domestic audiences, and

[9] Mass atrocity crimes, which are typically planned, authorized, and executed by small groups of elites, is a notable exception (Valentino, 2005); however, these crimes occur during civil wars or when regime survival is at stake, contexts where international scrutiny is less likely to outweigh other factors driving human rights violations.

transnational advocacy networks affect the establishment of truth commissions (Zvobgo, 2020), the prosecution of state officials (Sikkink, 2011), and the release of political prisoners (Gruffydd-Jones, 2021). However, some transitional justice policies do not cast states in a positive light; truth commissions can uncover new evidence of past abuse and amnesties of state officials can fuel accusations of impunity. Following the strategic logic of reactivity, I argue that states respond to the social pressure of UPR by implementing policies that signal compliance with human rights norms and avoiding those that risk new allegations of non-compliance.

States' ability to respond strategically to UPR depends on the extent to which decision-makers can act independently. An established literature shows that domestic institutions shape patterns of repression (Davenport, 2007*b*), the implementation of international laws and norms (Cortell and Davis, 1996), and the efficacy of naming and shaming campaigns (Hendrix and Wong, 2013). Domestic legal institutions are especially important (Hill and Jones, 2014). Strong judiciaries can reduce repression by restraining executive action and holding executives accountable for legal violations (Powell and Staton, 2009; Mitchell, Ring and Spellman, 2013). Democracies with robust legal institutions are thus most respectful of human rights.

This insight has a counterintuitive implication for where UPR is most likely to have the largest impact on state behavior. When domestic courts are weak, executives have more discretion over policy outcomes and there tends to be more need for transitional justice measures to address past legacies of abuse and repression, increasing the marginal effect of international law on state repression (Conrad and Ritter, 2019). As Conrad and Moore (2010) argue, governments are better able to curtail the use of torture when there are fewer veto players in the decision-making process. I argue that similar dynamics affect government responses to UPR. When executives govern under the guise of weak courts, they are able to implement policies selectively and on short-notice in an effort improve their performance at

UPR. As a result, UPR should have a greater effect on the implementation of accountability measures when executives operate independently of the judiciary.

4 Observable Implications

Each theoretical account provides distinct predictions concerning how international monitoring in the form of UPR affects human rights. The first set of distinctions concerns how UPR affects government respect for human rights, such as aggregate measures of physical integrity violations as well as indicators of specific crimes like the use of torture or extrajudicial killings. While one account predicts UPR to improve state behavior across these indicators, the backlash hypotheses anticipates a null or negative relationship between monitoring and government respect for human rights. The theory of conditional reactivity does not expect UPR to impact state human rights practices, because bureaucratic slack renders these outcomes non-responsive to short-term changes in social pressure.[10]

Unlike the alternative explanations, the theory of conditional reactivity expects monitoring to have different effects on different types of accountability measures. This argument indicates that states respond strategically to the social pressure generated at UPR. States often advocate against impunity and for the releasing political prisoners during interactive dialogue, whereas the topic of truth commissions is rarely broached.[11] The concept of reactivity suggests that states should respond to this pressure by granting more amnesties to opposition figures while increasing the prosecution of and reducing amnesties for state officials for past abuse. At the same time, states should scale back the creation of truth commissions in order to limit the availability of incriminating evidence.

[10] Hafner-Burton also argues that the efficacy of naming and shaming campaigns is limited by state capacity; consequentially, she argues these campaigns have a negative impact on government human rights practices but can have a positive effect on institutional structures, such as "government's abuse of the electoral process, its level of political pluralism and participation, and functioning government" (2008, 696). I test this argument alongside measures of state human rights practices below.

[11] During the first two cycles of UPR, only 75 comments made during interactive dialogue mentioned a "Commission of Inquiry" and only six mentioned a "Truth Commission," whereas 503 comments mentioned "impunity." Likewise, the stem/term "inquir∗" and "truth" were raised 152 and 128 times, respectively, while "prosecut∗" and "prison∗" occur 1,373 and 1,218 times.

The two alternative explanations do not anticipated UPR to have mixed effects on the implementation of legal accountability measures. The first account suggests that review should have a positive impact on the advancement of human rights, such as promoting truth commissions, opposition amnesties, and justice for past abuse by state officials. The backlash account, in contrast, expects the opposite: that UPR will either have no effect or cause states to double-down on past abuse.

The theories also differ in terms of how domestic institutions moderate the effect of international monitoring. My argument indicates that states' ability to react strategically to monitoring depends on policymaker autonomy. The impact of UPR is greatest when executives have more discretion to implement policy reforms independently; institutional checks on executive authority inhibit policy responses to UPR. In contrast, research on the impact of international law on human rights provides the opposite expectation: that monitoring is most effective when domestic political structures offer formal channels for mobilized domestic constituencies and veto players to hold executives accountable.

Some existing research hypothesizes that non-democracies are more responsive to international monitoring. Hendrix and Wong (2013) argue that non-democracies are more responsive to naming and shaming campaigns, because these campaigns introduce novel compliance information into domestic authoritarian contexts, whereas they spark backlash in democracies. Bisbee et al. (2019) argue that unaccountable governments – those with opaque or non-democratic political systems – respond to global performance indicators on education by substituting resources toward measured outcomes (primary education enrollment rates) and away from non-measured outcomes (secondary education enrollment rates).

My argument differs from these accounts both in terms of expectations and theoretical mechanisms. First, my theory expects non-response, not backlash, when international monitoring targets states whose political systems have robust checks-and-balances. Second, while novel information concerning government human rights practices may activate pro-compliance domestic political actors, I argue that UPR also impacts state behavior via transnational pressure and elite shaming. As a result, decision-makers attempt to implement policies to mitigate these pressures, and not just to please domestic audiences. Furthermore, few resources are required to free prisoners, initiate prosecutions, or authorize commissions of inquiry. While all policy choices involve tradeoffs, in the case of UPR, an explanation rooted in the concept of reactivity is more compelling than one anchored in policy substitution under recourse constraints.

5 Data and Measurement

I construct a panel dataset of all UN member states to test these competing hypotheses. My unit of analysis is, accordingly, the country-year. My panel begins in 2008, the year UPR first occurred, and ends in 2020, the final year outcome measurements are available. My treatment variable in all analyses measures whether a state underwent UPR in a given year, taking 1 in review years and otherwise 0.

The different theoretical expectations require different outcome measures. To measure the implementation of domestic accountability measures, I use a new dataset developed by the Transitional Justice Evaluation Tools (TJET) project (Dancy et al., 2024).[12] TJET is a comprehensive, global database of transitional justice policies implemented from 1970 through 2020. Since the creation of UPR in 2008, TJET includes data on 137 amnesties policies, 34 truth commissions, and 2,589 domestic human rights prosecutions of state officials.

I use these data to construct several measures of whether a state implements a new transitional justice policy in a given year. I create four measures of government amnesty policies. Opposition Amnesties and State Amnesties equal 1 if a government issues an amnesty to political opposition figures or government officials in a given year, and are otherwise 0. I further distinguish States Amnesties based on whether the amnesty is issued specifically for past violations of human rights (State HR Amnesties), the type of amnesty most often condemned at UPR. I also measure changes in the prosecution of state agents and the

[12] The full dataset is available at www.transitionaljusticedata.org.

implementation of truth commissions. The variable Truth Commissions takes 1, and is otherwise 0, if a new truth commission is created in a given year. Domestic HR Trials equals 1, and is otherwise 0, if a state begins prosecuting government officials for human rights abuse in a given year and had no prosecutions in the prior year.

I drawn on several sources to measure variation in state respect for human rights. For physical integrity violations, I use the latent variable measure of respect for human rights developed by Fariss (2014) and Fariss, Kenwick and Reuning (2020) as well as measures of freedom from torture and freedom from political killings developed by the Varieties of Democracy Project (Coppedge et al., 2024). These are continuous variables with standard-normal distributions. I also examine the effect of UPR on institutional structures, because Hafner-Burton (2008) finds that naming and shaming affects the electoral process, political pluralism, and participation. I assess whether UPR affects civil society participation, participatory democracy and direct popular vote as measured by the Varieties of Democracy Project. These are continuous variables ranging from 0 to 1 (e.g., low to high participation). Descriptive statistics for all variables are provided in Appendix A.

6 Research Design: The UPR Lottery

Research on the effects of international laws and institutions on state behavior is often confounded by the problem of selection on unobservables (Lupu, 2013*b*; Chaudoin, Hays and Hicks, 2018). States choose when to ratify treaties and enter international organizations, making it difficult to discern whether associations between ratification and outcomes are due to institutions themselves or states' pre-existing preferences and characteristic (von Stein, 2005; Simmons and Hopkins, 2005). Two unique features of the design and implementation of UPR help address these problems: its universality and a lottery used to assign states' review dates. Here, I explain these features and how I leverage them to estimate the effect of UPR on government human rights practices.

On March 15, 2006, the United Nations General Assembly adopted resolution 60/251, establishing the Human Rights Council and the mechanism of Universal Periodic Review. The resolution charged the HRC to undertake UPR in a manner that "ensures universality of coverage and equal treatment with respect to all States." In pursuit of these twin goals, the HRC held a lottery on September 21, 2007, to determine the order states would undergo review.[13] This lottery differed from a simple random ordering in three respects.

First, the HRC block randomized by region. The first cycle of UPR occurred over 12 sessions, beginning in April 2008 and ending October 2011; 16 UN member states underwent review during each session.[14] To ensure geographic representation within each session, the HRC set a fixed number of states to be reviewed from each UN regional grouping during each session.[15] For example, during UPR's first session, four states were reviewed from Africa, two from Asia, three from Latin America and the Caribbean, two from Eastern Europe, and three from Western Europe and Other States. This approach to the lottery not only ensures equal geographic treatment, it also improves the precision of statistical estimates, because region is associated with many factors – socio-economic development, regime type, conflict legacy, and more – that affect government human rights practices.

Second, following HRC resolution 5/1 of June 2007, states that were HRC members in 2007 or elected to join the HRC in 2008 were required to undergo review during the first session of the last year of their term on the HRC. For example, states whose HRC terms ended in 2007 or 2008 were assigned the earliest sessions in 2008, and states whose term ended in 2010 were assigned to the earliest sessions in 2010.[16] This rule determined the year

[13] The modalities of the lottery are detailed in the HRC's report on its sixth session; UN Human Rights Council, "Report of the Human Rights Council on its 6th session," 2007, Geneva, paragraphs 291-294 (page 132) and Annex V (pages 205-209), available at, https://digitallibrary.un.org/record/626770?ln=en. On June 17, 2011, the HRC adopted decision 17/119 fixing this order for all subsequent UPR cycles.

[14] The number of states reviewed each session was reduced to 14 after the first cycle; however, the order was not changed.

[15] See Appendix B for block randomization schedule and states' regional groupings.

[16] States that were HRC members in 2007 are bolded in the list of regional groupings provided in Appendix B, and the year their term expired is provided in parentheses.

that 56 states underwent UPR. Critically, this rule was set *after* these states joined the HRC and, I argue, is exogenous to their human rights practices.

Third, the order of some states was not determined by lottery. The HRC invited states to volunteer to undergo UPR first. Only two states accepted this invitation: Switzerland and Colombia. These states were reviewed in the third UPR session. Haiti was moved to the end of the first cycle following the 2010 earthquake. South Sudan was then placed in the order after Haiti at the end of the second cycle once it gained independence in 2011.

All states have complied with their assigned review dates. Only once – Israel during the second cycle – has a state refused to show up on its assigned review date. Even in this case, however, Israel's noncompliance was short lived: it skipped it's initial assigned review during the 15th UPR session on January 29, 2013, but then agreed to participate in a make-up review during the 17th session on October 29, 2013. As a result, these events did not alter the year of Israel's review and, thus, did not affect its treatment status. During the third cycle Israel returned to its place in the original, randomly assigned order.

I argue that this lottery renders states' review dates independent of other determinants of their human rights practices. Consequentially, differences between reviewed and non-reviewed states can be interpreted as causal effects of UPR. I estimate causal effects using a two-way fixed-effects model of the form,

$$Y_{i,t} = \alpha_i + \gamma_t + \tau D_{i,t} + \sum_j \eta_j X_{j,i,t} + \epsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ is the outcome variable of state i in year t, $D_{i,t}$ is an indicator that takes 1 if state i underwent review in year t, α_i and γ_t are state and year fixed effects, $X_{j,i,t}$ are the j control variables measured country i in year t, and $\epsilon_{i,t}$ is an error term. In analyses of the effect of UPR on human rights accountability measures, control variables are states' UNGA voting ideal points (Bailey, Strezhnev and Voeten, 2017), electoral democracy index (Coppedge et al., 2024), level of respect for human rights (Fariss, 2014), and the log of states' GDP per capita and log of population (Fariss et al., 2022); in analyses of the effect of UPR on government human rights practices or institutional structures, I exclude controls for electoral democracy and respect for human rights. Given the random assignment of states review dates, identification does not require these controls or fixed effects; however, their inclusion increases the precision of estimates (Gerber and Green, 2012).[17] τ, the effect of review on the respective outcome, is the quantity of interest.

I also leverage the UPR lottery to conduct sensitivity analysis via randomization inference in which I compute the exact distribution of null results from a large number of placebo lotteries. I replicate the lottery in two ways. First, I conduct 10,000 placebo lotteries following precisely the steps used by the HRC to implement the lottery on September 21, 2007, as described above. Critically, this approach holds fixed the treatment assignment of the 56 states that were either serving on the HRC when the lottery was conducted in 2007 or that had been elected to begin serving on the HRC in 2008. This is a conservative test, because these 56 states receive their observed treatment assignment in all permutations. Second, I conduct the same number of placebo lotteries, but ignore the rule that states serving or elected to serve as HRC members in 2007 must be reviewed in the final year of their term; all other rules and deviations from simple random ordering (e.g., regional distributions, preferences of Switzerland and Colombia, and the disruption of Haiti's earthquake) are followed. This approach permits the treatment assignment of these 56 states to vary. I note the results of these analyses below; I present full results in Appendix D.

The validity of this research design depends on the comparability of reviewed and non-reviewed states. In Figure 1, I implement the equivalence approach to balance tests as described by Hartman and Hidalgo (2018) to assess the comparability of the treatment groups produced by the lottery. Black diamonds represent the standardized mean difference between the treated and control states for the control variables, regional indicators, and

[17] The use of two-way fixed-effects to estimate average treatment effects has come under recent scrutiny due to the possibility of observations taking negative weights (de Chaisemartin and D'Haultfœuille, 2020; Imai and Kim, 2021); I guard against this threat to inference by conducting non-parametric sensitivity analyses, described below.

several additional state attributes. Grey bars are 95% equivalence confidence intervals – the largest difference at which the null hypothesis of difference is rejected at $\alpha = 0.05$. The vertical lines represent the $\pm 0.36\sigma$ equivalence range proposed by Hartman and Hidalgo (2018, 1006). The absolute difference between treated and control groups is greatest for Log Population, where it is only a difference of 0.088 standard deviations. These analyses show that states under review are comparable to states not under review across a wide-range of observable characteristics. In Figure 7 of Appendix D I test covariate balance in the observed lottery to the 10,000 placebo lotteries used to conduct randomization inference. The placebo lotteries produce treatment and control groups comparable to those in the true lottery.

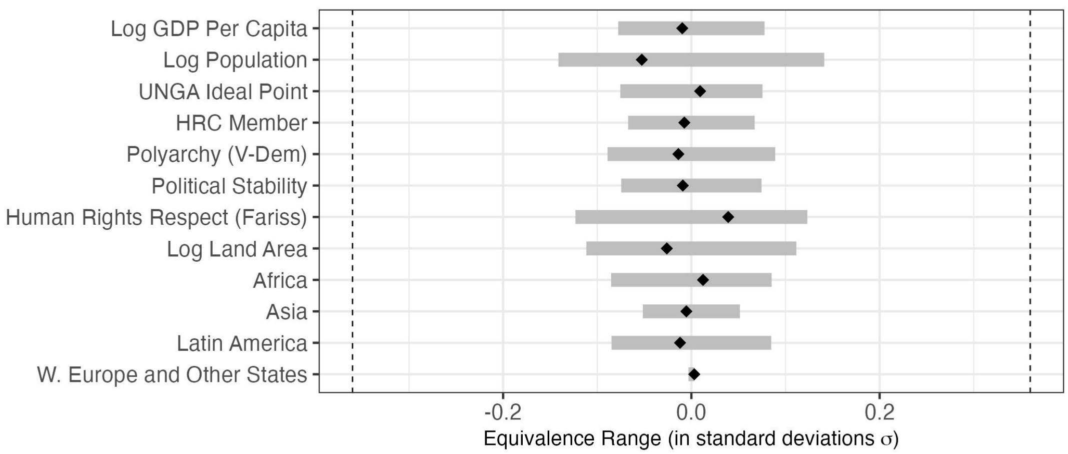

Figure 1: *Balance of Reviewed and Non-reviewed States.* Equivalence tests comparing covariate balance of states under review to those not under review. Equivalence range of $\pm 0.36\sigma$ denoted by vertical dashed lines. Standardized mean difference between treated and control states represented by black diamonds. 95% equivalence confidence intervals represented by gray bars.

7 Empirical Findings

I present my quantitative findings in three parts. First, I assess how UPR affects the implementation of legal accountability measures, showing that states respond to review in a manner that is consistent with the concept of reactivity. Second, I test whether UPR worsens government human rights practices or prompts institutional reform. I find that it does not.

Third, I assess how the effects of UPR on human rights accountability varies by the degree to which executives exercise policymaking authority independently.

7.1 Evidence of Reactivity

I begin my analysis with Figure 2. In the gray shaded region of each facet, I plot estimates – computed by equation (1) – of the effect of UPR on four legal accountability measures: opposition amnesties, state amnesties for past human rights violations, truth commissions, and domestic prosecutions of government agents.[18] In the unshaded regions of each plot, I present analyses in which I move artificially states' year of review forward or backward by 1, 2, or 3 years to assess if there is evidence that outcomes change in anticipation of or in the wake of reviews. Statistically significant estimates in these analyses would suggest that the observed effect could be either a result of temporal trends or changes that precede or follow review, not a result of review itself.

The results presented in Figure 2 are consistent with the concept of reactivity. These analyses show that states issue more amnesties to opposition figures and are more likely to initiate the prosecution of government officials in years when states undergo UPR. In contrast, they are less likely to issue amnesties to their own officials for past human rights violations and less likely to create truth commissions, which is consistent with the theory that states attempt to withhold new evidence of non-compliance when they come under the guise of the international spotlight. Furthermore, in no case do these changes precede review or persist in its wake. This suggests that UPR prompts an immediate change in the implementation of human rights accountability. These results also deviate from the expectations of the alternative explanations. Both predict a uniform pattern of response, either toward or away from compliance, and not an improvement in some indicators (opposition amnesties, state human rights amnesties, and prosecutions) and concurrent deterioration in others (truth commissions).

[18] Analyses of UPR's impact on state amnesties more generally are presented in Appendix D.

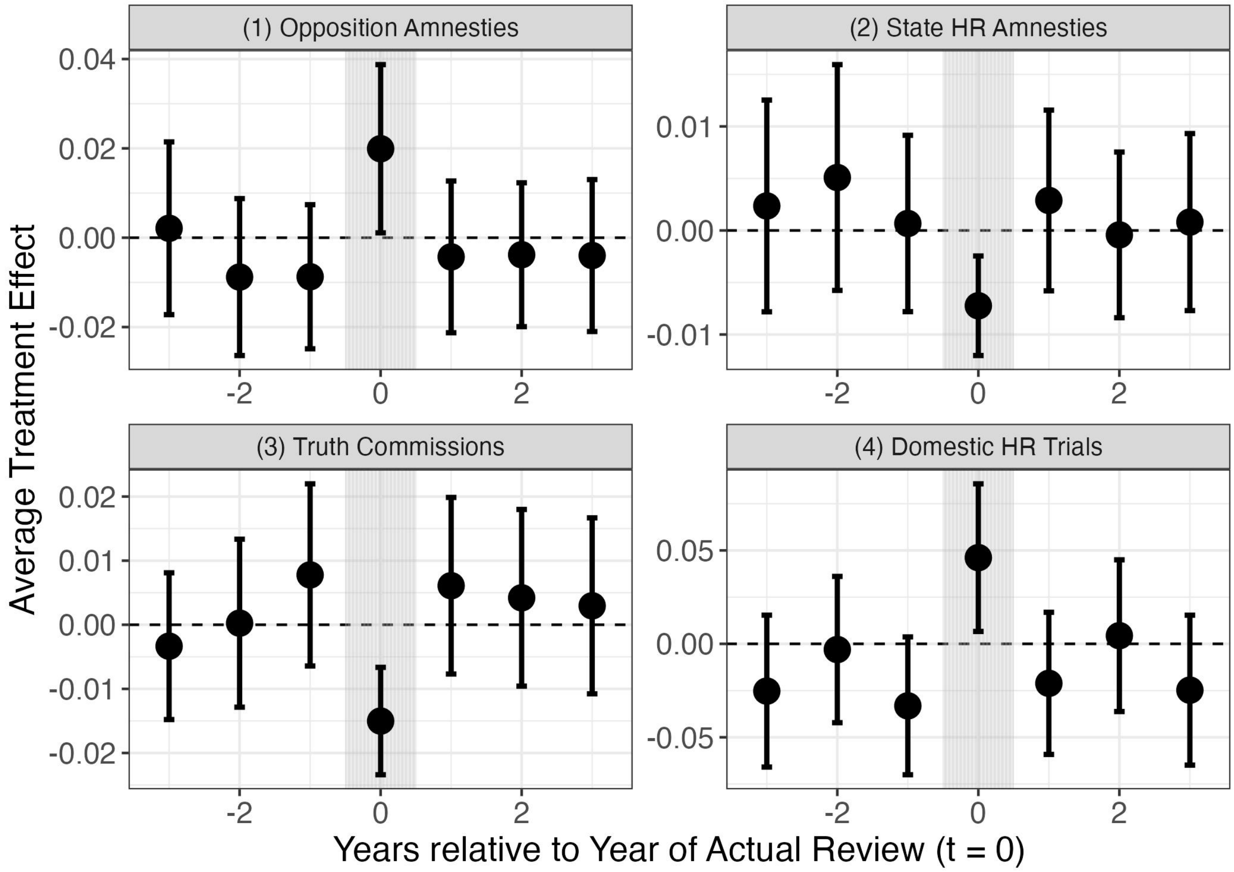

Figure 2: *Effect of UPR on Human Rights Accountability.* The impact of UPR on human rights accountability measures – estimated by equation (1) – is presented in the shaded region at $t = 0$. Analyses in which the year of review is moved artificially 1, 2, or 3 years before or after the actual year of review are presented in the unshaded region. Standard errors are clustered at the country-year.

The magnitude of these effects are small, but they are nevertheless substantively significant. For example, opposition amnesties are rare, occurring in 2.43% of country-years without review; the incidence of opposition amnesties nearly doubles in country-years undergoing review, rising to 4.44%. Likewise, review increases the probability that a government will initiate trials against its own officials from 0.135 to 0.182, an increase of over 4.72 percentage points or more than 35 percent. While 1.55% of country-years not under review created truth commissions (a total of 28 observations), only 0.21% of country-years under review (1 observation) do so. UPR nearly precludes the establishment of new truth commissions.

I present several supplementary analyses in Appendix D to assess the robustness of these findings. I use randomization inference to locate the observed estimates presented in Figure 2 in the distribution of null estimates computed from fictitious lotteries. These analyses show that the finding that UPR causes states to issue opposition amnesties is especially strong; even holding in the case of permuted lotteries that hold fixed the treatment status of the 56 HRC member and member-elect states. The effect of UPR on the initiation of truth commissions and domestic human rights prosecutions is also robust, but sensitive when the treatment status of the 56 HRC members held fixed. Furthermore, I also show in these analyses that the effect on amnesties for government officials is driven by variation in a reduction for amnesties for past human rights violations specifically, and not amnesties of government officials more generally.

Second, I use randomization inference to conduct nonparametric sensitivity analysis, comparing the observed relative risk of outcomes in states under review to the distribution of this statistic in across placebo lotteries. A nonparametric approach is instructive in this context because it avoids arbitrary modeling decisions that have long been at the heart of contradictory findings in the empirical literature and is robust when outcomes are rare (Imai and Lo, 2021).[19] The results of this nonparametric sensitivity analysis are consistent with the findings presented here.

7.2 No Evidence of Backlash or Improved Human Rights Practices

Next, in Figure 3, I assess whether UPR impacts government human rights practices or the nature of domestic political institutions as expected by the alternative explanations. Across the board, I find no evidence that UPR affects these outcomes. UPR impacts neither aggregate measures of state respect for human rights nor specific indices measuring government use of torture or political killings. Likewise, I find no evidence that UPR causes states to adopt more open forms of democracy, such as increasing participatory democracy or the use of

[19] See, for example, Green, Kim and Yoon (2001); King (2001); Gartzke (2007); Dafoe (2011).

direct popular vote, or to create more space for civil society actors to participate in the political process.

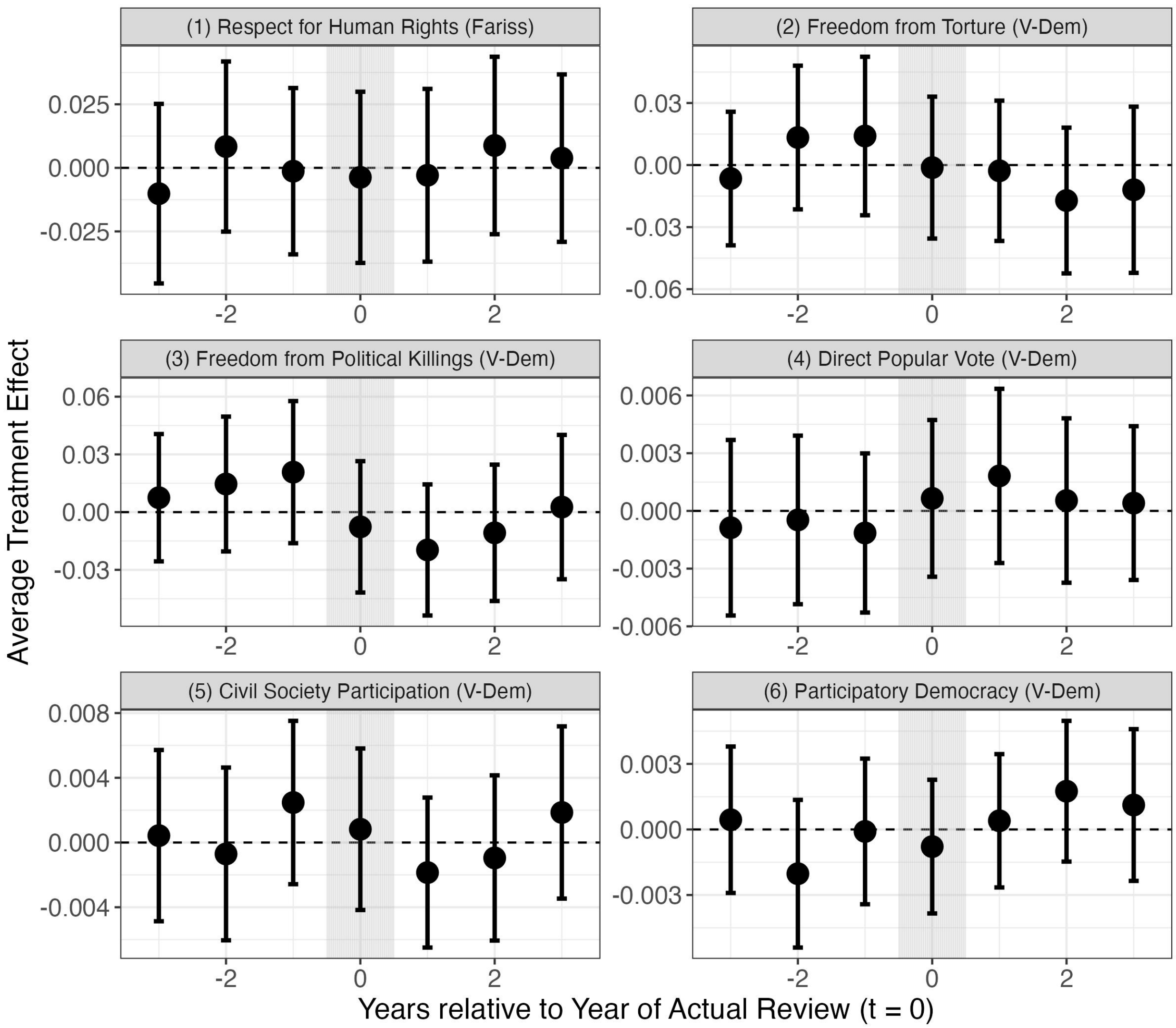

Figure 3: *Effect of UPR on Human Rights Practices and Political Institutions.* The impact of UPR on a selection of human rights practices and political institutions and outcomes – estimated by equation (1) – is presented in the shaded region at $t = 0$. Analyses in which the year of review is moved artificially 1, 2, or 3 years before or after the actual year of review are presented in the unshaded region. Standard errors are clustered at the country-year.

This analysis could be interpreted as consistent with a weaker version of the backlash hypothesis, in which states become recalcitrant and double-down on current practices in the face of external pressure. I contend, however, that when viewed in combination with the evidence presented in Figure 2, these findings are more consistent with a theory of conditional reactivity. Human rights abuse and political outcomes are the product of bureaucratic processes – rife with misgovernance and principal-agent problems – that are unlikely to be quickly affected by the type of short-term, exogenous change in social pressure produced by UPR. As a result, UPR does not affect these outcomes.

7.3 Legal Constraints Moderate Reactivity

I now examine how institutional constraints on executive authority moderate the effect of UPR.[20] I focus specifically on the role of judicial constraints, because the nature of the legal accountability measures under study – prosecutions, amnesties, and public inquiries in the form of truth commissions – all affect the jurisdiction or docket of the courts. I use the kernel estimator developed by Hainmueller, Mummolo and Xu (2019, 173-175) to estimate the effect of UPR on human rights accountability across different values of the V-Dem Judicial Constraints on the Executive Index. Higher values of this index indicate that an executive is more constrained by the judiciary.

This estimation strategy allows local effects to vary in a nonlinear manner. This would be necessary, for example, if outcomes in states with high and low levels of judicial constraints are not impacted by UPR, but states with intermediate levels of judicial constraints are. In these models, I include the same covariates and fixed-effects as in equation (1) and add terms for the moderator and the interaction of review and the moderator.[21]

The results of this analysis are presented in Figure 4. Each pane show how the effect of UPR on the implementation of a human rights accountability varies by judicial constraints on the executive; stacked histograms at the bottom of each pane demonstrate common support across the distribution of the moderator. Evidence in Figure 4 is broadly consistent with the theory of conditional reactivity. It suggests that the effect of UPR on opposition amnesties

[20] That is, I aim to understand how the effect of a single treatment varies across levels of a baseline covariate; what Keele and Stevenson (2021) call an "effect modification."

[21] The Hainmueller, Mummolo and Xu (2019) estimator estimates an optimal kernel bandwidth; however, these estimated bandwidths are unstable across analyses with different outcomes, ranging from 0.05 to 0.66. Consequentially, I use a fixed bandwidth of 0.1 to facilitate comparison across analyses also while striking a balance between stable and flexible estimation. See Appendix E for discussion.

and truth commissions is most pronounced in states where the executive faces few judicial constraints; there is no evidence of an effect in states with strong judicial constraints on the executive. Likewise, estimates indicate that UPR affects the initiation of domestic human rights trials of and amnesties for government officials in states with intermediate levels of judicial constraints. These governments typically impose weak constraints on executive authority, ranging from states like Saudi Arabia in 2013 (whose judicial constraints on the executive index is 0.264) and Kyrgyzstan in 2010 (0.319) through Morocco in 2017 (0.659) and Kenya in 2010 (0.762).

The findings in Figure 4 deviate from the theoretical expectations of conditional reactivity in one respect: it does not predict an increase in domestic trials in states with the most robust judicial institutions. This finding is consistent with research showing that the impact of international law on human rights outcomes is greatest in states with robust legal systems (Simmons, 2009). This finding also goes against research showing democracies are more likely to exhibit backlash to international naming and shaming campaigns (Hendrix and Wong, 2013). The evidence here does not supports this inference. Despite these deviations, the overall pattern that emerges from these analysis is most consistent with a theory of conditional reactivity.

I validate these findings in Appendix E. First, I present scatterplots and descriptive statistics of outcomes and the moderator by treatment status to assess functional form and common support across analyses. Second, I use randomization inference to compute the distribution of null results using the UPR lottery. These analyses are consistent with the results presented here when the lottery varies the assignment status of HRC; the impact of UPR on opposition amnesties and the initiation of domestic human rights trials is sensitive to holding fixed the treatment assignment of these 56 states. Third, I assess whether a state undergoing review impacts its judicial constraints on the executive index, which could introduce post-treatment bias; I find that it does not.

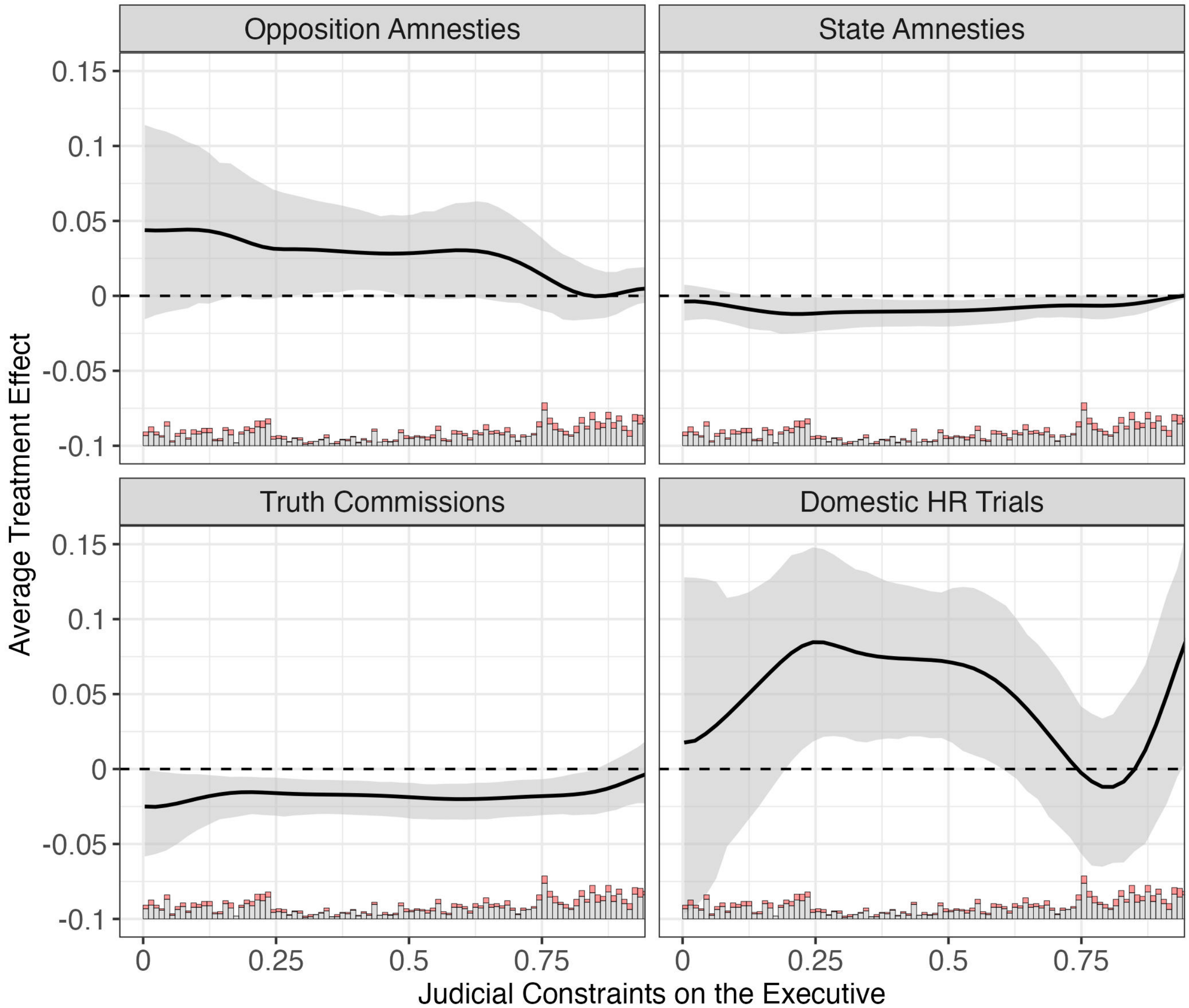

Figure 4: *Effect of UPR on Human Rights Accountability Varies by Judicial Constrains on Executive.* Solid lines are kernel estimates of the effect of review on a given human rights accountability measure by the level of judicial constrains on the executive. Bootstrapped country-year cluster-robust 95% confidence intervals are provided by the gray region. The margin of each plot provides a stacked histogram of the number of reviewed (red) and non-reviewed (gray) states at each value of the moderator.

8 Illustrative Evidence

The most direct evidence of UPR's impact on human rights reforms occurred when Israel's delegation announced during its presentation at the beginning of its second review that Israel would be releasing a group of Palestinian prisoners "that night."[22] The lead-up to this review

[22] The UN summarizes the Israeli presentation at UPR as follows: "[T]he [Israeli] delegation pointed out that Israel had agreed to release Palestinian prisoners as a confidence-building measure. A second group

was marked by intense social pressure (Cumming-Bruce, 2013). Israel's delegation had failed to appear for it's originally scheduled review on January 29, 2013. This was the first and only time a state had skipped a review, prompting a months-long push to secure Israeli participation.

Israel's release of Palestinian prisoners was a concession to this pressure. The US helped broker an initial release of prisoners in advance of Israel's review, which Israel characterized as a "confidence-building measure"; the timing and location of the announcement of Israeli's second amnesty suggest it was motivated, in part, to earn goodwill during its review at UPR. Israel had reason to anticipate criticism for its treatment of Palestinian prisoners: it was a frequent focus of HRC resolutions and the topic of several NGO reports submitted in advance of Israel's review (Human Rights Council, 2013*b*). During review, several states – including Libya, China, Qatar, Syria, Egypt, Pakistan, and Oman – commended Israel's announcement that it would release some prisoners that night, while still calling on it to release more.

Like Israel, other states highlighted opposition amnesties in their national reports and presentations at UPR. In its remarks before interactive dialogue in April 2014, Albania highlighted how a new amnesty law, implemented earlier that year, "had resulted in a significant reduction in prison overcrowding" (Human Rights Council, 2014). Mauritania similarly emphasized an amnesty policy in its presentation before UPR in 2010 (Human Rights Council, 2011). And, in 2009, Morocco praised Chad for granting amnesty before UPR while also encouraging it to implement recommendations of its truth commission (Human Rights Council, 2009). These examples indicate that states view amnesty policies as pro-active legal measures they can use to signal compliance with international norms at UPR.

Evidence also indicates that the social pressure generated by UPR can lead states to implement policies that they might otherwise forgo. For instance, the only recommendation Pakistan did not accepted in 2012 concerned oppression in Balochistan and impunity for

of prisoners was being released that night. Their release illustrated the determination of Israel to reach an agreement with its Palestinian neighbours that would, once and for all, end the conflict" (Human Rights Council, 2013*a*).

torture and disappearances nationwide, indicating its unique sensitivity of these specific issues.[23] These issues also arose during Pakistan's first review in 2008. Despite their sensitivity, Pakistan took some action on both fronts in the intervening period. Before its review in 2012 Pakistan issued an amnesty for Balochi rebels and rebel leaders, and after the topic of impunity was raised repeatedly in 2008, Pakistan initiated nine trials of government officials for human rights violations in 2010, a substantial increase given that the most recent trials were two in 2007 and one in 2000.

9 Reactivity and the Struggle for Human Rights

These analyses provide strong evidence that UPR causes states to adopt human rights accountability measures in a manner that is consistent with the theory of reactivity. UPR prompts states to issue more amnesties for opposition figures, fewer amnesties for government officials, and increases the likelihood that states will begin prosecuting their own officials for past crimes. However, UPR also causes states to create fewer truth commissions, indicating an effort to obfuscate evidence of wrongdoing.

This collection of findings raise broader questions about the efficacy of UPR in the struggle to advance human rights. Are states merely practicing gamesmanship, adopting accountability strategically to ward off international opprobrium? Or does UPR "work" by prompting states to implement measures they otherwise would not?

The theory of reactivity suggests these are not mutually exclusive explanations; policy change can follow from both strategic and sincere motives. Indeed, the empirical findings presented here align with both possibilities. In Figure 2, for example, states adopting accountability measures in the years preceding review, and forgoing them afterward, would be more consistent with gamesmanship, while effects persisting in the years after review occurs would be more indicative of lasting policy change. The evidence, however, does not point

[23] Pakistan did not accept the recommendation from the United States that it "[h]alt operations aimed at silencing dissent in Balochistan and ensure laws are fully equally enforced to investigate and prosecute those responsible for torture and enforced disappearances nationwide" (Human Rights Council, 2012).

decisively in either direction. Likewise, the qualitative evidence shows that UPR leads states to implement policies they otherwise would not, but Figure 4 shows the states are more impacted by UPR when they have weaker judicial institutions, precisely the contexts where legal accountability adopted to appease international pressure is often "window-dressing" (Levitsky and Murillo, 2009, 120).

The consequences of UPR for human rights practices, however, does not necessarily hinge on the intentions of those implementing reforms. There is a substantial body research showing how modest policy changes, rhetorical commitments, or symbolic acts cause lasting and often unintended policy shifts. Scholarship in both international relations and American politics has shown that new policies can "lock in" government action, making policy reversal politically and practically difficult (Moravcsik, 2000; Patashnik, 2009). Policy proclamations can tie governments to set courses of actions, raising the risks of hypocrisy and reputational costs (Schimmelfennig, 2001). As a result, even minor policy change can set governments on courses of action that are "almost impossible to reverse" (Pierson, 2000). This research suggests that the policy reforms caused by UPR, especially policies involving change, like the releasing of opposition figures and the prosecution of government officials for past abuse, are not so easy to unwind.

These findings have important implications for our understanding of the politics of UPR and the efficacy of international monitoring. As recent scholarship has consistently documented, the UN Human Rights Council is deeply politicized (Carraro, 2017; Terman and Voeten, 2018). At UPR, in particular, as Terman and Byun so eloquently write, "principled neutrality is compromised in favor of political discretion" (Terman and Byun, 2022, 385). But states are political actors, and no institutional arrangement is capable of getting them to check their politics at the door. The question, then, is what institutional arrangement work despite politicization? The evidence presented here indicates that the social processes unlocked by UPR have the power to affect government accountability practices even in contexts rampant politicization. And even if modest or incremental, any act of legal accountability is a step toward addressing past injustice and advancing human rights.

References

Adair, John G. 1984. "The Hawthorne Effect: A Reconsideration of the Methodological Artifact." *Journal of Applied Psychology* 69(2):334–345.

Annan, Kofi. 2005. "In larger freedom: towards development, security and human rights for all." A/59/2005/Add.1. New York, NY: United Nations.

Bailey, Michael, Anton Strezhnev and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61(2):430–456.

Bisbee, James H., James R. Hollyer, B. Peter Rosendorff and James Raymond Vreeland. 2019. "The Millennium Development Goals and Education: Accountability and Substitution in Global Assessment." *International Organization* 73(3):547–578.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival*. MIT Press.

Campbell, Donald T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54:297–312.

Carey, Sabine C., Neil J. Mitchell and Will Lowe. 2013. "States, the security sector, and the monopoly of violence: A new database on pro-government militias." *Journal of Peace Research* 50(2):249–258.

Carnegie, Allison and Austin Carson. 2018. "The Spotlight's Harsh Glare: Rethinking Publicity and International Order." *International Organization* 72(3):627–657.

Carraro, Valentina. 2017. "The United Nations Treaty Bodies and Universal Periodic Review: Advancing Human Rights by Preventing Politicization?" *Human Rights Quarterly* 39(4):943–970.

Carraro, Valentina. 2019. "Promoting Compliance with Human Rights: The Performance of the United Nations' Universal Periodic Review and Treaty Bodies." *International Studies Quarterly* 63(4):1079–1093.

Chaudoin, Stephen, Jude Hays and Raymond Hicks. 2018. "Do We Really Know the WTO Cures Cancer?" *British Journal of Political Science* 48(4):903–928.

Chayes, Abram and Antonia Handler Chayes. 1993. "On compliance." *International Organization* 47(2):175–205.

Conrad, Courtenay R. and Emily Hencken Ritter. 2019. *Contentious Compliance: Dissent and Repression under International Human Rights Law*. Oxford University Press.

Conrad, Courtenay R. and Will H. Moore. 2010. "What Stops the Torture?" *American Journal of Political Science* 54(2):459–476.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Cecilia Borella, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerlow, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Joshua Krusell, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Natalia Natsika, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Oskar Rydén, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Yi ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2024. "V-Dem Country-Year Dataset v14.". Varieties of Democracy (V-Dem) Project. https://doi.org/10.23696/mcwt-fr58.

Cortell, Andrew P. and James W. Davis. 1996. "How Do International Institutions Matter? The Domestic Impact of International Rules and Norms." *International Studies Quarterly* 40(4):451–478.

Creamer, Cosette D. and Beth A. Simmons. 2019. "Do Self-Reporting Regimes Matter? Evidence from the Convention Against Torture." *International Studies Quarterly* 63(4):1051–1064.

Cumming-Bruce, Nick. 2013. "Israel Skips U.N. Review on Rights, a New Move." *The New York Times*. January 29.

Dafoe, Allan. 2011. "Statistical Critiques of the Democratic Peace: Caveat Emptor." *American Journal of Political Science* 55(2):247–262.

Dai, Xinyuan. 2002. "Information Systems in Treaty Regimes." *World Politics* 54(4):405–436.

Dai, Xinyuan. 2007. *International Institutions and National Policies*. Cambridge University Press.

Dancy, Geoff and Christopher J. Fariss. 2024. "The Global Resonance of Human Rights: What Google Trends Can Tell Us." *American Political Science Review* 118(1):252–273.

Dancy, Geoff, Phuong Pham, Kathryn Sikkink, Oskar Timo Thoms and Patrick Vinck. 2024. "Introducing the Transitional Justice Evaluation Tools (TJET) Database.". Working Paper.

Davenport, Christian. 2007*a*. "State Repression and Political Order." *Annual Review of Political Science* 10(1):1–23.

Davenport, Christian. 2007*b*. *State Repression and the Domestic Democratic Peace*. Cambridge University Press.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110(9):2964–2996.

Espeland, Wendy and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1):1–40.

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., Therese Anders, Jonathan N. Markowitz and Miriam Barnum. 2022. "New Estimates of Over 500 Years of Historic GDP and Population Data." *Journal of Conflict Resolution* 66(3):553–591.

Fariss, Christopher, Michael Kenwick and Kevin Reuning. 2020. "Latent Human Rights Protection Scores Version 4." Harvard Dataverse. DOI: 10.7910/DVN/RQ85GK.

Finnemore, Martha. 1993. "International organizations as teachers of norms: the United Nations Educational, Scientific, and Cultural Organization and science policy." *International Organization* 47(4):565–597.

Finnemore, Martha and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization* 52(4):887–917.

Franklin, James C. 2008. "Shame on You: The Impact of Human Rights Criticism on Political Repression in Latin America." *International Studies Quarterly* 52(1):187–211.

Gartzke, Erik. 2007. "The Capitalist Peace." *American Journal of Political Science* 51(1):166–191.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton.

Green, Donald P., Soo Yeon Kim and David H. Yoon. 2001. “Dirty Pool.” *International Organization* 55(2):441–468.

Greenhill, Brian. 2010. “The Company You Keep: International Socialization and the Diffusion of Human Rights Norms.” *International Studies Quarterly* 54(1):127–145.

Greenhill, Brian. 2015. *Transmitting Rights: International Organizations and the Diffusion of Human Rights Practices*. Oxford University Press.

Gruffydd-Jones, Jamie J. 2021. “International Attention and the Treatment of Political Prisoners.” *International Studies Quarterly* 65(4):999–1011.

Guess, Andrew and Alexander Coppock. 2020. “Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments.” *British Journal of Political Science* 50(4):1497–1515.

Gurowitz, Amy. 1999. “Mobilizing International Norms: Domestic Actors, Immigrants, and the Japanese State.” *World Politics* 51(3):413–445.

Guzman, Andrew. 2008. *How International Law Works: A Rational Choice Theory*. New York, NY: Oxford University Press.

Haas, Peter M. 1992. “Epistemic Communities and International Policy Coordination.” *International Organization* 46(1):1–35.

Hafner-Burton, Emilie M. 2008. “Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem.” *International Organization* 62(4):689–716.

Hafner-Burton, Emilie M. 2013. *Making Human Rights a Reality*. Princeton, NJ: Princeton University Press.

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. and Kiyoteru Tsutsui. 2005. “Human Rights in a Globalizing World: The Paradox of Empty Promises.” *American Journal of Sociology* 110(5):1373–1411.

Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. “How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.” *Political Analysis* 27(2):163–192.

Hartman, Erin and F. Daniel Hidalgo. 2018. "An Equivalence Approach to Balance and Placebo Tests." *American Journal of Political Science* 62(4):1000–1013.

Hendrix, Cullen S. and Wendy H. Wong. 2013. "When Is the Pen Truly Mighty? Regime Type and the Efficacy of Naming and Shaming in Curbing Human Rights Abuses." *British Journal of Political Science* 43(3):651–672.

Hill, Daniel W, Jr. and Zachary M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(3):661–687.

Honig, Dan. 2019. "When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation." *International Organization* 73(1):171–201.

Human Rights Council. 2009. "Report of the Working Group on the Universal Periodic Review: Chad." Twelfth session, Agenda item 6, Universal Periodic Review, A/HRC/12/5*, Geneva. October 5.

Human Rights Council. 2011. "Report of the Working Group on the Universal Periodic Review: Mauritania." Sixteenth session, Agenda item, Universal Periodic Review, A/HRC/16/17, Geneva. January 4.

Human Rights Council. 2012. "Report of the Working Group on the Universal Periodic Review: Pakistan." Twenty-second session, Agenda item, Universal Periodic Review, A/HRC/22/12, Geneva. December 26.

Human Rights Council. 2013*a*. "Report of the Working Group on the Universal Periodic Review: Israel." Twenty-fifth session, Agenda item 6, Universal Periodic Review, A/HRC/25/15, Geneva. December 19.

Human Rights Council. 2013*b*. "Summary prepared by the Office of the High Commissioner for Human Rights in accordance with paragraph 5 of the annex to Human Rights Council resolution 16/21." Working Group on the Universal Periodic Review, A/HRC/WG.6/15/ISR/3, Fifteenth session, Geneva. January 21 – February 1.

Human Rights Council. 2014. "Report of the Working Group on the Universal Periodic Review: Albania." Twenty-seventh session, Agenda item 6, Universal Periodic Review, A/HRC/27/4, Geneva. July 7.

Imai, Kosuke and In Song Kim. 2021. "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29(3):405–415.

Imai, Kosuke and James Lo. 2021. "Robustness of Empirical Evidence for the Democratic Peace: A Nonparametric Sensitivity Analysis." *International Organization* 75(3):901–919.

James, Ron. 2002. Territoriality and Plausible Deniability: Serbian Paramilitaries in the Bosnian War. In *Death Squads in Global Perspective: Murder with Deniability*, ed. Bruce B. Campbell and Arthur D. Brenner. Palgrave McMillan pp. 286–312.

Johnston, Alastair Iain. 2001. "Treating international institutions as social environments." *International Studies Quarterly* 45(1):487–515.

Keck, Margaret E. and Kathryn Sikkink. 1998. *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.

Keele, Luke and Randolph T. Stevenson. 2021. "Causal interaction and effect modification: same model, different concepts." *Political Science Research and Methods* 9(3):641–649.

Kelley, Judith and Beth Simmons. 2015. "Politics by Number: Indicators as Social Pressure in International Relations." *American Journal of Political Science* 59(1):55–70.

Kelley, Judith G. 2017. *Scorecard Diplomacy: Grading States to Influence their Reputation and Behavior*. Cambridge University Press.

Kelley, Judith G. and Beth A. Simmons. 2019. "Introduction: The Power of Global Performance Indicators." *International Organization* 73(3):491–510.

Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.

King, Gary. 2001. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data." *International Organization* 55(2):497–507.

Koremenos, Barbara. 2016. *The Continent of International Law: Explaining Agreement Design*. New York, NY: Cambridge University Press.

Koremenos, Barbara, Charles Lipson and Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(4):761–799.

Lebovic, James H. and Erik Voeten. 2006. "The Politics of Shame: The Condemnation of Country Human Rights Practices in the UNCHR." *International Studies Quarterly* 50(4):861–888.

Levitsky, Steven and María Victoria Murillo. 2009. "Variation in Institutional Strength." *Annual Review of Political Science* 12:115–133.

Lupu, Yonatan. 2013*a*. "Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements." *International Organization* 67(3):469–503.

Lupu, Yonatan. 2013*b*. "The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects." *American Journal of Political Science* 57(4):912–925.

Lupu, Yonatan. 2015. "Legislative Veto Players and the Effects of International Human Rights Agreements." *American Journal of Political Science* 59(3):578–594.

McCall, George J. 1984. "Systematic Field Observation." *Annual Review of Sociology* 10:263–282.

Mitchell, Neil. 2004. *Agents of Atrocity: Leaders , Followers , and the Violation of Human Rights in Civil War*. Palgrave MacMillan.

Mitchell, Sara McLaughlin, Jonathan J. Ring and Mary K. Spellman. 2013. "Domestic Legal Traditions and States' Human Rights Practices." *Journal of Peace Research* 50(2):189–202.

Moravcsik, Andrew. 2000. "The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe." *International Organization* 54(2):217–252.

Morse, Julia C. and Tyler Pratt. 2022. "Strategies of Contestation: International Law, Domestic Audiences, and Image Management." *The Journal of Politics* 84(4):2080–2093.

Murdie, Amanda M. and David R. Davis. 2012. "Shaming and Blaming: Using Events Data to Assess the Impact of Human Rights INGOs." *International Studies Quarterly* 56(1):1–16.

Nutt, Cullen G. and Reid B.C. Pauly. 2021. "Caught Red-Handed: How States Wield Proof to Coerce Wrongdoers." *International Security* 46(2):7–50.

Patashnik, Eric M. 2009. *Reforms at Risk: What Happens After Major Policy Changes Are Enacted*. Princeton University Press.

Pierson, Paul. 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94(2):251–267.

Poe, Steven C. and C. Neal Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88(4):853–872.

Powell, Emilia J. and Jeffrey K. Staton. 2009. "Domestic Judicial Institutions and Human Rights Treaty Violation." *International Studies Quarterly* 53(1):149–174.

Rejali, Darius. 2007. *Torture and Democracy.* Princeton University Press.

Risse, Thomas, Stephen C. Ropp and Kathryn Sikkink, eds. 2013. *The Persistent Power of Human Rights: From Commitment to Compliance.* Cambridge University Press.

Ritter, Emily Hencken and Courtenay R. Conrad. 2016. "Human rights treaties and mobilized dissent against the state." *The Review of International Organizations* 11(4):449–475.

Schimmelfennig, Frank. 2001. "The Community Trap: Liberal Norms, Rhetorical Action, and the Eastern Enlargement of the European Union." *International Organization* 55(1):47–80.

Sikkink, Kathryn. 2011. *The Justice Cascade: How Human Rights Prosecutions Are Changing World Politics.* New York, NY: WW. Norton & Company.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics* . New York, NY: Cambridge University Press.

Simmons, Beth A. and Daniel J. Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99(4):623–631.

Slough, Tara and Christopher Fariss. 2021. "Misgovernance and Human Rights: The Case of Illegal Detention without Intent." *American Journal of Political Science* 65(1):148–165.

Snyder, Jack. 2020. "Backlash against human rights shaming: emotions in groups." *International Theory* 12(1):109–132.

Strezhnev, Anton, Judith G. Kelley and Beth A. Simmons. 2021. "Testing for Negative Spillovers: Is Promoting Human Rights Really Part of the "Problem"?" *International Organization* 75(1):71–102.

Terman, Rochelle. 2023. *The Geopolitics of Shaming: When Human Rights Pressure Works—and when it Backfires.* Princeton University Press.

Terman, Rochelle and Erik Voeten. 2018. "The relational politics of shame: Evidence from the universal periodic review." *The Review of International Organizations* 13(1):1–23.

Terman, Rochelle and Joshua Byun. 2022. "Punishment and Politicization in the International Human Rights Regime." *American Political Science Review* 116(2):385–402.

United Nations General Assembly resolution 60/251. 2006. "Human Rights Council." A/RES/60/251. New York, NY: United Nations.

Valentino, Benjamin A. 2005. *Final Solutions: Mass Killing and Genocide in the 20th Century*. Cornell University Press.

Velasco, Kristopher. 2023. "Transnational Backlash and the Deinstitutionalization of Liberal Norms: LGBT+ Rights in a Contested World." *American Journal of Sociology* 128(5):1381–1429.

Vinjamuri, Leslie. 2017. Human Rights Backlash. In *Human Rights Futures*, ed. Stephen Hopgood, Jack Snyder and Leslie Vinjamuri. Cambridge University Press pp. 114–134.

von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *The American Political Science Review* 99(4):611–622.

Zvobgo, Kelebogile. 2020. "Demanding Truth: The Global Transitional Justice Network and the Creation of Truth Commissions." *International Studies Quarterly* 64(3):609–625.

Appendix

A Summary Statistics

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Amnesty of Opposition Figure	2,259	0.029	0.167	0	1
Amnesty of State Official	2,259	0.010	0.098	0	1
Amnesty of State Official for HR Violation	2,259	0.005	0.073	0	1
Truth Commission Onset	2,272	0.013	0.112	0	1
Trial Onset	2,258	0.145	0.352	0	1
Respect for Human Rights	2,257	0.694	1.582	−2.676	5.496
Freedom from Torture	2,272	0.916	1.492	−3.150	3.509
Freedom from Political Killings	2,272	1.277	1.468	−2.973	3.504
Direct Popular Vote	2,259	0.105	0.125	0.000	0.747
Civil Society Participation	2,259	0.668	0.241	0.022	0.988
Participatory Democracy	2,259	0.492	0.187	0.019	0.882
Review	2,272	0.207	0.405	0	1
Judicial Constraints on the Executive	2,259	0.582	0.312	0.003	0.991
Liberal Democracy	2,272	0.532	0.256	0.015	0.919
Log GDP Per Capita	2,085	3.010	1.163	0.319	5.396
Log Population	2,085	6.918	1.678	2.238	11.907
UNGA Ideal Point	2,232	−0.155	0.846	−2.147	2.767
Political Stability	2,232	−0.195	0.965	−3.315	1.656
Log Land Area	2,060	11.891	2.143	3.551	16.639
HRC Member	2,233	0.272	0.445	0	1
Africa	2,233	0.313	0.464	0	1
Asia	2,233	0.262	0.440	0	1
Latin America	2,233	0.146	0.353	0	1
W. Europe and Other States	2,233	0.146	0.353	0	1
COW Country Code	2,272	468.037	236.057	2	950
Year	2,272	2,014.007	3.740	2,008	2,020

B UPR Lottery Details

Figure 5 provides the distribution of the number of states from each regional group assigned to each session during the first cycle. This is the schedule used for block randomization by region during the UPR lottery.

Session/ year	African Group	Asian Group	GRULAC	WEOG	EEG	Total
1-2008	4	4	3	3	2	16
2-2008	5	5	2	2	2	16
3-2008	4	4	3	3	2	16
4-2009	5	5	2	2	2	16
5-2009	4	4	3	3	2	16
6-2009	5	5	3	2	1	16
7-2010	4	5	3	2	2	16
8-2010	4	4	3	3	2	16
9-2010	4	5	3	2	2	16
10-2011	5	4	3	2	2	16
11-2011	4	5	2	3	2	16
12-2011	5	4	3	2	2	16
Total	53	54	33	29	23	

Figure 5: *Regional Distribution of States per Session.* Table provides the number of states that were required to be from each region when drawing lots (A/HRC/6/22, page 209).

Regional Groupings

The states categorized in each regional grouping are listed below. Bolded states are those that were serving or had been elected to serve on the HRC in 2007 when the lottery was conducted. The year these states were required to undergo their first review is noted in parentheses.

African Group (53 States)

- **Algeria (2007)**; **Angola (2010)**; Benin; Botswana; Burkina Faso; Burundi; **Cameroon (2009)**; Cape Verde; Central African Republic; Chad; Comoros; Congo; Côte d'Ivoire; Democratic Republic of the Congo; **Djibouti (2008)**; **Egypt (2010)**; Equatorial Guinea; Eritrea; Ethiopia; **Gabon (2008)**; Gambia; **Ghana (2008)**; Guinea; Guinea-Bissau; Kenya; Lesotho; Liberia; Libyan Arab Jamahiriya; **Madagascar (2010)**; Malawi; **Mali (2008)**; Mauritania; **Mauritius (2009)**; **Morocco (2007)**; Mozambique; Namibia; Niger; **Nigeria (2009)**; Rwanda; Sao Tome and Principe; **Senegal (2009)**; Seychelles;

Sierra Leone; Somalia; **South Africa (2007)**; Sudan; Swaziland; Togo; **Tunisia (2007)**; Uganda; United Republic of Tanzania; **Zambia (2008)**; Zimbabwe.

Asian Group (54 States)

- Afghanistan; **Bahrain (2007)**; **Bangladesh (2009)**; Bhutan; Brunei Darussalam; Cambodia; **China (2009)**; Cyprus; Democratic People's Republic of Korea; Fiji; **India (2007)**; **Indonesia (2007)**; Iran (Islamic Republic of); Iraq; **Japan (2008)**; **Jordan (2009)**; Kazakhstan; Kiribati; Kuwait; Kyrgyzstan; Lao People's Democratic Republic; Lebanon; **Malaysia (2009)**; Maldives; Marshall Islands; Micronesia (Federated States of); Mongolia; Myanmar; Nauru; Nepal; Oman; **Pakistan (2008)**; Palau; Papua New Guinea; **Philippines (2007)**; **Qatar (2010)**; **Republic of Korea (2008)**; Samoa; **Saudi Arabia (2009)**; Singapore; Solomon Islands; **Sri Lanka (2008)**; Syrian Arab Republic; Tajikistan; Thailand; Timor Leste; Tonga; Turkmenistan; Tuvalu; United Arab Emirates; Uzbekistan; Vanuatu; Viet Nam; Yemen.

Latin American and Caribbean States (GRULAC) (33 States)

- Antigua and Barbuda; **Argentina (2007)**; Bahamas; Barbados; Belize; **Bolivia (2010)**; **Brazil (2008)**; Chile; Colombia; Costa Rica; **Cuba (2009)**; Dominica; Dominican Republic; **Ecuador (2007)**; El Salvador; Grenada; **Guatemala (2008)**; Guyana; Haiti; Honduras; Jamaica; **Mexico (2009)**; **Nicaragua (2010)**; Panama; Paraguay; **Peru (2008)**; Saint Kitts and Nevis; Saint Lucia; Saint Vincent and the Grenadines; Suriname; Trinidad and Tobago; **Uruguay (2009)**; Venezuela (Bolivarian Republic of).

Western Europe and Other States (WEOG) (29 States)

- Andorra; Australia; Austria; Belgium; **Canada (2009)**; Denmark; **Finland (2007)**; **France (2008)**; **Germany (2009)**; Greece; Iceland; Ireland; Israel; **Italy (2010)**; Liechtenstein; Luxembourg; Malta; Monaco; **Netherlands (2007)**; New Zealand; Norway; Portugal; San Marino; Spain; Sweden; **Switzerland (2009)**; Turkey; **United Kingdom of Great Britain and Northern Ireland (2008)**; United States of America.

Eastern European States (EEG) (23 States)

- Albania; Armenia; **Azerbaijan (2009)**; Belarus; **Bosnia and Herzegovina (2010)**; Bulgaria; Croatia; **Czech Republic (2007)**; Estonia; Georgia; Hungary; Latvia; Lithuania; Moldova; Montenegro; **Poland (2007)**; **Romania (2008)**; **Russian Federation (2009)**; Serbia; Slovakia; **Slovenia (2010)**; The former Yugoslav Republic of Macedonia; **Ukraine (2008)**.

C Impact of UPR on Salience of Human Rights

Figure 6 presents the results of an analysis assessing how UPR impacts Google searches for human rights (that is, variation in the Google Trends "human rights" topic) within countries in the months before, during, and after a country undergoes UPR during the 2008-2020 time period. Formally the analysis is,

$$Y_{i,m} = \alpha_i + \gamma_m + \beta D_{i,m} + \varepsilon_{i,m}$$

where $Y_{i,m}$ is the value of the Google Trends "human rights" topic for country i in month m, α_i and γ_m are country and month fixed effects, $D_{i,m}$ is an indicator that takes 1 if country i under went review in month m, and $\varepsilon_{i,m}$ is an error term. The effect of review on human rights searches is estimated by $\hat{\beta}$. I present the effect of review in the shaded region at $t = 0$ as well as placebo estimates in which I move the month of review earlier or later by -12 to 12 months. Standard errors are clustered at the country-month. This analysis suggests that UPR increases domestic Google searches for human rights related topics, which is consistent with the claim that UPR exerts social pressure via a domestic politics mechanism.

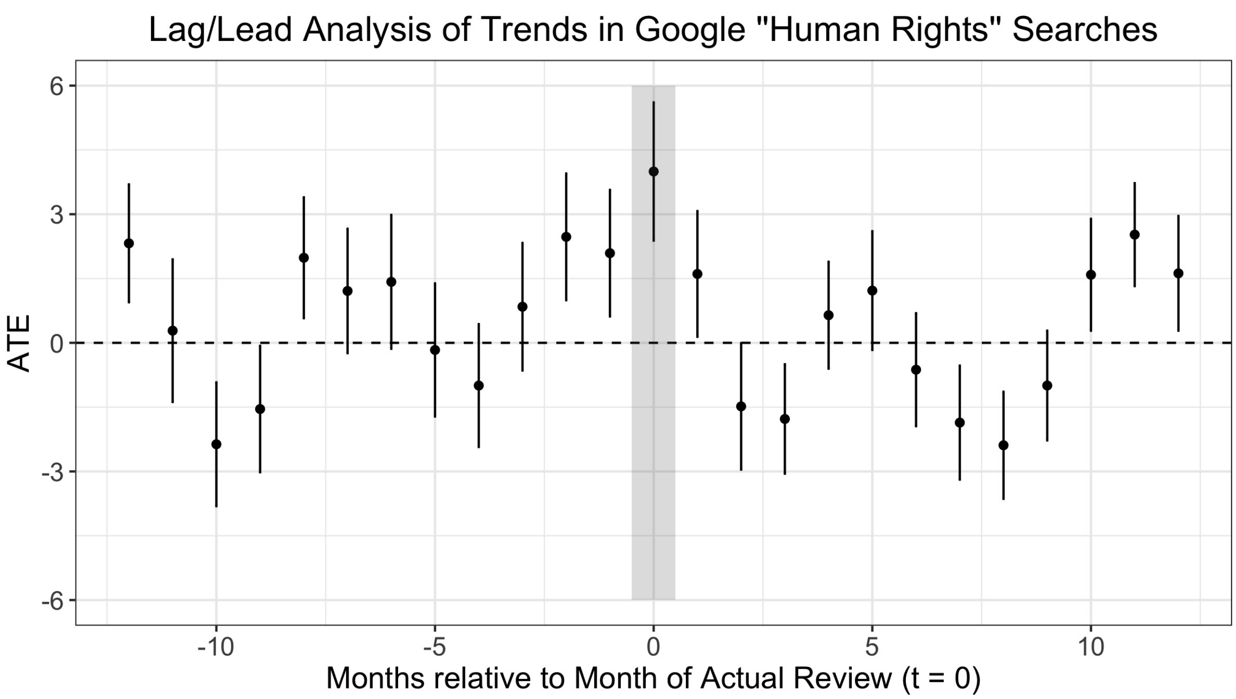

Figure 6: *Effect of UPR on Google Searches of Human Rights Topic.*

D Randomization Inference

I conduct randomization inference in two ways. First, I conduct 10,000 placebo lotteries following precisely the steps used by the HRC to implement the lottery on September 21, 2007. Critically, this approach holds fixed the treatment assignment of the 56 states that were either serving on the HRC when the lottery was conducted in 2007 or that had been elected to begin serving on the HRC in 2008.[24] This is a conservative test, because in all permutations these 56 states receive their observed treatment assignment. Consequently, this approach does not produce a distribution centered around zero, but instead one in which some effect is observed. Second, I conduct the same number of placebo lotteries, but ignore the rule that states serving or elected to serve as HRC members in 2007 must be reviewed in the final year of their term; all other rules (e.g., regional distributions, preferences of Switzerland and Colombia, and the disruption of Haiti's earthquake) are followed. This approach permits the treatment assignment of these 56 states to vary.

I present my results as follows. First, in Figure 7, I assess the comparability of covariate balance in the observed lottery to the 10,000 placebo lotteries used to conduct randomization inference. The placebo lotteries produce treatment and control groups that are very similar to those observed in the true lottery. This suggests that the outcome of the observed lottery is not an outlier. The t-statistic for each covariate in the observed lottery (represented by the red dashed vertical lines) is between the 2.5 and 97.5 percentiles of the distribution of t-statistics computed for each placebo lottery (represented by the solid green and orange vertical lines) with the sole exception of Log Population, where the lottery imposing the HRC rule tends to produce more a more populous treatment group than the lottery that omits this rule.[25] This difference suggests that larger, more powerful states were more likely to be serving or elected to serve on the HRC when it was first created and when the lottery was conducted. I address this difference by controlling for population and GDP per capita in my regression-based analyses and by assessing the robustness of my results using non-parametric sensitivity analysis described below.

Second, in Figure 8, I use randomization inference to assess the robustness of the results presented in Figure 2. Models (2), (4), (5), and (6) present the same analyses as those

[24] These states' names and the years of their initial reviews are bolded in the regional lists above.

[25] Formally, for each variable Y, the t-statistic is,

$$T^{\text{t-stat}} = \frac{\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}}{\sqrt{s_t^2/N_t + s_c^2/N_c}} \tag{2}$$

where s_T^2 and s_C^2 are the variances of Y for states assigned to treatment and control, $\bar{Y}_t^{\text{obs}}$ and $\bar{Y}_c^{\text{obs}}$ are the averages of Y for states assigned to treatment and control, and N_t and N_c are the number of states assigned to treatment and control.

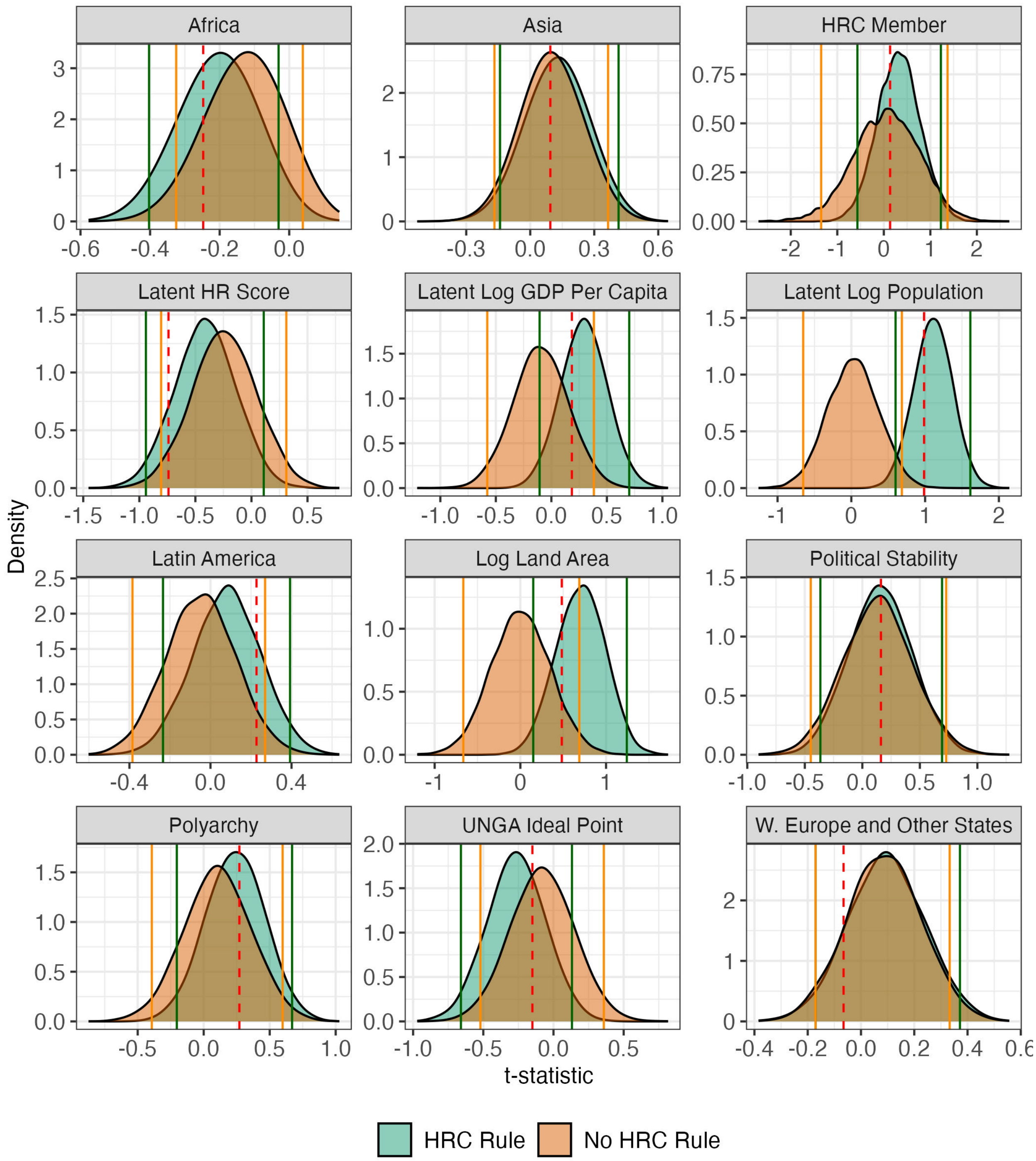

Figure 7: *Covariate Balance Across Permutations.* Each facet plots the density of t-statistics of differences between reviewed and non-reviewed states for given covariate across 10,000 placebo lotteries. Distributions of t-statistics computed from lotteries that hold fixed HRC members' treatment assignments are green; distributions computed from lotteries that ignore this rule are orange. Vertical solid lines mark the 2.5 and 97.5 percentiles of each distribution. Red dashed vertical lines mark the observed t-statistics from the true lottery.

provided at $t = 0$ in Figure 2; model (1) excludes covariates to assess model dependence and model (3) assess whether UPR impacts all amnesties issued governments for their own officials, rather than just the set of amnesties targeting past human rights violations. The distribution of null estimates computed by replicating the UPR lottery 10,000 times are represented by the two densities. The green density impose the rule that the 56 states serving or elected to the HRC in 2007 underwent UPR first, which holds these states' treatment assignments fixed; the orange densities ignore this rule. In each plot I report the empirical cumulative distribution function (eCDF), $F(t)$, of the observed estimate in the distribution of placebo estimates.[26] By conventional standards, values of $F(t)$ at the 95% (90%) significance level are less than 0.025 (0.05) or greater than 0.975 (0.95).

Next, in Figure 9, I extend the randomization inference to estimates in which I artificially move earlier or later states assigned review dates. Again, the only estimates that falls outside the 0.025 and 0.975 quantiles of the permuted estimates are those of the observed actual treatment presented in the shaded regions of each pane. The analyses in Figures 8 and 9 thus validate the results presented in Figure 2.

Third, in Figure 10, I use nonparametric statistics to evaluate the robustness of the results presented in Figure 2. Specifically, I compare the true observed relative risk to the observed relative risk computed across the set of placebo lotteries.[27] Again, I find strong evidence that states respond in a manner consistent with the concept of reactivity. Consistent with the regression-based results, these analyses show that states under review are twice as likely to issue opposition amnesties and less than half as likely to create new truth commissions. Estimates of the effect of UPR on the initiation of domestic human rights prosecutions is slightly more robust in this analysis than in Figure 8.

[26] Formally, the eCDF is estimated as

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq t)$$

where t is the quantity of interest computed from the observed data, x_i are test statistics computed from each of the $n = 10,000$ permutations of the treatment assignment, and $I()$ is an indicator function that takes 1 if $x_i \leq t$ and is 0 otherwise.

[27] Formally, observed relative risk is,

$$RR^{Obs}_{XD} = \frac{P(Y=1|D=1)}{P(Y=1|D=0)} = \frac{\frac{1}{n_t}\sum_{i=1}^{n_t} Y_{i\in t}}{\frac{1}{n_c}\sum_{i=1}^{n_c} Y_{i\in c}}$$

where n_t and n_c are the number of states in the treatment and control conditions, and $Y_{i\in t}$ and $Y_{i\in c}$ are the outcomes for states in these conditions.

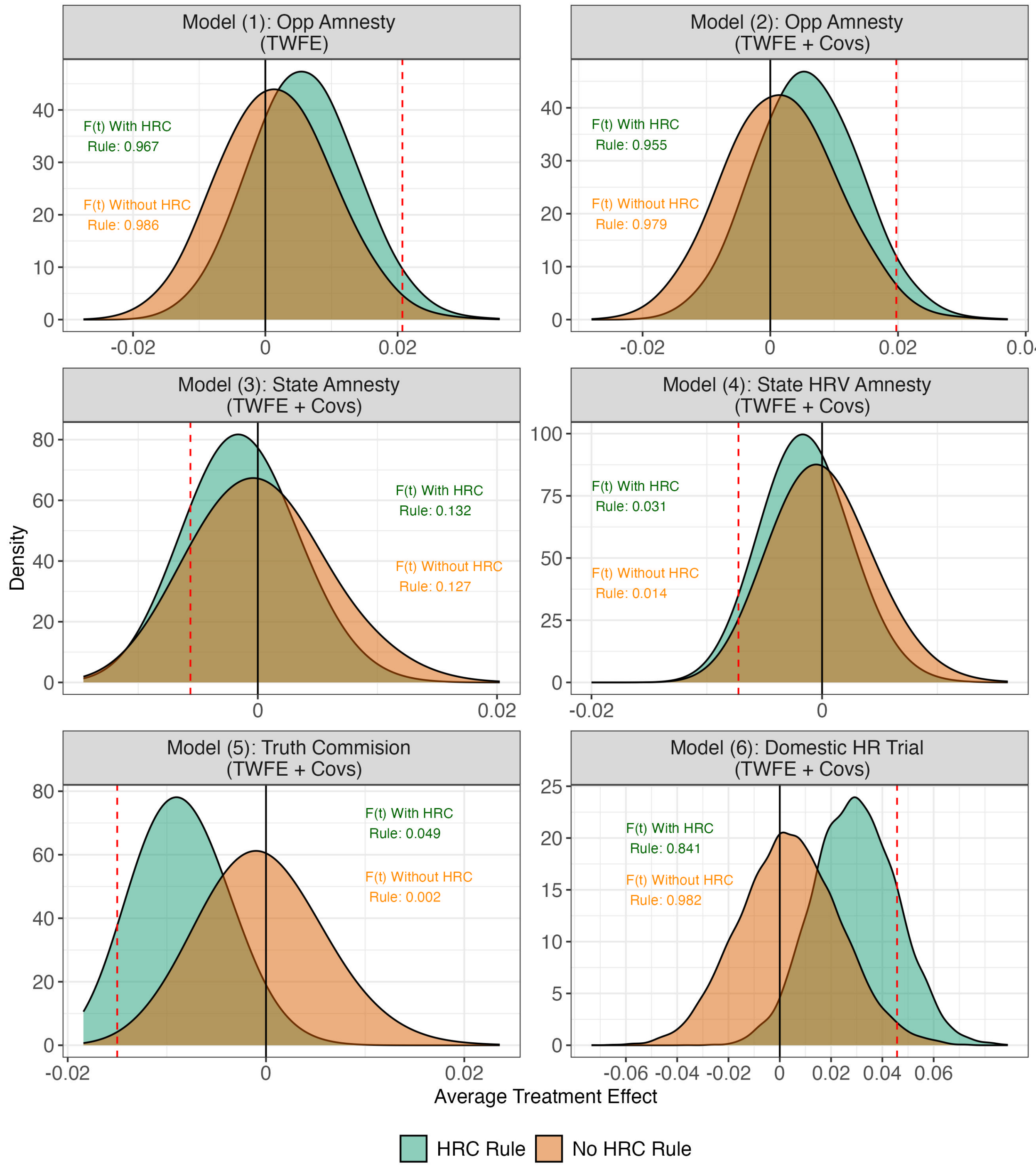

Figure 8: *Randomization Inference.* Each facet plots the distribution of 10,000 placebo ATEs ($\hat{\tau}$) computed by equation (1). As in Figure 7, distributions of ATEs computed from lotteries that hold fixed HRC members' treatment assignments are green; distributions computed from lotteries that ignore this rule are orange. Red dashed vertical lines mark the observed ATE from the true lottery. The eCDF of the observed ATE in the distribution of placebo lotteries is noted in each plot.

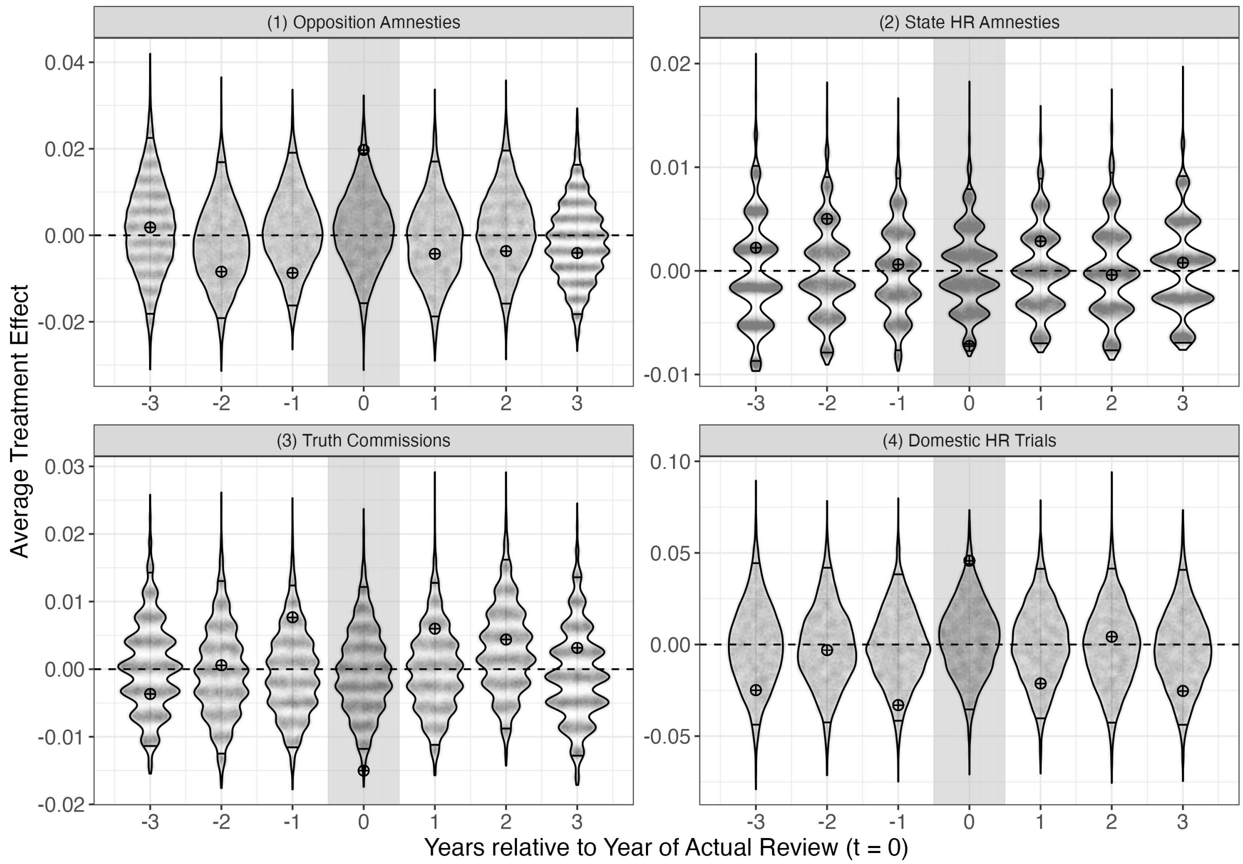

Figure 9: *Randomization Inference with Lag and Lead of Review.* Each facet plots the distribution of 10,000 placebo ATEs ($\hat{\tau}$) computed by equation (1). Circles with crosshairs are the point estimates reported in Figure 2. The distribution of placebo estimates is represented by the violins and scatterplots. Black lines at the top and bottom of each distribution represent 0.025 and 0.975 quantiles in the distribution of placebo estimates.

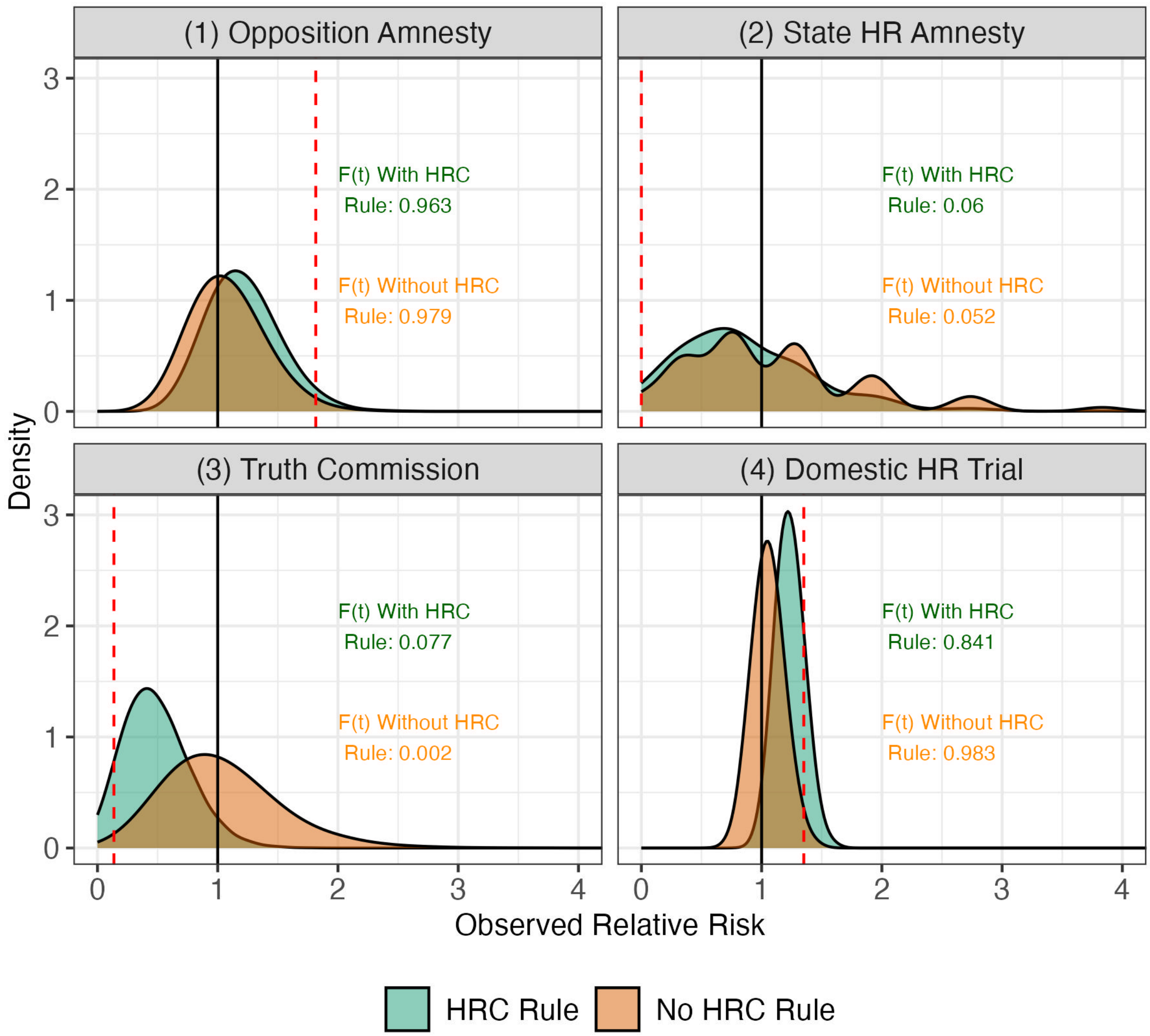

Figure 10: *Nonparametric Sensitivity Analysis.* Facets plot the density of 10,000 placebo estimates of observed relative risk. Distributions of estimates from both lotteries are presented and the true observed test statistic is marked by red dashed vertical lines. The eCDF of the observed relative risk in the distribution of placebo lotteries is noted in each plot.

E Multiplicative Interaction Models

Here I show that the effect of UPR is greatest among states in with executives that are less constrained by the judiciary. To do so, I estimate a series of two-way fixed-effect multiplicative interaction models of the form:

$$Y_{i,t} = \alpha_i + \gamma_t + \delta D_{i,t} + \eta X_{i,t} + \beta(D_{i,t} \times X_{i,t}) + \epsilon_{i,t}$$

where $Y_{i,t}$ is, as before, the relevant outcome for state i in year t, $D_{i,t}$ is an indicator that takes 1 if state i underwent review in year t, α_i and γ_t are state and year fixed effects, and $X_{i,t}$, the moderator, is the V-Dem Judicial Constraints on the Executive Index for country i in year t.

To conduct this analysis, I follow the best-practices recommended by Hainmueller, Mummolo and Xu (2019). First, in Figures 11 and 12, I present scatterplots of the value of the outcome by the moderator for each treatment group along with LOESS (red) and linear regression lines (blue). The overlap of the LOESS and OLS lines is evidence consistent with the linear interaction effect assumption, suggesting that a simple linear interaction model is well-suited to the data and not likely prone to model dependence (Hainmueller, Mummolo and Xu, 2019, 165). In the case of opposition amnesties and domestic human rights prosecution, there is evidence of a non-linear treatment effect, suggesting that this assumption may be violated. Therefore, I estimate main results using a semi-paramentric kernel weighting estimator. These plots also provide preliminary descriptive evidence of an interaction effect. We observe greater variance in outcomes among more states with more executive independence from the judiciary. Finally, these plots demonstrate common support in the values of the moderator between these groups, which is to be expected given the random assignment of treatment.

Second, in Figure 13, I present four interaction analyses using randomization inference to construct null distributions. Figure 14 replicates this analysis using the lottery that restricts the treatment assignment status of the 56 states serving on or elected to the HRC in 2007. The black line in these figures is the same estimate presented in Figure 4; however, instead of estimating uncertainty with a bootstrapped 95% confidence interval, I plot 10,000 null estimates computed from replications of the lottery. Dashed lines in Figures 13 and 14 mark the 95th and 90th percentiles of these null distributions for each value of the moderator.

Finally, in Table 2, I demonstrate that review has no effect on the V-Dem Judicial Constraints on the Executive Index, which would suggest post-treatment bias. In these analyses, I estimate a series of linear regression models: Model (1) is a simple bivariate regression of the index on the treatment variable; Model (2) adds country and year fixed

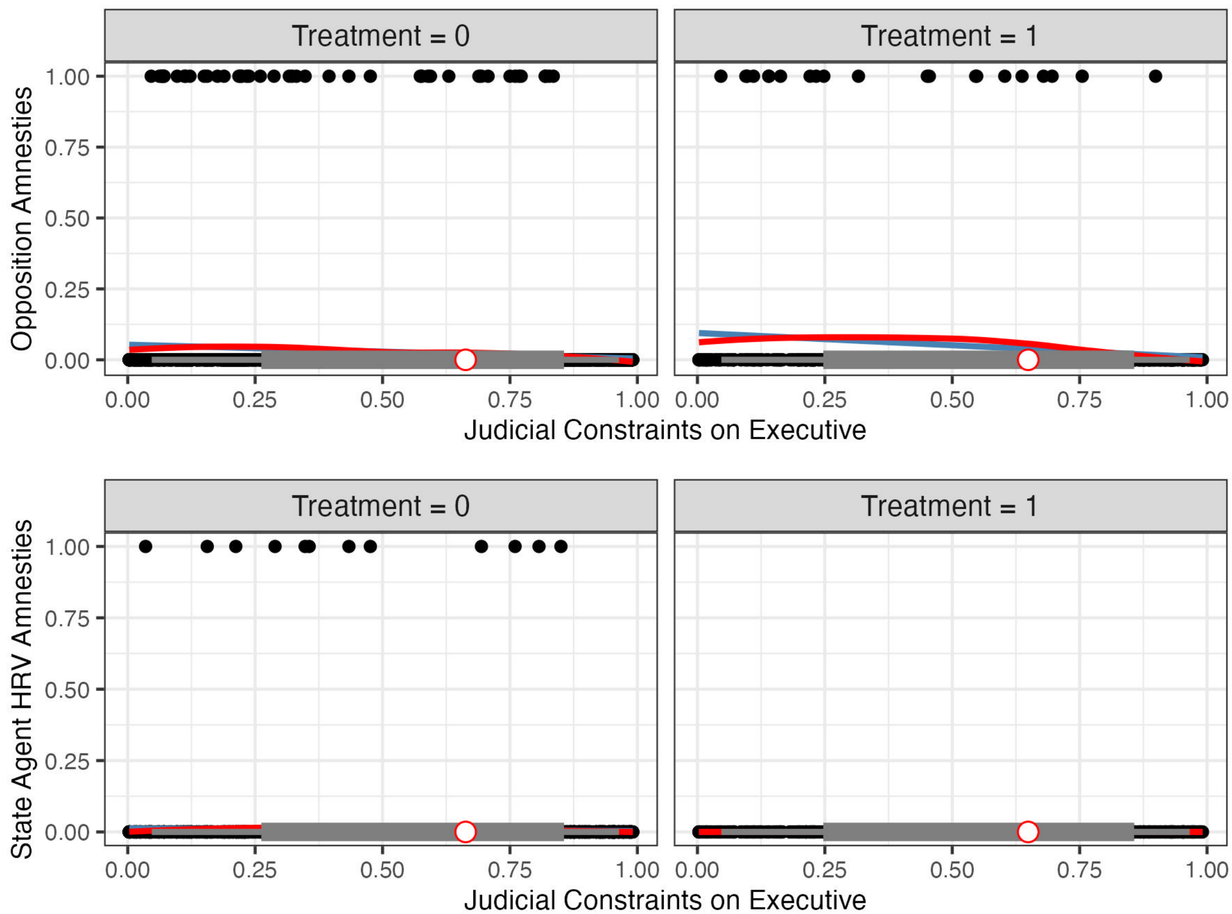

Figure 11: *Linear Interaction Diagnostic Plots (1 of 2).* Figures present scatterplots of outcome by treatment assignment across values of the moderator, the V-Dem Judicial Constraints on Executive Index. Red lines are LOESS curves. Blue lines are OLS regression lines.

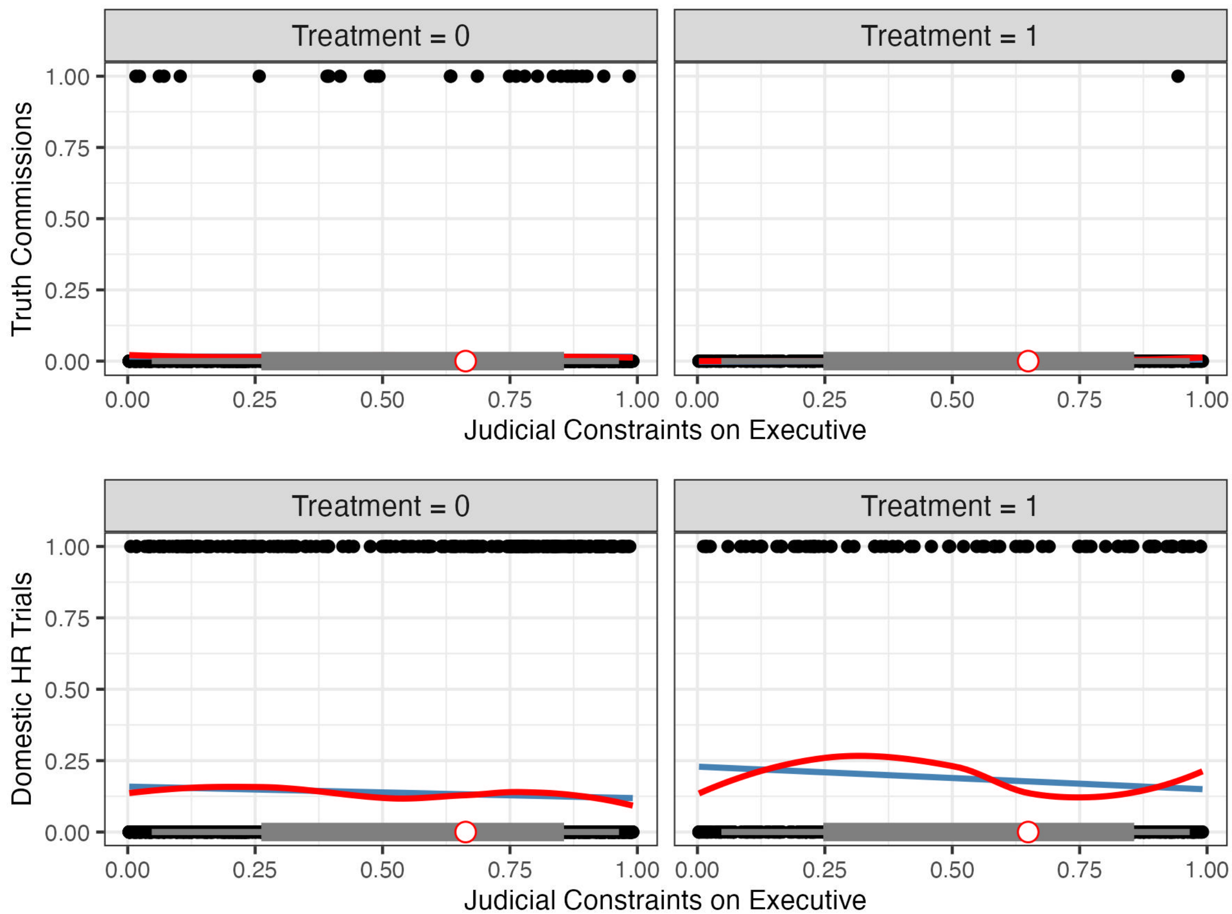

Figure 12: *Linear Interaction Diagnostic Plots (2 of 2).* Figures present scatterplots of outcome by treatment assignment across values of the moderator, the V-Dem Judicial Constraints on Executive Index. Red lines are LOESS curves. Blue lines are OLS regression lines.

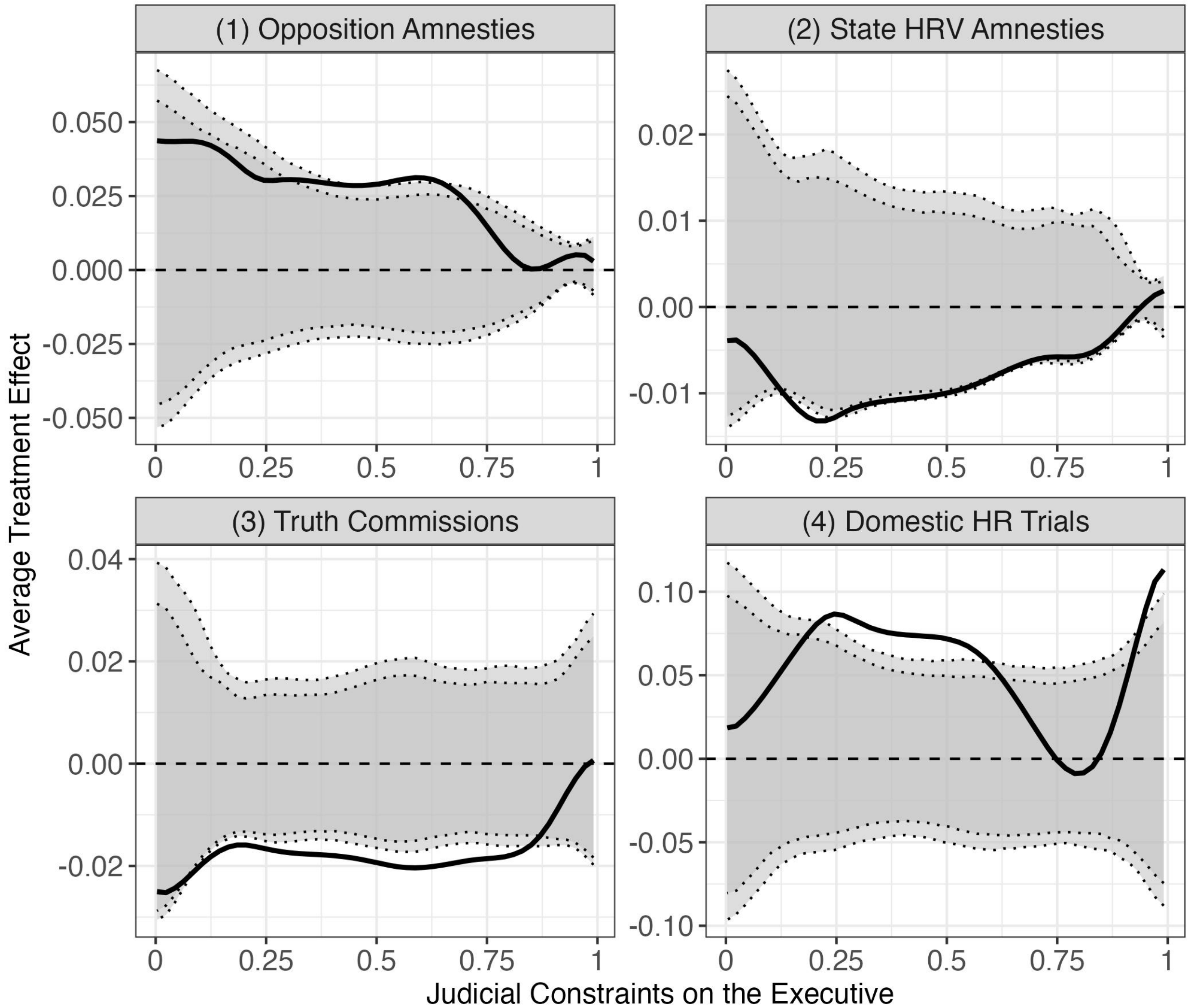

Figure 13: *Interaction of Review and Judicial Constrains on Executive, Replication with Randomization Inference, No HRC Rule.* Each pane presents an interaction model of how the relationship between the treatment and outcome variables depends on the level of executive independence from the judiciary. The distribution of the 95% and 90%-tiles of imputed estimates are represented by the gray shaded regions and dashed lines.

effects; and Model (3) adds the same set of control variables used in prior analyses. Across the board, there is no evidence that UPR impacts the moderator, mitigating concerns about post-treatment bias.

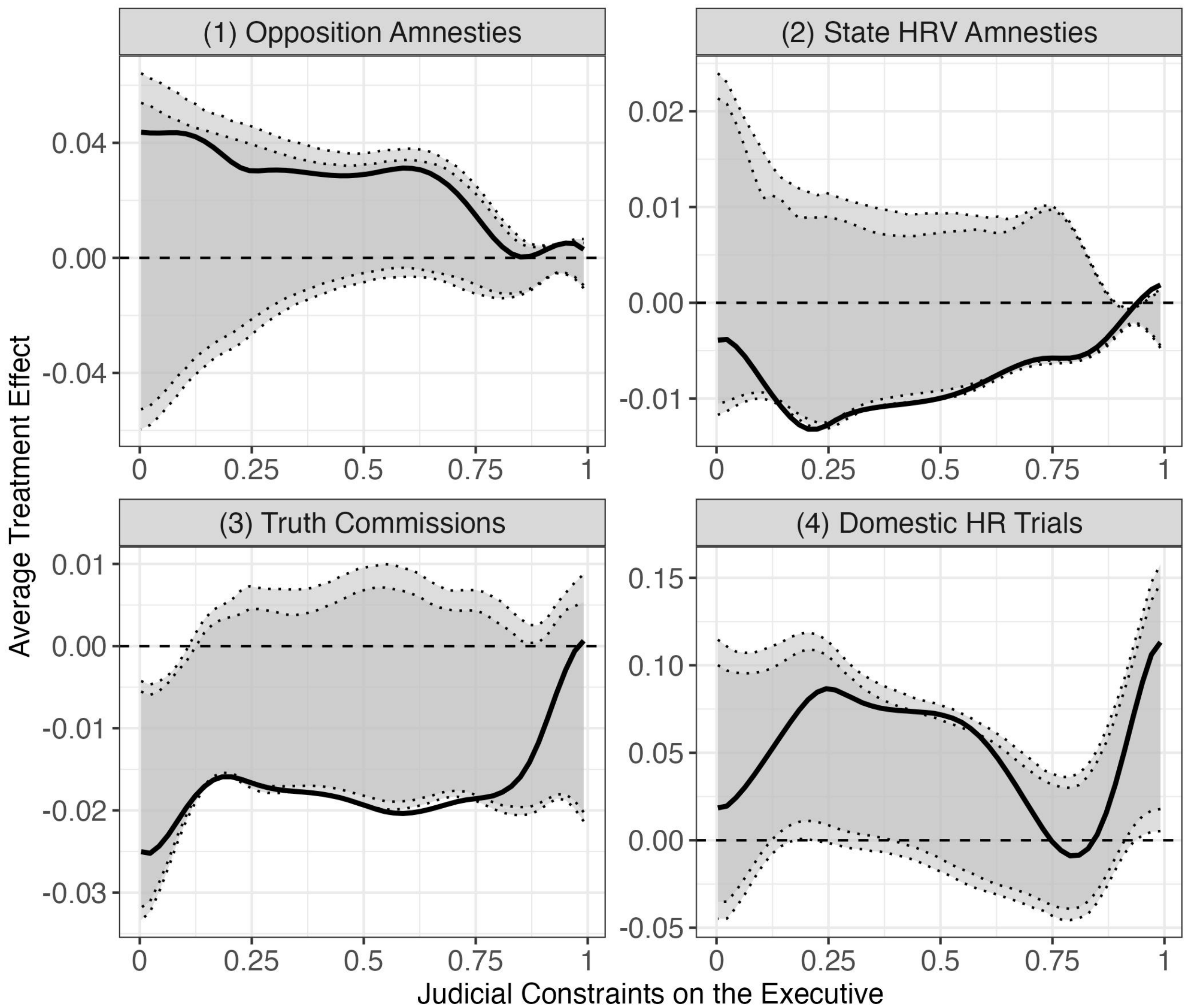

Figure 14: *Interaction of Review and Judicial Constrains on Executive, Replication with Randomization Inference, HRC Rule.* Each pane presents an interaction model of how the relationship between the treatment and outcome variables depends on the level of executive independence from the judiciary. The distribution of the 95% and 90%-tiles of imputed estimates are represented by the gray shaded regions and dashed lines.

Table 2: *Treatment Does Not Impact Moderator*

	Dependent variable:		
	Judicial Constraints on the Executive		
	(1)	(2)	(3)
Review	−0.007 (0.016)	−0.001 (0.004)	−0.004 (0.002)
Polity Score			−0.037 (0.029)
Log GDP Per Capita			0.008 (0.007)
Log Population			0.757*** (0.059)
UNGA Ideal Point			0.002 (0.005)
HRC Member			−0.050*** (0.014)
Country + Year FE	–	✓	✓
Covariates	–	–	✓
Observations	2,259	2,259	2,059
Adjusted R^2	−0.0004	0.955	0.976

Notes: Country-year cluster robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01.