Trojan Horses in Liberal International Organizations?: How Democratic Backsliders Undermine the UNHRC

Anna M. Meyerrose[*]and Irfan Nooruddin[†]

Abstract

Liberal democracy is facing renewed challenges from a growing group of states undergoing democratic backsliding. While entrenched autocrats have long resented and contested the established liberal order, we know far less about how newer backsliding states behave on the international stage. We argue these states, who joined prominent western liberal institutions prior to their backsliding, will use their established membership in these organizations both to protect themselves from future scrutiny regarding adherence to liberal democratic values, and to oppose the prevailing western liberal norms that increasingly conflict with their evolving interests. Using voting data from the UN Human Rights Council (UNHRC) from 2006–2021, we show that backsliding states are more likely to vote against targeted resolutions that name and shame specific countries. We supplement this analysis with detailed data from the Universal Periodic Review (UPR) and combine regression analysis and a structural topic model (STM) to show that backsliding states are more critical in their UPR reports when evaluating advanced western democracies, and more likely to emphasize issues that align with their own interests while de-emphasizing ones that might threaten government power and control over citizens.

---

[*]Assistant Professor, School of Politics and Global Studies, Arizona State University.

[†]Hamad bin Khalifa Al-Thani Professor of Indian Politics, School of Foreign Service, Georgetown University.

Liberal democracy faces its most severe global challenge since the fall of the Soviet Union: in 2021, the number of liberal democracies around the world declined to levels not seen since 1989 and currently represent only 13% of the world's population (Alizada et al., 2022). Liberal democracy's decline has accelerated precipitously over the last decade in large part due to democratic backsliding, a historically unprecedented phenomenon that occurs when democratically elected officials intentionally weaken or erode institutional checks on government power (Bermeo, 2016; Waldner and Lust, 2018; Haggard and Kaufman, 2021). While research has begun to explore the consequences of backsliding for domestic democratic institutions and seeks to provide both domestic- (Haggard and Kaufman, 2016; Mudde and Rovira Kaltwasser, 2017; Ziblatt and Levitsky, 2018) and international-level (Meyerrose, 2020, 2021; Kelemen, 2020) explanations to account for this emerging phenomenon, less attention has been paid to the implications of backsliding for international outcomes.

There exists extensive work on the international interests and behavior of *entrenched* autocrats, with an emphasis on how these autocrats use both new and existing multilateral international organizations (IOs) to advance their agendas, protect themselves from international intervention and scrutiny, and more broadly challenge established western liberal values (Finnemore and Sikkink, 1998). Powerful autocratic states such as Russia and China have long contested western liberal values, particularly in the area of human rights (Binder and Payton, 2022), from within established western IOs (Lebovic and Voeten, 2006; Johnston, 2007; Donno, 2010; Smith, 2016; Carraro, 2017). More recently, these powerful autocrats have even begun to create their own regional IOs (Cottiero and Haggard, 2021) that promote the spread of authoritarian rules and regimes in their neighborhoods (Libman and Obydenkova, 2018; Kneuer et al., 2019) and provide an ideological counter-narrative to the current international environment (Ginsburg, 2020; Debre, 2021; Hafner-Burton, Pevehouse and Schneider, 2023). These and related developments have led pundits and politicians to warn the liberal international order is under threat (Ikenberry, 2018; Börzel and Zürn, 2021; Way, 2022).

Backsliding states are fundamentally altering the composition of long-established western liberal IOs, yet we know little about how backsliding states behave within these organizations. In the post-Cold War era, western-dominated multilateral organizations increasingly sought to widen their membership to bring emerging and fragile democracies into the fold (Pevehouse, 2005; Donno, 2013; Genna and Hiroi, 2014; Poast and Urpelainen, 2018). However, this tactic might now be backfiring. Today, many of the backsliding states of greatest concern are established members of these very western liberal organizations. What impact do these backsliding member states have on the liberalizing goals and overall functioning of these western IOs? How do they leverage their membership in these institutions to protect and advance their evolving and increasingly illiberal interests?

Democratic backsliding is unique when compared to historic cases of democratic collapse. Today, moves away from democracy occur much more gradually than in the past (Bermeo, 2016; Waldner and Lust, 2018), and contemporary cases of backsliding suggest the outcome of ongoing erosions is an illiberal democratic or semi-autocratic regime (Luhrmann and Lindberg, 2019), rather than a full-fledged autocracy. Backsliding states, while increasingly illiberal, heed incentives to appear at least minimally democratic internationally (Hyde, 2011) and also to maintain the patina of democracy to assuage their domestic constituencies. Therefore, we expect backsliding states to display illiberal interests and behaviors that are increasingly similar to those of fully entrenched autocrats even if at times and in some arenas they remain distinct.

We argue backsliding states will use their established membership in western multilateral organizations both to protect themselves from future scrutiny regarding their adherence to liberal democratic values, and also more directly to oppose the prevailing western liberal norms that dominate the international system and threaten the domestic legitimacy of these newly emerging regimes. As these states move away from democracy, they can expect to face increased scrutiny for their pseudo-democratic and illiberal domestic practices. In anticipation, these regimes have incentives to oppose international efforts to enforce western

liberal values and to protect regimes similar to them from international sanctions in the hope that this behavior will be reciprocated in the future. At the same time, these countries are part of a growing global wave of backsliding states, all of which have illiberal interests advocated by powerful would-be autocrats. This emboldens them to challenge the West on the universality of liberal values, and also to promote an alternative set of norms that better align with their evolving illiberal interests. In these ways, we argue the very international fora created, subsidized, and promulgated by the West to promote liberal hegemony might now be used by backsliding states to stymie and undo those efforts.

We test our argument in the context of the international human rights regime, which originated with the United Nations' (UN) Universal Declaration of Human Rights in 1948 and eventually grew into a global framework to promote and monitor states' human rights practices. Human rights are a defining feature of the liberal international order, and are arguably among the most invasive and intrusive elements of this regime, enshrining at the extreme the principle of responsibility-to-protect (R2P). The challenge to unconstrained sovereignty embedded in the human rights regime is particularly objectionable to illiberal regimes, and research shows these norms in particular are heavily contested by illiberal regimes.[1] Specifically, we evaluate our argument using data from the UN Human Rights Council (UNHRC), founded in 2006, and the Universal Periodic Review (UPR), which was introduced in 2008 as an additional tool to promote human rights. The UPR is arguably one of the most elaborate multilateral institutions devoted to promoting and protecting human rights (Terman and Byun, 2022) and naming and shaming states that violate these rights (Kim, 2023).[2]

To our knowledge, no fully updated dataset of UNHRC resolutions exists; therefore,

---

[1]This feature arguably suggests a scope condition for our argument. Backsliding and other illiberal states should work hardest to undermine international organizations that focus on human rights while being satisfied to maintain the status quo in multilateral fora that focus on more technical and less intrusive issues less associated with the liberal international order.

[2]Although the UNHRC's predecessor, the UN Commission on Human Rights (UNCHR) was founded in 1946, we focus on the time period from 2006 to the present since democratic backsliding is a historically recent phenomenon that began in the early- to mid-2010s (Haggard and Kaufman, 2021).

we create a comprehensive dataset that includes voting behavior for all members of the UNHRC from 2006 through 2021. Using these data, we show backsliding states are more likely to vote against resolutions that target specific states, suggesting these countries are using their voting powers in the UNHRC to reinforce the notion of sovereign immunity and non-interference while also undermining the credibility of these resolutions that are being passed in an attempt to guard themselves against future sanctions. We supplement this analysis with detailed data on UPR reports and their content. Regression analysis shows that backsliding states use these reports to target consolidated Western democracies for their own alleged domestic human rights abuses. We also use these UPR data to estimate a structural topic model (STM), finding that backsliding states' reports are more likely to emphasize non-threatening human rights topics related to protecting vulnerable groups and human development, while de-emphasizing ones that might threaten government power and control over citizens, such as civil and political liberties. Taken together, these findings suggest backsliding states are using international human rights fora — a critical component of the liberal international order — simultaneously to protect their own interests, to promote a new set of international human rights practices, and increasingly to scrutinize and counter efforts by Western democracies to enforce international liberal norms and values.

1 Western Multilateral Fora and the Liberal International Order

The liberal international order (LIO) has a long history that can be traced to the years after World War I, but expanded significantly after World War II with the founding of prominent multilateral institutions, with the United Nations (UN) system, the Bretton Woods institutions, and the General Agreement on Tariffs and Trade at the core (Ikenberry, 2009; Börzel and Zürn, 2021). The founders of these and subsequent western international organizations (IOs) viewed them in part as tools to expand and entrench the LIO. While the Bretton Woods institutions promoted open trade and embedded liberalism, the UN was

crucial in managing the transition from empires to sovereign states and founding the international human rights regime. After the end of the Cold War, western-based IOs proliferated, with politically oriented ones (such as the United Nations) in particular becoming inextricably linked to the construction of a liberal global order centered on democracy, elections, rule of law, and human rights (Barnett and Finnemore, 2021).

Indeed, the LIO expanded significantly in the post-Cold War era, a period when western liberal democracies increasingly worked through IOs with the overarching goal of creating "an international 'space' for liberal democracy, reconciling the dilemmas of sovereignty and interdependence, seeking protections and preserving rights within and between states" (Ikenberry, 2018, 8). Throughout the 1990s and early 2000s, these international institutions gained significantly more authority and became decisively more liberal with a strong emphasis on human rights, rule of law, democracy promotion, the free movement of people, global cooperation, liberal values, and material progress via free markets (Pevehouse, 2005; Barnett and Finnemore, 2021; Börzel and Zürn, 2021). Indeed, as democracy promotion became a core foreign policy objective for western powers (Collier and Levitsky, 1997; Carothers, 2002; Levitsky and Way, 2010; Johnstone and Snyder, 2016), membership in these IOs was increasingly conditioned on regime type and the holding of internationally certified free and fair elections (Hyde, 2011). It was in this context that countries in east Asia, eastern Europe, and Latin America transitioned to democracy and sought out and secured membership in these western liberal IOs.

These IOs have been credited with encouraging transitions to democracy, and even facilitating democratic consolidation, among these emerging democratic states in the 1990s and 2000s (Pevehouse, 2005; Donno, 2010; Poast and Urpelainen, 2018).[3] The belief in the power of the liberal international order was so great that the West even believed that tougher cases for democracy, such as China and Russia, could eventually liberalize politically as they became more integrated in these liberal international institutions (Walker, 2016); there-

---

[3]More recent evidence, however, suggests these IOs, and globalization more broadly, can unintentionally contribute to backsliding in new democracies (Meyerrose, 2020, 2021).

fore, their membership and participation in these organizations was also encouraged and expanded.

The LIO sought to promote liberal values such as free-and-fair elections, enhanced rule of law, and more expansive human rights (Barnett and Finnemore, 2021; Finnemore and Sikkink, 1998; Flores and Nooruddin, 2016*b*; Hafner-Burton, 2012; Hyde, 2011). The UN agencies, in particular, were pivotal in these efforts. The UN was founded with the goal of saving "succeeding generations from the scourge of war..., to reaffirm faith in fundamental human rights..., to establish conditions under which justice...can be maintained..., [and] to promote social progress and better standards of life in larger freedom" (*Charter of the United Nations*, 1945). This emphasis on the importance of protecting and spreading human rights led to the creation of the UN Commission on Human Rights (UNCHR) in 1946. The UNCHR was widely viewed as a platform for democratic states to advance liberal norms via socialization (Finnemore and Sikkink, 1998) and an institution that could be used to formalize commitments to norms of liberal behavior (Voss, 2019). Indeed, scholars found that new democracies joined these human rights institutions and treaties at higher rates than both dictatorships and consolidated democracies, using the sovereignty costs associated with membership to commit credibly to upholding international law and to becoming consolidated democracies (Moravcsik, 2000; Landman, 2005; Hafner-Burton, 2012).

Nevertheless, the UNCHR faced significant criticism throughout its lifespan, with many arguing it was too heavily exposed to political influence, and that its rules for choosing members left open the possibility (too often verified by reality) that states with poor human rights records could become members and de-legitimize its work (Edwards et al., 2008). In light of these issues, the UNCHR was dissolved and replaced by the UN Human Rights Council (UNHRC) in 2006. In its current form, the Council consists of 47 member states elected for 3-year terms. The UNHRC reports directly to the UNGA, and therefore has higher status than its predecessor. While few amendments were made to membership criteria (Cox, 2010; Hug and Lukács, 2014), in other ways, the UNHRC implemented significant changes:

the Council now meets more often (Edwards et al., 2008) and, in 2008, initiated a new tool — the Universal Periodic Review (UPR) — to assess human rights situations in *all* UN member states.

Over time, western liberal democracies have used and expanded these formal human rights institutions with other tools to promote human rights and, when necessary, to name and shame states that violate them (Koliev, 2020; Kim, 2023). Indeed, the UNCHR's activities increased significantly following the end of the Cold War, becoming particularly far-reaching in the mid-1990s, and its activities increasingly targeted and in particular punished the worst human rights violators (Lebovic and Voeten, 2006). Scholars also find peer review groups, such as the UPR, can be effective tools for naming and shaming (Carraro, Conzelmann and Jongen, 2019), though the particular human rights norms that are enforced in this context are contingent on the nature of the relationship between the reviewing and target states (Terman and Byun, 2022). Western democracies have also found less direct ways of punishing states that violate liberal norms, for example by tying human rights rules directly to market access via preferential trade agreements (Hafner-Burton, 2005). Scholars disagree, however, about the utility of these tactics. While there is evidence that naming and shaming is an effective tool used by IOs and NGOs to punish human rights violations (Lebovic and Voeten, 2006; Risse and Sikkink, 1999; Woo and Murdie, 2017; Terman and Voeten, 2018), other research suggests human rights treaties and naming-and-shaming are unable to influence human rights adherence in target states meaningfully Tsutsui and Wotipka (2004); Hafner-Burton and Tsutsui (2005); Neumayer (2005); Vreeland (2008) and in some cases might even fuel additional human rights abuses (Hafner-Burton, 2008).

2 The Global Democratic Recession and Autocratic International Cooperation

Despite extensive efforts to promote democracy, human rights, and other liberal values via IOs in the post-Cold War environment, the last decade and a half has witnessed a sharp

rise in cases of democratic backsliding and rampant illiberalism in new democracies whose transitions to democracy were once heavily influenced and supported by the international community (Meyerrose, 2020), and in particular by the West.

We focus here on backsliding states in particular. Luhrmann and Lindberg (2019) identify three waves of "autocratization," each of which followed their corresponding waves of democracy (Huntington, 1991). The time period on which we focus is part of the third wave of autocratization, which began around 1993 and continues into the present. Autocratization denotes moves away from democracy (or, toward autocracy) that can occur in any type of regime. When applied to democracies, autocratization can denote both abrupt regime changes, such as those that result from a classic coup d'état, as well as more gradual and incremental erosions that result in either an illiberal, diminished democracy or involve an eventual transition to autocracy. The current third wave of autocratization is distinct from previous waves in that moves away from democracy are occurring more gradually (Bermeo, 2016; Waldner and Lust, 2018), and these moves — which we term 'backsliding' — are increasingly evident in an unprecedented number of erstwhile democracies (Luhrmann and Lindberg, 2019).

Given the unique characteristics of the third wave of autocratization, our focus is on *democracies* that are inexorably sliding away from democracy. In other words, we are concerned with cases of democratic backsliding, which we consider a subset of autocratization. Democratic backsliding occurs when democratically elected officials weaken or erode institutional checks on government power (Bermeo, 2016), including the constitution, rule of law, civil and minority rights, the independence of the judiciary and the media, and separation of power within governments (Meyerrose, 2020; Haggard and Kaufman, 2021). Democratic backsliding is a historically recent and unprecedented phenomenon that has impacted all types of democracies, including those ones that joined the UN and other liberal IOs in the post-Cold War era, and it has become increasingly frequent: as of 2021, the number of liberal democracies around the world had declined to levels not seen since 1989 (Alizada et al.,

8

2022). Contemporary cases suggest the outcome of ongoing erosions, at least in the medium term, is a sort of illiberal democracy or semi-autocratic regime, rather than full-fledged authoritarianism (Luhrmann and Lindberg, 2019). Thus, as democratic backsliding has become increasingly common, illiberal democracies and semi-autocracies have proliferated, expanding the number of non-liberal democratic states both in the world and also within the very western liberal IOs charged with promoting and maintain the liberal international order.

Concurrent with this global democratic recession, autocracies and backsliding states taken active measures at the international level to evade western pressures to adhere to liberal values. To-date, most research on autocratic cooperation has focused in particular on autocratic regional IOs, or on IOs dominated by autocratic members (Cottiero and Haggard, 2021). Powerful autocratic states such as Russia, China, Venezuela, and Saudi Arabia have created regional organizations such as the Shanghai Cooperation Organization, the Bolivarian Alliance for the Peoples of Our America, and the Gulf Cooperation Council as alternatives to established western liberal IOs. By providing member states with economic and military support, increasing regime legitimacy, and challenging international pressures to democratize (Weyland, 2017), these regional organizations serve as one tool for powerful autocratic states to promote the spread of authoritarian rules and regimes in their neighborhoods (Libman and Obydenkova, 2018; Kneuer et al., 2019) and legitimize an ideological counter-narrative to liberal hegemony (Ginsburg, 2020; Debre, 2021). These regional organizations not only provide an alternative to the prevailing liberal democratic values, but can also be used to help autocratic leaders maintain power at home, in part by appeasing international actors committed to promoting democracy, good governance, and liberal values (von Soest, 2015; Hafner-Burton, Pevehouse and Schneider, 2023). By passing regional Human Rights Charters or sending accommodating election observers to member states, such regional organizations allow autocratic leaders to signal to their domestic audiences that (western liberal) international accusations of flawed elections or abuses of human rights are unfounded (Debre and Morgenbesser, 2017; Debre, 2021; Donno et al., 2023; Bush, Cottiero and Prather, 2023).

While these autocratic regional organizations will undoubtedly continue to gain influence in the coming decades, western-based IOs still dominate the multilateral landscape. Yet, even there, evidence is growing that powerful, established autocracies use their membership in these IOs to subvert from within institutional efforts to promote liberal democratic values. Binder and Payton (2022) show that rising powers such as Brazil, Russia, India, China, and South Africa have begun to form a bloc of dissatisfied states, as exemplified by their voting behavior in the UN General Assembly. Focusing on the UNCHR, Hug and Lukács (2014) similarly find authoritarian states pursued membership on the Commission to dilute international human rights norms and to advance illiberal ones; they also worked to shift the focus from civil and political rights to more purely economic issues. Repressive states also used their membership in the UNCHR to insulate themselves from investigation and deflect attention and blame toward other states (Edwards et al., 2008; Cox, 2010; Seligman, 2011; Hug and Lukács, 2014; Vreeland, 2019). More recent data from the current UNHRC also suggests states from the Global South, including Russia, use amendments to UNHRC resolutions to protect their own human rights preferences and also to challenge the existing international order (Voss, 2019). Looking beyond established autocracies, there is also evidence that new democracies, which are over-represented in contemporary cases of backsliding, exhibit distinct voting patterns in western liberal IOs, such as the European Parliament (Meyerrose, 2018).

The post-Cold War push to expand the LIO resulted in then-emerging and fragile democracies —many of which, more recently, have begun backsliding toward autocracy (Luhrmann and Lindberg, 2019)— also becoming members of western liberal IOs. Today, these backsliding and autocratizing states are joining fully autocratic states as established members of the very IOs that, historically, have been at the forefront of promoting and supporting the LIO. While initially this membership expansion was viewed favorably by proponents of liberal democracy, broadening the reach and potential influence of western liberal IOs via expanded or universal membership may be incompatible with deepening, or, the ability of

these organizations to accomplish their goals of Western liberal values.[4] We develop our argument in the following section.

3 Undermining the LIO from Within: Backsliding States in Western Liberal IOs

The fall of the Soviet Union left the United States as the undisputed hegemon of the international system in the 1990s. In the preceding decades the US and its partners had spearheaded a proliferation of international organizations, covering every topic imaginable, from the purely political to the very technical. The decisions and rulings rendered by this dense web of IOs generated a robust body of international law, the enforcement of which, while uneven, was backed by uncontested US power. A particular focus of this new liberal international order was the promotion of democracy, which in turn involved a more intrusive examination of member state's domestic politics.

The power of the LIO and the absence of an alternative champion left developing countries with scant choice but to sign on to the corresponding international organizations, and also to accept the greater scrutiny of their domestic policy records. As Susan Hyde has argued, this spawned a generation of 'pseudo-democrats' who hoped that being seen to hold free and fair elections could unlock material benefits, while fearing that explicitly violating this new norm might invite punishment (Hyde, 2011). Elsewhere, as documented in the literature reviewed above, membership in the LIO meant subjecting oneself to new pressures to improve human rights records, to protect labor and the environment from corporate predation, and, even, to accept new international accounting standards to facilitate trade and investment.

For the developing world, being admitted to the liberal club was important symbolically for the domestic legitimacy and international rewards it conferred but also risky because of the challenges to unfettered domestic sovereignty. But with democracy ascendant and unrivaled, these countries had little choice but to go along with the wave. However, this status

---

[4]Pahre (1995) illustrates this theoretical tension in the context of the European Union's expansion.

quo changed faster than anyone had anticipated, and a mere two decades after the 'end of history' (Fukuyama, 1989) was celebrated, the world was on the cusp of a serious democratic recession. The United States was distracted by its wars in Iraq and Afghanistan, and the moral suasion power it exercised was severely undermined by the loss of legitimacy caused by those wars, and particularly the false pretenses under which the Iraq war was waged. The staggering growth of China's economy provided both alternative sources of aid and investment for many developing countries chafing under Western rules, but also legitimated a different model of governance that privileged stability and performance over multi-party elections. Democratic backsliding grew more common, as the elections that had become commonplace in the 1990s no longer yielded any gains in democracy but, if anything, often harmed the broader democratic project (Flores and Nooruddin, 2016a; Meyerrose, Flores and Nooruddin, 2019). A key point to remember is that backsliding did not necessarily mean new leadership; rather it was the incumbent governments that engineered the undermining of electoral integrity, indulged in electoral violence, and repressed their oppositions to ensure their continued hold on power (Flores and Nooruddin, 2022). These democrats-turned-autocrats, having cemented their rule at home, now had to limit and deflect international criticism of their actions.

Membership in IOs provided democratic backsliders an ideal forum in which to undermine the LIO's pressure on them. Because they were already members, having been admitted to these IOs either at independence in the case of the universal IOs, or on the backs of their democratizing credentials for the more exclusive ones, they could exercise their voting privileges in these organizations to limit the efficacy of the IOs. The UNHRC is a perfect example. Here is an organization whose sole purpose is to use the power of the United Nations system to embarrass countries that are violating human rights and to issue condemnations that have the force of the global Human Rights Council. Yet, membership in the Council is not predicated on a country's own democratic or human rights records, but rather the UNHRC's membership is elected by regional peers. Over time, this meant that the global democratic

recession also resulted in a backsliding of democracy *within* the Council's membership.

We argue that backsliding countries should be less likely to vote in favor of UNHRC resolutions targeting other states for alleged human rights violations. These countries have the most to gain from undermining the ability of IOs to scrutinize the domestic politics of nations, and have the most to gain from reinforcing the notion of sovereign immunity within the international system. If correct, this has an important consequence for the UNHRC more generally because it means that increasingly its resolutions, even if passed, will do so with narrower majorities than before due to a larger number of 'nay' votes.[5] This in turn undermines the moral force of UNHRC resolutions and, as such, their influence. This logic suggests the following testable hypothesis:

Hypothesis 1: *Backsliding states are less likely to vote in favor of UNHRC resolutions.*

Voting on specific resolutions allows states to register visibly their support for or opposition to the principles of the LIO. Indeed, recent research suggests states interested in weakening the international human rights monitoring system are increasingly using the UNHRC to escalate tensions between the North and the South (Lakatos, 2022). However, their ability to do so in a setting governed by majoritarian voting rules is constrained. The creation of the UPR mechanism inadvertently provides an additional means for non-democracies to challenge prevailing liberal norms and western dominance. Because the UPR allows countries to comment on other countries' human rights situations in a bilateral setting, it serves not only as a way for liberal democracies to promote human rights, but also as a tool for backsliding and autocracy-leaning states to challenge directly their would-be critics. One way to do this is to place these western democracies under increased scrutiny for their own alleged domestic human rights shortcomings.[6] This helps deflect attention from non-democracies and also undermines the credibility of states at the forefront of the promotion of human rights. This suggests the following testable hypothesis:

---

[5]For a related project that is consistent with our approach, see Lipps and Jacob (2023).

[6]A fascinating case-study of this approach is provided by Baturo (2023) who shows how Russia has used the language of anti-Nazism to dull criticism of its domestic policies.

Hypothesis 2: *Backsliding states will be more critical in UPR reports targeting western liberal democracies.*

The UPR also allows countries to reveal the issues they think are most worthy of discussion by the international community. Autocratic-leaning states have long argued that western states' emphasis on individual political and civil liberties is at best partial and not reflective of their country's needs. Rather they argue that greater focus should be paid to economic and social rights. We do not disagree that all rights are worthy of elevation by the international community, but see in the UPR mechanism an opportunity to understand whether autocratic and backsliding states are more likely to engage a different set of issues than their consolidated democracy counterparts. In particular, we seek to test the following hypothesis:

Hypothesis 3: *Backsliding states are less likely to emphasize democratic rights and civil liberties in their UPR recommendations.*

Taken as a set, if confirmed by data, these hypotheses shed new light on how the global trend in democratic backsliding undermines existing IOs such as the UNHRC in promulgating the core ideas of the LIO. If such states are less likely to vote in favor of UNHRC resolutions condemning human rights violations in particular countries, thereby generating fewer unanimous condemnations, and if such states are more likely to de-emphasize political and civil liberties in favor of other framings of rights in their use of the UPR mechanism, then the overall efficacy of the UNHRC is undermined since the organization can no longer claim to speak with one voice. Over time, the toll of such internal division within the UNHRC must be to undermine its very reason for existence. In the next section, we analyze country voting data in the UNHRC and a text dataset from UPR recommendations to assess if these concerns are warranted, or hyperbolic.

4 Data and Results

We test our argument using data from human rights organizations within the UN. Given its universal membership, the UN is particularly vulnerable to backsliding states interfering with its efforts to support the liberal values of its founding Charter. We focus on two tools for human rights promotion and protection in particular: the UN Human Rights Council (UNHRC) and the Universal Periodic Review (UPR).

4.1 Voting in the UNHRC

We test our theory first in the context of the UNHRC. The UNHRC was founded in 2006 after its predecessor the UNCHR was dissolved (Hug and Lukács, 2014). The present-day UNHRC consists of 47 member states that are elected by the UN General Assembly for three-year terms. However, there are no requirements for membership; as such, even backsliding states —and states with abysmal human rights records— have the opportunity to be elected to the UNHRC (Edwards et al., 2008; Hug and Lukács, 2014). As Figure 1 illustrates, while the average democracy score of the UNHRC and its predecessor has always been relatively low, this average has declined even more recently, reflecting the ongoing global democratic recession.

Building on work that finds regime type influences how states vote on targeted UNHRC resolutions, we test if these backsliding states exhibit distinct voting patterns. Existing research suggests states use international human rights institutions to advance norms that characterize their domestic politics (Edwards et al., 2008), and that democracies are more likely to vote in favor of targeted UNHRC resolutions than their non-democratic counterparts (Seligman, 2011). Therefore, we expect backsliding states will seek to protect state sovereignty, and also to dilute efforts at human rights promotion by voting against targeted UNHRC resolutions.

We explore how backsliding states behave in the UNHRC with a dataset that includes all
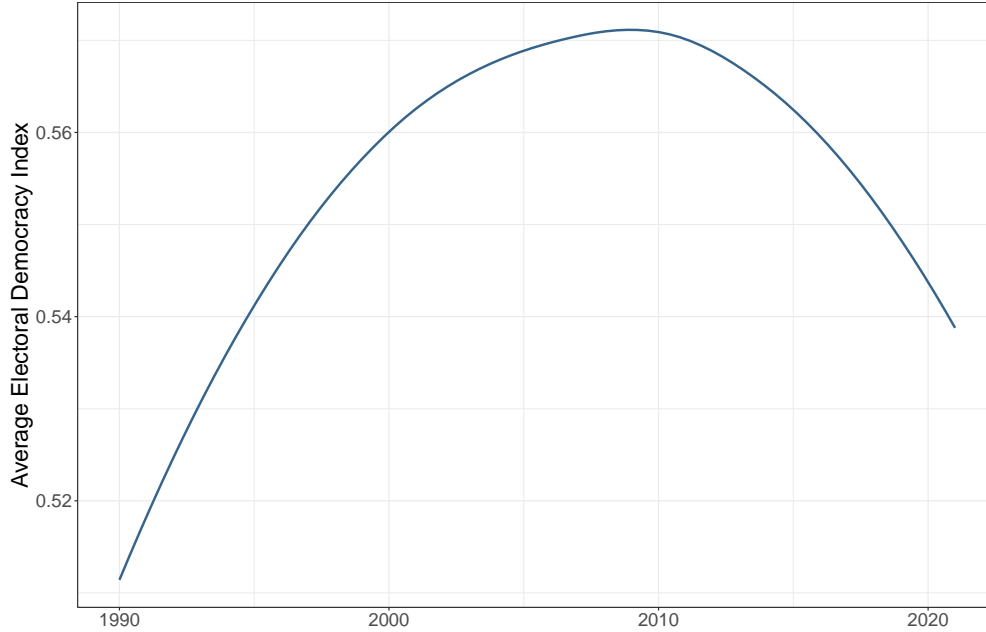
Figure 1: Average electoral democracy index scores of UNHRC members by year. Note that the electoral democracy index is continuous between 0 and 1. Data source: Varieties of Democracy dataset (Coppedge et al., 2021).

UNHRC resolutions that were a) targeted at a specific state and b) decided using a recorded roll-call vote among all members of the UNHRC between 2006 and 2021. We focus exclusively on this time period since it limits the sample to resolutions voted on in the current UNHRC, rather than its predecessor organization. From a theoretical perspective, this time period is also of particular relevance since democratic backsliding is a recent phenomenon that began in the early-2000s.

To our knowledge, no fully updated dataset of votes on targeted UNHRC resolutions currently exists. Therefore, we scrape this information directly from the UNHRC online library for all resolutions from 2006 through 2021[7] and combine this with existing data on targeted resolutions in the UNCHR prior to 2006. The resulting dataset contains information on how each member of the UNHRC voted on any given targeted resolution[8] and the state that is the subject of the resolution. Therefore, our dataset contains observations at the

---

[7]https://searchlibrary.ohchr.org/search?ln=en&cc=Voting. Last accessed 9-March-2023.

[8]The possible votes are "yes," "no," or "abstain"
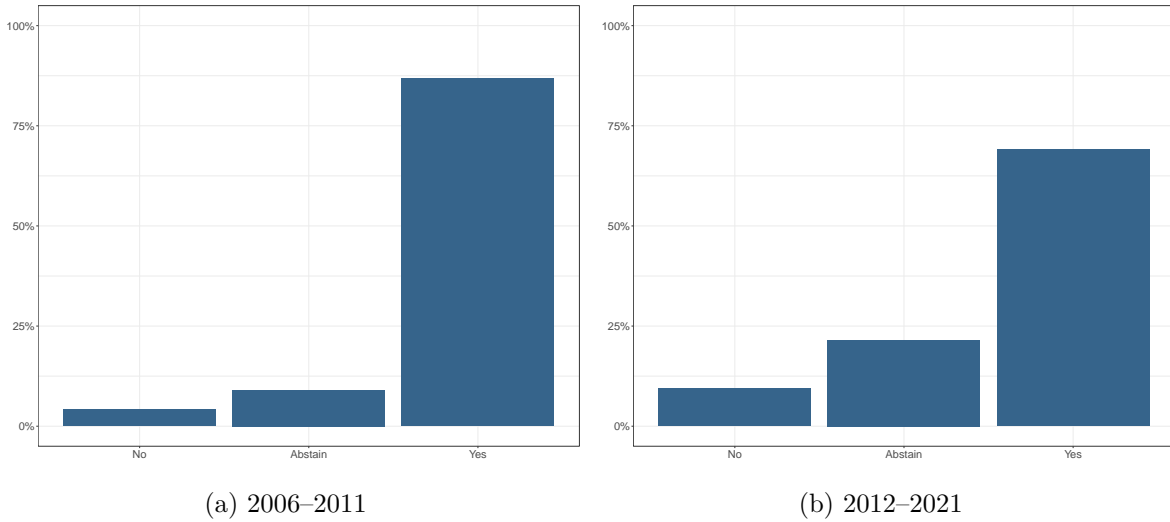
(a) 2006–2011   (b) 2012–2021

Figure 2: Distribution of Votes in UNHRC, 2006–2021

resolution-UNHRC member state level of analysis and, for the time period of interest, consists of 13,721 observations covering 120 unique resolutions for which votes were recorded and adopted by the UNHRC targeting 18 states between 2006 and 2021.[9]

We predict that backsliding states will be less likely to vote in favor of targeted UNHRC resolutions than their more democratic counterparts. Figure 2 shows that between 2006 and 2021 "yes" votes are by far the most common type of vote. However, when we compare votes in the period before backsliding became more common (2006–2011) to the more recent back-sliding era of 2012–2021, we see a marked increase in the proportion of "no" and "abstain" votes in the UNHRC. Figure 3 illustrates this starkly and plots the percent of "yes" votes in the UNHRC per year, making clear the precipitous decline in consensus on resolutions that begins around the same time when backsliding became increasingly common both globally and also amongst UNHRC members, as illustrated in Figure 1 above.

The dependent variable in our analysis is a UNHRC member state's vote choice on any given targeted resolution. Since UNHRC membership rotates, not all states vote on targeted

---

[9]The states targeted in these resolutions, and the number of times they were targeted, are: Belarus (10), Burundi (7), Democratic People's Republic of Korea (1), Democratic Republic of the Congo (1), Eritrea (3), Ethiopia (1), Georgia (4), Iran (7), Israel (29), Myanmar (6), Nicaragua (3), Philippines (1), South Sudan (1), Sri Lanka (2), Syria (30), Ukraine (6), Venezuela (5), and Yemen (3).
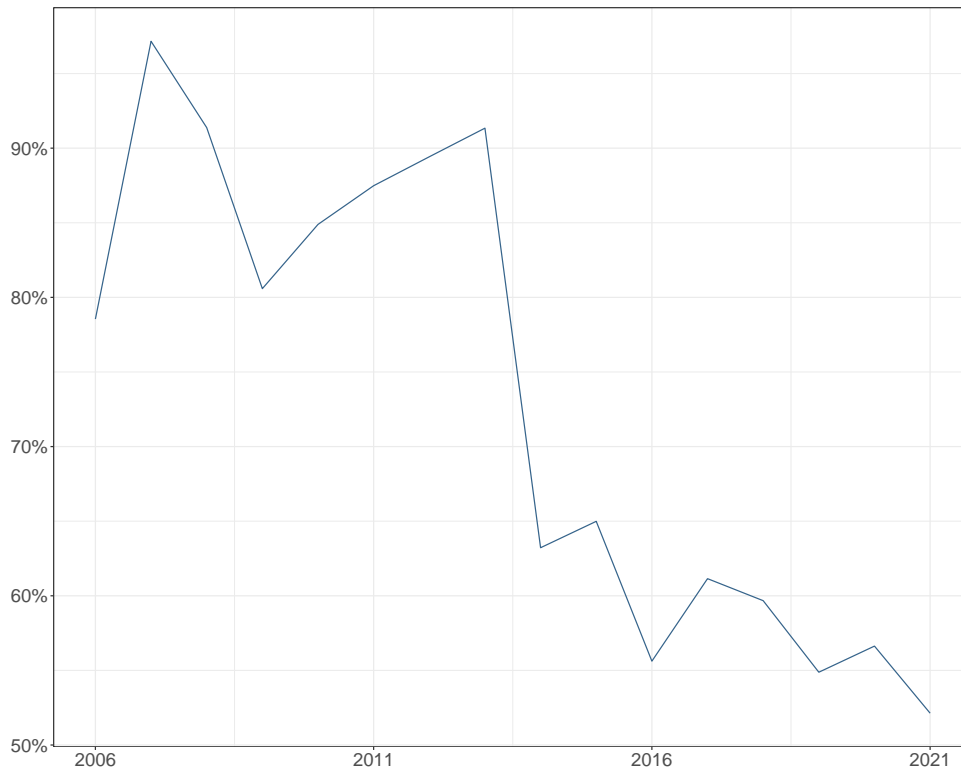
Figure 3: Percent of yes votes in the UNHRC by year, 2006–2021

resolutions, and there is variation amongst those who do with respect to the number of resolutions on which they vote. Figure 4 shows the number of targeted resolutions each UNHRC member state voted on between 2006 and 2021. During this time period, 131 states cast a vote on at least one targeted resolution.

Our main independent variable is a binary indicator for whether or not the voting state has shown evidence of democratic backsliding. To create this variable, we use the electoral democracy index (EDI) from the Varieties of Democracy (V-Dem) dataset (Coppedge et al., 2021). The EDI captures the extent to which a state meets Dahl (1971)'s definition of polyarchy, defined by the presence of universal suffrage, free and fair elections, varying sources of information, and freedom of expression and association. The EDI is a continuous variable taking values between 0 and 1, with higher values indicating more democratic states; this indicator has been used in a number of recent studies to capture backsliding (Jee, Lueders
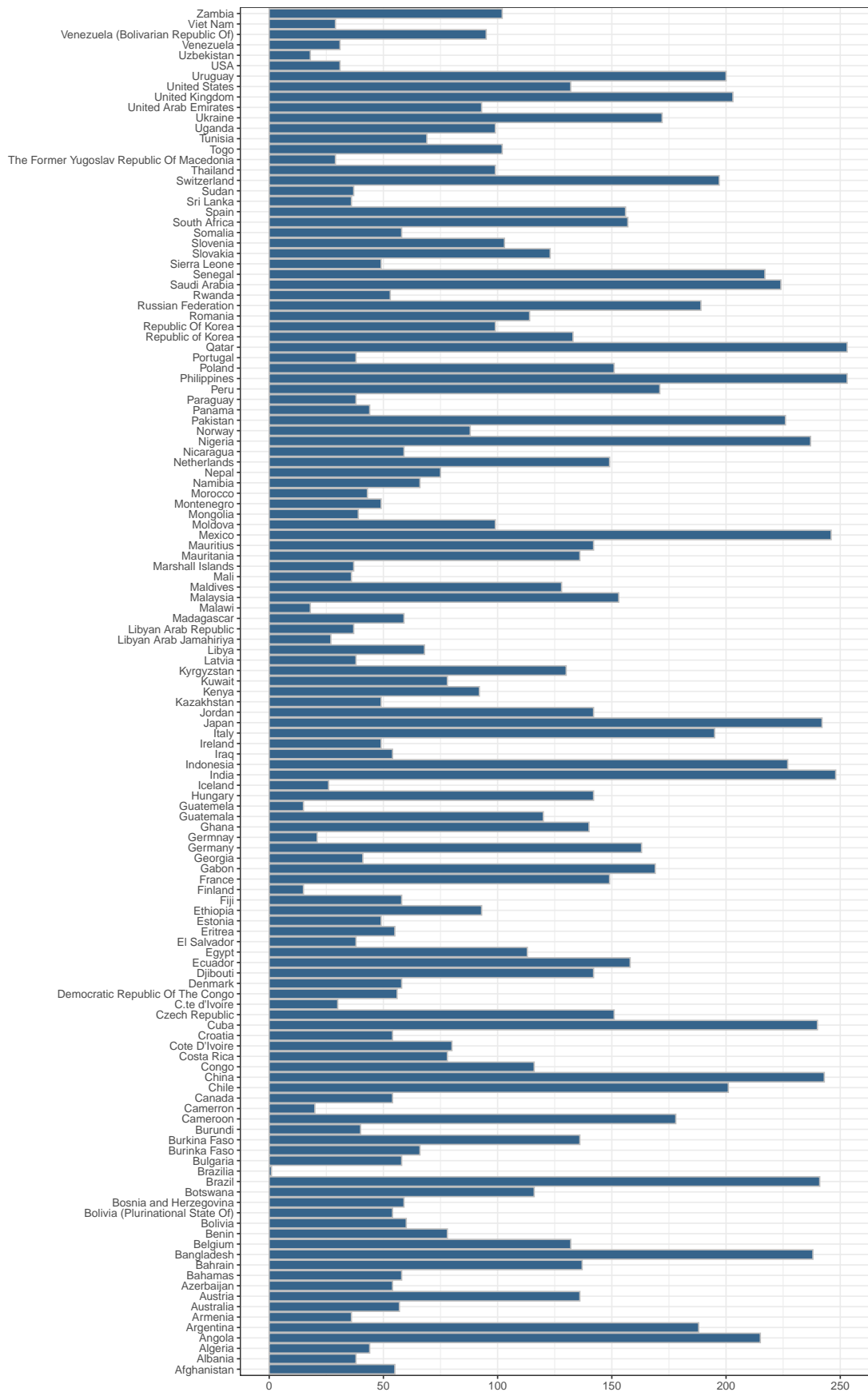
Figure 4: Total number of votes in the UNHRC per country, 2006–2021

and Myrick, 2022). As noted above, we define backsliding as a process that uniquely occurs (or at least begins) in democratic states; therefore, as a starting point we identify all countries that were democracies in 1995; any country with an EDI score greater than 0.5[10] in 1995 is coded as a democracy in that year.[11] Among these democracies, we identify cases that have experienced democratic decline since 1995. Since small declines along the EDI may simply be due to measurement error or other idiosyncratic factors, we follow Luhrmann and Lindberg (2019) and set a threshold to identify substantial changes in a state's level of democracy. Specifically, we operationalize backsliding as any country-year observation for which a UNHRC member state's EDI has declined by 0.1 (e.g. 10%) or more since 1995. Our main variable is coded as 1 for all country-year observations where the voting state's EDI was above 0.5 in 1995 and has decreased by 0.1 or more since 1995; otherwise, it takes a value of 0.[12]

We also include a relevant set of control variables for both the voting and target states. On the voting state side, we control for its current electoral democracy score, in addition to the backsliding indicator. While we expect backsliding to make a state more likely to

---

[10]The 0.5 threshold on the EDI is frequently used in recent work on backsliding to distinguish democracies from autocracies (e.g., (Haggard and Kaufman, 2021)).

[11]We choose 1995 as our starting point for several reasons. First, 1995 roughly corresponds with the end of the third wave of democracy (Huntington, 1991; Alizada et al., 2022). Furthermore, 1995 also marks the beginning of what scholars have termed the third wave of autocratization, which followed the third wave of democracy and continues into the present. One of the defining features of this third wave of autocratization is that, unlike previous waves, democracies are over-represented in the universe of countries moving toward autocracy (Luhrmann and Lindberg, 2019); in other word, democratic backsliding is a core characteristic of the third wave of autocratization.

[12]The following UNHRC member states meet these criteria for backsliding in one or more of the years between 2006 and 2021: Bangladesh, Bolivia, Brazil, Ecuador, Fiji, Hungary, India, Madagascar, Nicaragua, Philippines, Poland, Sri Lanka, Ukraine, and Venezuela. While our rationale for using 1995 as a baseline to identify these cases of backsliding states in the UNHRC is related to the end of the third wave of democracy and the start of the subsequent wave of autocratization, there is no single year that clearly marks the end of the former and beginning of the latter. To explore the effect of using 1995 as our baseline year, in Appendix Figure A1 we show the democratic trajectory of the states we identify as cases of backsliding among UNHRC members from 1985 until 2021, finding that overall these states' levels of democracy increased in the 10 years prior to 1995. Furthermore, with a few exceptions, these states experienced steady, monotonic declines in their democracy scores sometime after 1995. Furthermore, in Appendix Table A1 we show what our universe of cases of backsliding members of the UNHRC would be if we set the baseline years as 2000 and 2006, the year in which the UNHRC was founded. Here, we find that most of the cases of backsliding we identify with the 1995 baseline are also coded as backsliding with these alternative baseline years. This gives us further confidence in our choice of 1995 as the baseline.

vote against UNHRC resolutions, backsliding can result in an illiberal democratic, semi-autocratic, or fully autocratic regime; this endpoint is likely also relevant in predicting how a state will vote. We also control for the proportion of UN General Assembly (UNGA) votes on which the voting state agreed with the US in a given year (Bailey, Strezhnev and Voeten, 2017), as well as for whether or not the voting state is located in the same (UN) geographical region as the target state.

Our model includes relevant characteristics of the state targeted by a UNHRC resolution as well. First, we include an indicator for whether or not the resolution is targeted at Israel, as research suggests voting patterns against Israel in the UNHRC are distinct (Seligman, 2011). In addition, we include variables to capture the target state's current electoral democracy score, its agreement score with US votes in the UNGA, and its human rights score (Herre and Roser, 2016).[13] Summary statistics for all variables included in our analysis are reported in Appendix 2.

To test the extent to which backsliding influences voting behavior on targeted UNHRC resolutions, we estimate a multinomial logistic regression model with year fixed effects. As noted above, the voting outcome can take one of three values: yes, abstain, or no. We designate "yes" votes as the baseline category for the dependent variable in our model and report the results in Table 1. Although we estimate a single model, we report the log-odds for abstentions in the first column, and the same for "no" votes in the second. Since most states vote on multiple resolutions, the observations in our data are not independent. Therefore, we also cluster standard errors by voting state.

Overall, we find evidence that backsliding states are more likely to vote against targeted UNHRC resolutions. Specifically, we find that, for a democracy that has backslid since 1995, its log-odds of voting "no" relative to "yes" increases by 0.70. To facilitate interpretation, we also calculate the relative risk of a backsliding state voting "no" rather than "yes" by exponentiating the reported coefficient: a backsliding state is roughly 2 times more likely to

---

[13]While many studies use data on human rights violations from the CIRI human rights data project, those data only go through 2011. We use the Our World in Data dataset, which provides coverage through 2019.

Table 1: Democratic backsliding and vote choice in the UNHRC, 2006–2021

	Dependent variable:	
	Abstain	No
	(1)	(2)
Backslide since 1995	−0.12	0.70***
	(0.11)	(0.13)
Current EDI score	−0.32**	−2.62***
	(0.13)	(0.21)
Target state = Israel	−0.11	0.06
	(0.19)	(0.31)
Target dem score	1.83***	2.12***
	(0.31)	(0.48)
Voting and target in same region	−0.90***	−0.66***
	(0.13)	(0.16)
Voting UNGA US agree	−1.16***	1.08***
	(0.26)	(0.41)
Target UNGA US agree	−1.50***	−2.79***
	(0.35)	(0.58)
Target state HR score	0.25***	0.05
	(0.04)	(0.07)
Year fixed effects	Yes	Yes
Clustered standard errors	Yes	Yes
Model type	Multinomial logit	Multinomial logit
Observations	11,809	11,809
K	3	3
Akaike Inf. Crit.	13,238.23	13,238.23

Note: *p<0.1; **p<0.05; ***p<0.01

vote "no" as opposed to "yes" on a targeted UNHRC resolution. The negative coefficient for the current democracy score variable in the "no" column of Table 1 also suggests that as a state's level of democracy decreases, it is more likely to vote "no," rather than "yes," on targeted UNHRC resolutions.[14]

Since the magnitude of backsliding may influence how a state votes on targeted UNHRC resolutions, we also re-estimate the model from Table 1 using a continuous, rather than

---

[14]We define backsliding as a process unique to democracies in that the starting point for any case of backsliding is within a democratic state. However, the literature recently has focused more broadly on autocratization, which is defined as any move away from democracy in both democracies *and* autocracies (Luhrmann and Lindberg, 2019). When we operationalize backsliding as a 10% or greater shift away from democracy in any state, regardless of starting point, we find similar results. See Appendix 3.

binary, variable to capture backsliding. When backsliding is operationalized as the change in a state's EDI since 1995, we again find that backsliding states are more likely to vote against and less likely to abstain relative to the baseline yes.[15]

To summarize, Table 1 provides support for our expectations regarding the behavior of backsliding states as outlined in Hypothesis 1: backsliding democracies are more likely to vote against targeted UNHRC resolutions. This effect is robust to alternate definitions and thresholds of democratic backsliding. We next test the extent to which these states use the UN's human rights mechanisms to challenge the western liberal democracies who work to promote and sustain the liberal international order more directly.

4.2   Targeting Advanced Democracies in the UPR

As part of its efforts to reform the UNCHR in 2006, the UN also introduced the Universal Periodic Review (UPR), which went into effect in 2008. Under the UPR, *all* UN member states, rather than simply ones identified by members of the UNHRC, are subject to periodic reviews of their domestic human rights practices. Specifically, all states are reviewed every four and a half years, and these review sessions take place three times per year in Geneva. In these sessions, three randomly assigned states lead the discussion of the state under review. Reviews are based on information provided by the state under review, independent human rights commissions and experts, and national human rights or non-governmental groups. The reviews are conducted by the 47 sitting members of the UNHRC, but any state can be involved in the discussion. Reviewer states identify issues in the state under review and make suggestions for how to address them. The reviewed state can either accept or reject these recommendations (Cox, 2010). Unlike most functions in the UNHRC, which are subject to regional affiliations and North-South conflicts, the UPR process emphasizes bilateral state-to-state relations (McMahon and Ascherio, 2012). We take advantage of this feature of the

---

[15]The opposite signs on voting 'no' and 'abstaining' is mechanical since states can only vote once. As Figure 2 above illustrates, what's happening since 2012 is that states are moving away from voting "yes". Backsliding states in particular are now more likely to vote "no". See also Appendix 4.

UPR to observe more directly the individual behavior of backsliding states.

Data on the content of these reviews are publicly available online.[16] We use these to create a dataset that consists of 32,598 individual recommendations made for all UN member states from 2008 through 2020. This dataset also includes information on the reviewing state. Similar to the UNHRC (Figure 1), we find the average democracy score for states writing UPR recommendations has also declined substantially over time: while in 2008 the average democracy score for recommending states in the UPR was 0.71, by 2020 that number had fallen by more than 10% to 0.59.[17] Reflecting this trend, as Figure 5 illustrates, the percent of UPR reviewers that meet our criteria for cases of backsliding has increased steadily over time from 3.7% in 2008 to 9.8% in 2020.
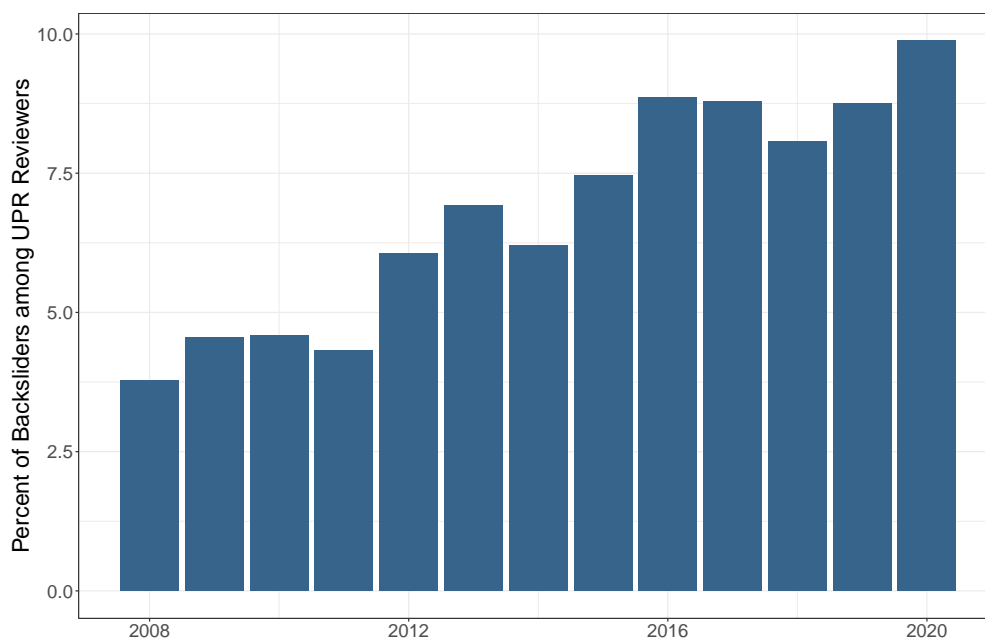

Figure 5: The percent of UPR reviewers that qualify as cases of backsliding has steadily increased over time.

We use these UPR data to further explore how backsliding states use UN human rights tools differently than their more stable, non-backsliding counterparts. Since all states are subject to UPR reviews, we begin by examining the extent to which these backsliding states

[16]https://upr-info-database.uwazi.io/en/. Last accessed 9-March-2023.
[17]See Figure A2 in Appendix 5.

use their UN membership to confront western liberal states, namely, by testing if they use these UPR reports to place more scrutiny and pressure on consolidated western liberal democracies regarding their own human rights practices.

We focus on a subset of our UPR data. Flores and Nooruddin (2016*a*, 85) define developing democracies as states "for whom a democratic system was not a certainty in 1946 or in the year of its birth as a sovereign country, whichever came second." This excludes —or in other words, designates as advanced, consolidated democracies— the following 19 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States. Since we are interested in how backsliding and non-liberal democratic states use the UPR tool to pressure and scrutinize consolidated western democracies, we restrict our data to include only observations where the state under review is one of these 19 advanced democracies and the recommending state is *not* one of these 19 countries.

To test Hypothesis 2, our main dependent variable is a count of the number of issues a reviewing state identifies in its UPR report. There is a finite (though substantially long) list of issues that can be highlighted in these reports.[18] The number of issues in any given report in our subset of the data ranges from 0 to 19, with a mean of 3.49 issues identified per report. The number of issues identified in a report is a plausible proxy for how critical that report is; more critical reports will identify more issues in the state under review.

As before, our main independent variable is a binary variable indicating whether the reviewing state was a democracy in 1995, and, if so, whether its level of democracy has deteriorated by 0.1 or more since 1995 for each year in our sample.[19] Furthermore, we use

---

[18]There are over 60 types of issues in our dataset, and they include topics such as: "death penalty," "elections," "extrajudicial executions," "freedom of the press," "labour rights," "minority rights," "trafficking," and "women's rights."

[19]It is important to note that since all UN member states, rather than just members of the UNHRC, can write UPR reviews, the subset of cases of backsliding in these models is larger than in the UNHRC models, which only includes states that were members of the Council between 2006 and 2020. The following states meet our operationalization of backsliding and wrote at least one UPR review of an advanced democracy between 2008 and 2021: Bangladesh, Belarus, Benin, Bolivia, Brazil, Ecuador, Fiji, Honduras, Hungary,

the same control variables as those reported in Table 1, with several exceptions. First, we no longer control for whether the state under review is Israel, since only reviews of the 19 democracies listed above are included in our models. We also control for whether or not the recommending state is also under review by the UNHRC in that same year to account for any incentives this may create.[20]

Since the count of issues identified in the UPR reports is under-dispersed,[21] we employ a quasi-Poisson model as our primary specification. Furthermore, to account for potential selection bias, we follow Terman and Byun (2022) and estimate an additional model that includes only observations where the recommending state was concurrently a member of the UNHRC. Since only oral comments are recorded in UPR reports, states with more representatives in Geneva —including UNHRC member states that are already present and expected to participate in UPR sessions— may be over-represented amongst UPR recommending states. However, there is no reason to expect the types of recommendations made by UNHRC members to be substantively distinct from those of other states, all else equal. Each of these models include year fixed effects. Since states can submit multiple reviews, the observations in our data are not independent. Therefore, the models also include standard errors clustered by recommending state. The results are reported in Table 2.[22] We find evidence that backsliding states identify a greater number of issues in reports targeting advanced western democracies. These results suggest that backsliding states, in addition to being more likely to vote against targeted resolutions in the UNHRC, also use the UPR mechanism deliberately to challenge the established liberal international order by placing western democracies under greater scrutiny for their human rights practices.

India, Mali, Mauritius, Nicaragua, Philippines, Poland, Sri Lanka, Turkey, Ukraine, Venezuela, and Zambia.

[20]Summary statistics for these data are reported in Appendix 6.

[21]The dispersion parameters for models 1 and 2 in Table 2 are 0.62 and 0.57, respectively.

[22]Alternative models that use the continuous measure of backsliding discussed in Section 4.1 return similar results. See Appendix 7.

Table 2: UPR Issues Identified against Advanced Democracies, 2008–2020

	Dependent variable:	
	Count of issues identified in country report	
	All Reports	Reports by UNHRC Members
	(1)	(2)
Backslide since 1995	0.16***	0.27***
	(0.02)	(0.03)
Current EDI score	0.04*	0.30***
	(0.03)	(0.04)
State under review EDI score	−0.45	0.03
	(0.31)	(0.49)
Same region	0.24***	0.31***
	(0.02)	(0.03)
Recommend UNGA US agree	−0.57***	−0.66***
	(0.05)	(0.08)
State under review UNGA US agree	0.61***	0.43***
	(0.10)	(0.15)
State under review HR score	0.02***	0.02**
	(0.01)	(0.01)
Recommending state under review	−0.07***	−0.15***
	(0.01)	(0.02)
Year fixed effects	Yes	Yes
Clustered standard errors	Yes	Yes
Model type	Quasi-Poisson	Quasi-Poisson
Observations	6,301	2,654

Note: *p<0.1; **p<0.05; ***p<0.01

4.3   The Content of UPR Recommendations

We have showed thus far that backsliding states exhibit distinct behaviors in the UN human rights institutions: they are more likely to vote "no" on targeted UNHRC resolutions and they identify a greater number of human rights issues in their UPR reviews of advanced western liberal democracies. Having established these distinct patterns of behavior, as an additional test of our theory, we examine the extent to which the substance of backsliding states' input into these human rights mechanisms reflects their changing and often increasingly illiberal domestic interests that directly contrast with the western liberal norms underlying the existing international order.

Human rights norms and topics vary with respect to how sensitive and threatening they

are to a state under review for its human rights practices. Typically, the more sensitive and threatening norms are ones that in some way directly challenge the power or legitimacy of the target regime. For example, governments condemned for violating the physical integrity rights of their citizens face potential international sanctions, such as loss of foreign aid, as well as challenges to their domestic legitimacy and support (Lebovic and Voeten, 2009; Krain, 2012). Similarly, civil and political liberties are typically viewed as particularly sensitive human rights since promoting them could pose a challenge to incumbents by opening space for political opposition (Linz and Stepan, 1996; Haggard and Kaufman, 2016).

Other types of human rights enforcement are less costly in that they do not pose a direct challenge to the target government. First, socioeconomic rights that emphasize human development in the form of improved health care, poverty reduction, and educational access, pose little threat to regime stability and, in fact, most regimes have an interest in promoting these second generation human rights (Teets, 2014). Similarly, women and children's rights —such as protections against domestic violence and human trafficking— are typically viewed as "safe" norms since they do not invoke abuses committed by the state (Terman and Byun, 2022). We expect that backsliding states, which are by definition in a transitional state and thus seeking to consolidate their power, will increasingly seek to avoid challenges to their rule in part by de-emphasizing threatening norms in their reviews of other states, focusing instead on less sensitive human rights topics.[23]

To test this, we return to our UPR dataset. In addition to the number and type of issues identified and actions recommended in these reports, our dataset includes information on the recommendations a reviewing state makes to the state under review. Unlike the data on issues, these recommendations are free-form text responses. These allow us to examine the content and topics that backsliding states choose to emphasize in their UPR reports.

We draw on the recommendations contained in UPR reports to estimate a structural topic model (STM) with 6 topics.[24] The `stm` package generates common words associated

---

[23]See Winzen (2023) for a related argument in the EU context.

[24]To do so, we use the `stm` package in `R` to tune the number of topics and estimate parameters. We select

with each topic identified by the model. Using these common words in combination with sample excerpts from the UPR review documents, we manually label each of the 6 topics. These topics and their corresponding key words are shown in Figure 6.[25]

Topic	Representative words
1. Human development	Education, poverty, rural, quality, empower, health-care, strengthen, improve
2. Ratify international human rights conventions	Convention, international, protocol, ratify, inhumane, cruel, death, degrade
3. Establish national human rights institutions	Nation, human, right, protect, institution, implement, bodies, worker
4. Civil and political liberties	Ensure, freedom, right, law, independent, journalist, transparency, press, assembly, express
5 Women and LGBTQ+ rights	Violence, women, sexual, pregnancy, domestic, abortion, marriage, incest, orientation
6. Protect minorities and vulnerable groups	Combat, discrimination, children, trafficking, roma, racism, xenophobia, intolerance, hate, racial

Figure 6: This figure illustrates the labels we assign to each topic in our model and the most common words appearing in each STM topic.

In our STM model, we explore the types of issues backsliding states emphasize in their UPR reports. When estimating the model, we include the same set of independent variables as those included in Table 2. The results of this model can be found in Figures 7 and 8.[26]

---

the number of topics that maximizes semantic coherence and exclusivity (6 topics, see Appendix Figure A3). We pre-process all text by eliminating common English stopwords, numbers, and punctuation. We also lowercase all text.

[25]See Appendix 9 for representative excerpts from each topic.

[26]Confidence intervals in both figures are at the 95% level.
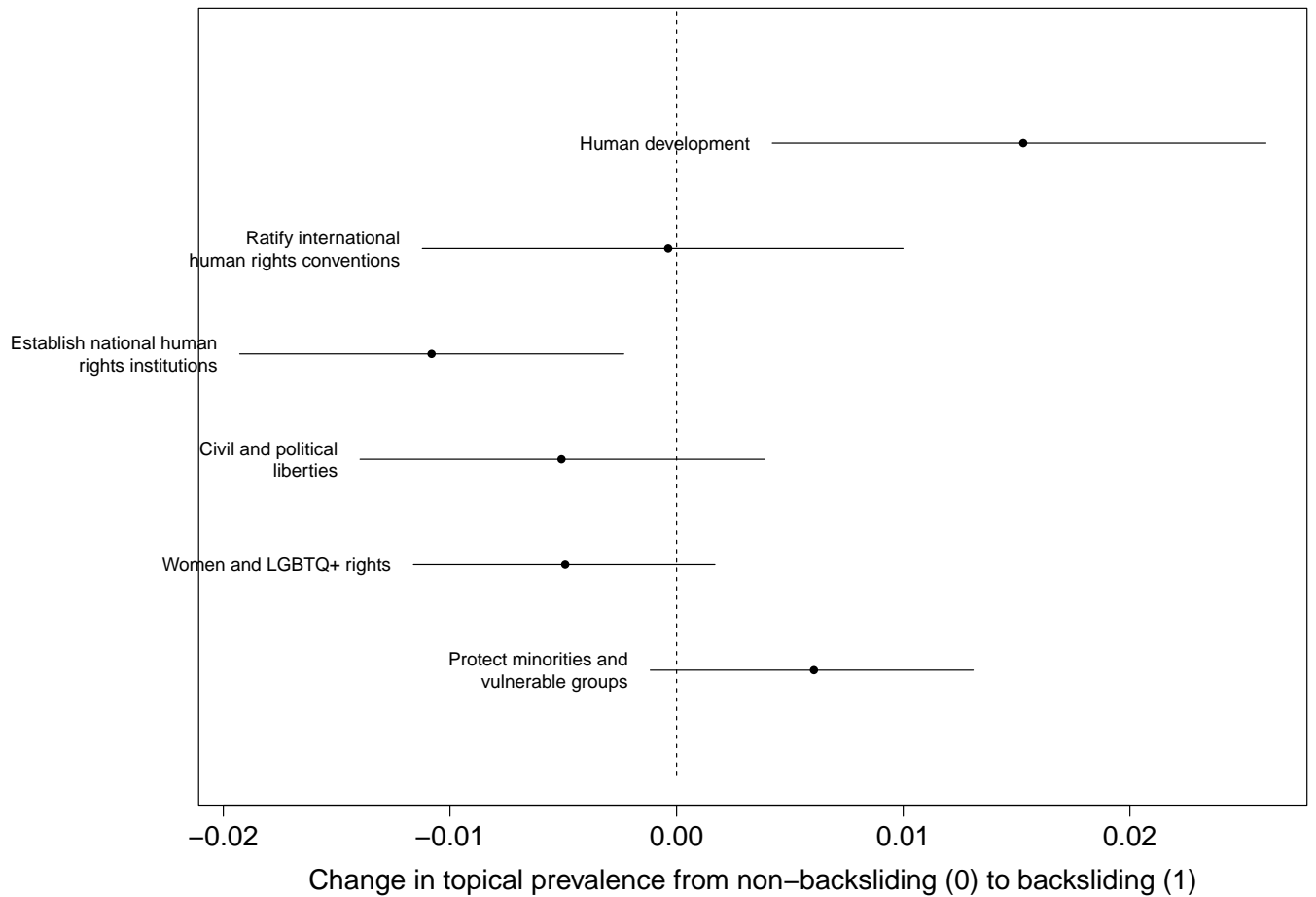
Figure 7: Structural Topic Model Results (Binary Indicator of Backsliding). Substantively, the effect size corresponds to the change in the proportion of the text relevant to a given topic when our indicator for backsliding changes from 0 (not backsliding) to 1 (backsliding).

Figure 7 plots the change in topic prevalence as our binary indicator for democratic backsliding changes from 0, which indicates a stable regime, to 1, which indicates a case of backsliding. The results in Figure 7 are suggestive of additional support for our theory: backsliding states are significantly more likely to emphasize human development than their non-backsliding counterparts. They are also significantly less likely to advocate establishing national human rights institutions. However, the confidence intervals for these topics are relatively large, and the results for the other four topics identified in the model are not statistically significant. This is likely due to the fact that the absolute number of backsliding states represented among the universe of UPR recommending states is smaller by construction. For a different perspective, therefore, we explore the change in topic prevalence as a recommending state's level of democracy declines along the electoral democracy index rather than requiring that they meet our higher threshold of democratic backsliding; these results are more precisely estimated and are reported in Figure 8.

As Figure 8 shows, as a reviewing state's level of democracy declines, it becomes significantly more likely to focus on non-threatening human rights topics. While the positive and large coefficient on human development is particularly noteworthy, we also find that these states are more likely to reference issues related to protecting minorities and vulnerable groups and efforts to establish national human rights institutions. This emphasis on national, rather than international human rights institutions, may reflect concerns related to sovereignty and growing aversion to international influence, both of which have been characteristics of contemporary backsliding regimes and their often populist leaders (Haggard and Kaufman, 2021).

At the same time, Figure 8 shows these states are less likely to emphasize issues that pose a threat to government control and stability. Most notably, as states' levels of democracy decline, they are significantly less likely to advocate civil and political liberties often associated with western liberalism, such as freedom of expression, religion, and association, as well as rights for political opposition groups that could serve as a challenge to the ruling
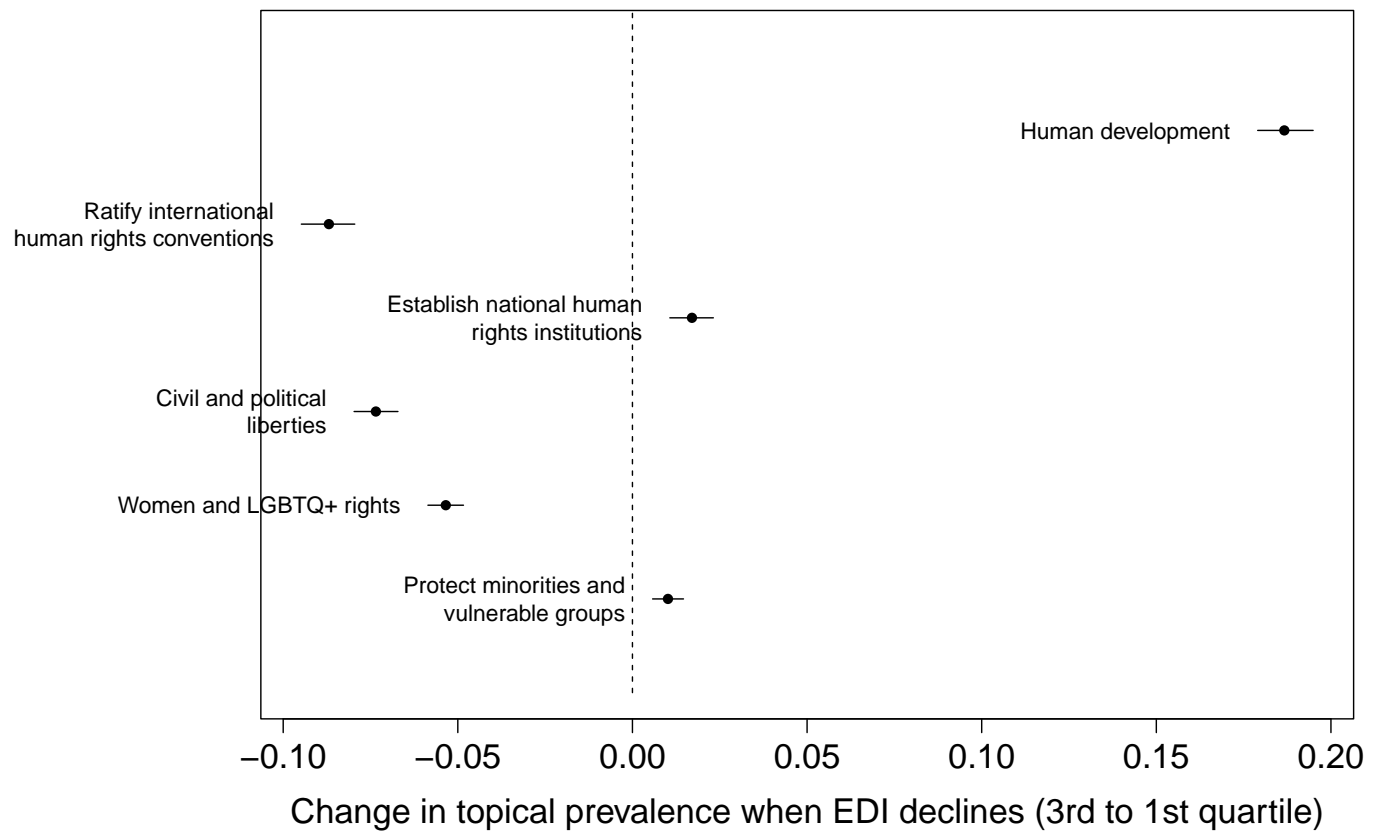
Figure 8: Structural Topic Model Results (Continuous Regime Measure). Substantively, the effect size corresponds to the change in the proportion of the text relevant to a given topic when a recommending state's electoral democracy index declines from the 3rd quartile (0.87) to the 1st quartile (0.33).

regime. These states are also significantly less likely to support ratifying international human rights conventions, which again may reflect issues related to sovereignty, and are less likely to advocate the protection of gender-based and LGBTQ+ rights. The latter is likely related to the fact that many contemporary cases of backsliding occur under the leadership of far-right, populist politicians who advocate returns to more traditional cultural and family values (Mudde and Rovira Kaltwasser, 2017). Taken together, these results provide further suggestive evidence that backsliding states are using international human rights fora to promote their own interests and to counter efforts by democracies to spread liberal democratic norms and values.

5 Conclusion

In this paper, we present evidence that backsliding states' representation in the United Nations' human rights organizations has increased over time, and find that they exhibit behavior distinct from that of their democratic counterparts. First, we show that backsliding states are more likely to vote against targeted UNHRC resolutions. We also draw on data from the Universal Periodic Review to show how these states are more actively using the UN human rights institutions to challenge western liberal democracies by identifying a larger number of issues in their UPR reports on these countries. Furthermore, the types of recommendations they make in all of their UPR reports illustrate how these states use existing international institutions to promote policies and norms more in line with their domestic environments and preferences. Specifically, we show that backsliding states are more likely to emphasize issues such as human development, while de-emphasizing topics that might threaten their power and control over citizens, such as civil and political liberties, in their UPR reports.

Taken together, these findings have several implications for the functioning and efficacy of the UN's efforts at promoting human rights, and suggest the possibility that similar dynamics could be at play in other western, liberal IOs whose established members are

similarly backsliding. First, our findings regarding voting behavior in the UNHRC suggest recent global backsliding has impacted the efficacy of this institution, most concretely by decreasing the unanimity of UNHRC resolutions (Figure 3 above). Resolutions with less unanimous support not only risk diminishing the amount of pressure target states feel to implement reforms, but at a more fundamental level also serve to further undermine the legitimacy of the UNHRC and the values it was created to uphold. Our findings illustrate the rise of illiberalism in IOs and its consequence for international cooperation, which is the focus of the proposed special issue of which this manuscript is a part.

The ways in which we show that these backsliding states strategically leverage recommendations in their UPR reports are perhaps even more troubling. The evidence suggests these backsliding states are not only using their voting rights to decrease the efficacy and legitimacy of existing institutions, but are also more actively working to undermine existing fundamental conceptions of human rights that are centered on civil and political liberties, and replacing them with an alternative set of softer, economic rights —such as human development— that better align with their domestic, illiberal interests. To the extent that these efforts succeed, the past three decades of progress made in expanding individual freedoms could potentially be undone.

We identify several areas for additional research. First, future studies should explore the extent to which these dynamics we identify within the UN human rights institutions are at play in other western liberal IOs. To what extent, for example, have states such as Hungary and Poland been able to leverage their membership privileges to undermine political and liberal aspects of the European Union and related European institutions from within? How, if at all, are these IOs working to counteract actions being taken by their now backsliding member states? Furthermore, additional research is needed to evaluate whether backsliding states' actions within existing liberal IOs are having the desired effects. This requires a better understanding of who these backsliding states view as their intended audiences when they employ these tactics, as well as the effect these efforts have on these audiences. From

whom do these backsliding states seek legitimacy? Do these backsliding states face any costs for their subversive behavior?

The implications of our findings should concern any supporters of the LIO agenda on human rights. The absence of any credible global leadership on these issues, coupled with the growing presence of autocracies and problematic democracies within the very IOs charged with protecting rights, spells the end of an era in which human rights were a meaningful topic of discussion internationally. Of particular concern is the inability of Western states to tackle democratic backsliding within their own borders and in their strategic partners globally. At least at this moment, it appears that human rights IOs are simply theater, but where the once-upon-a-time understudies are increasingly seizing center stage, and changing the script while they are at it.

References

Alizada, Nazifa, Vanessa A Boese, Martin Lundstedt, Kelly Morrison, Natalia Natsika, Yuko Sato, Hugo Tai, Staffan I Lindberg, Lisa Gastaldi, Sandra Grahn, Garry Hindle, Nina Ilchenko, Josefine Pernes and Johannes von Römer. 2022. "Democracy Report 2022: Autocratization Changing Nature?".

Bailey, Michael A., Anton Strezhnev and Erik Voeten. 2017. "Estimating dynamic state preferences from United Nations voting data." *Journal of Conflict Resolution* 61:430–456.

Barnett, Michael and Martha Finnemore. 2021. *The Power of Liberal International Organizations.* Johns Hopkins University Press.

Baturo, Alexander. 2023. "Autocrats and the Moral High Ground: Russia and the Politics of Anti-Nazism in International Organizations.". Working paper.

Bermeo, Nancy. 2016. "On Democratic Backsliding." *Journal of Democracy* 27(1):5–19.

Binder, Martin and Autumn Lockwood Payton. 2022. "With Frenemies Like These: Rising Power Voting Behavior in the UN General Assembly." *British Journal of Political Science* 52(1):381–398.

Börzel, Tanja A. and Michael Zürn. 2021. Contestations of the Liberal International Order: From Liberal Multilateralism to Postnational Liberalism. Vol. 75 Cambridge University Press pp. 282–305.

Bush, Sarah, Christina Cottiero and Lauren Prather. 2023. "Zombies Ahead: Explaining the Rise of Low-Quality Election Monitoring.". Working paper.

Carothers, Thomas. 2002. "The End of the Transition Paradigm." *Journal of Democracy* 13:5–21.

Carraro, Valentina. 2017. "The united nations treaty bodies and universal periodic review: Advancing human rights by preventing politicization?".

Carraro, Valentina, Thomas Conzelmann and Hortense Jongen. 2019. "Fears of peers? Explaining peer and public shaming in global governance." *Cooperation and Conflict* 54:335–355.

Charter of the United Nations. 1945.

Collier, David and Steven Levitsky. 1997. "Democracy with Adjectives: Conceptual Innovation in Comparative Research." *World Politics* 49:430–451.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Nazifa Alizada, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Allen Hicken, Garry Hindle, Nina Ilchenko, Joshua Krusell, Anna L uhrmann, Seraphine F. Maerz, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Pamela Paxton, Daniel Pemstein, Josefine Pernes,

Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Ei tan Tzelgov, Yi ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2021. "V-Dem Country-Year/Country-Date Dataset v11.".

Cottiero, Christina and Stephan Haggard. 2021. "The Rise of Authoritarian Regional International Organizations." *Working Paper, UC Institute on Global Conflict and Cooperation* .

Cox, Eric. 2010. "State Interests and the Creation and Functioning of the United Nations Human Rights Council." *Journal of International Law and International Relations* 6:87–120.

Dahl, Robert. 1971. *Polyarchy: participation and opposition.* Yale University Press.

Debre, Maria J. 2021. "The dark side of regionalism: how regional organizations help authoritarian regimes to boost survival." *Democratization* 28:394–413.

Debre, Maria J. and Lee Morgenbesser. 2017. "Out of the shadows: autocratic regimes, election observation and legitimation." *Contemporary Politics* 23:328–347.

Donno, Daniela. 2010. "Who Is Punished? Regional Intergovernmental Organizations and the Enforcement of Democratic Norms." *International Organization* 64:593–625.

Donno, Daniela. 2013. *Defending Democratic Norms: International Actors and the Politics of Electoral Misconduct.* Oxford: Oxford University Press.

Donno, Daniela, Kelly Morrison, Burcu Savun and Perisa Davutoglu. 2023. "Competing Verdicts: Multiple Election Monitors and Post-Election Contention.". Working paper.

Edwards, Martin S., Kevin M. Scott, Susan Hannah Allen and Kate Irvin. 2008. "Sins of commission? Understanding membership patterns on the United Nations human rights commission." *Political Research Quarterly* 61:390–402.

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

Flores, Thomas Edward and Irfan Nooruddin. 2016a. *Elections in hard times: Building stronger democracies in the 21st century.* Cambridge University Press.

Flores, Thomas Edward and Irfan Nooruddin. 2022. "Why incumbents perpetrate election violence during civil war." *Conflict Management and Peace Science* 0(0).

Flores, Thomas and Irfan Nooruddin. 2016b. *Elections in Hard Times: Building Stronger Democracies in the 21st Century.* Cambridge: Cambridge University Press.

Fukuyama, Francis. 1989. "The End of History?" *The National Interest* (16):3–18.

Genna, Gaspare M. and Taeko Hiroi. 2014. *Regional Integration and Democratic Conditionality: How Democracy Clauses Help Democratic Consolidation and Deepening.* Florence, GB: Routledge.

Ginsburg, Tom. 2020. "Authoritarian international law?" *American Journal of International Law* 114:221–260.

Hafner-Burton, Emilie, Jon Pevehouse and Christina Schneider. 2023. "Enlightened Autocrats? Good Governance in Autocratic International Organizations.". Working paper.

Hafner-Burton, Emilie M. 2005. "Trading Human Rights: How Preferential Trade Agreements Influence Government Repression." *International Organization* 59:593–629.

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

Hafner-Burton, Emilie M. 2012. "International regimes for human rights." *Annual Review of Political Science* 15:265–286.

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.

Haggard, Stephan and Robert Kaufman. 2021. *Backsliding: Democratic Regress in the Contemporary World.* Cambridge University Press.

Haggard, Stephan and Robert R Kaufman. 2016. "Democratization During the Third Wave." *Annual Review of Political Sciencel* 19(8):1–20.

Herre, Bastian and Max Roser. 2016. "Human Rights." *Our World in Data* . https://ourworldindata.org/human-rights.

Hug, Simon and Richard Lukács. 2014. "Preferences or blocs? Voting in the United Nations Human Rights Council." *Review of International Organizations* 9:83–106.

Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century.* Norman: University of Oklahoma Press.

Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became An International Norm.* Cornell University Press.

Ikenberry, G. John. 2009. "Liberal Internationalism 3.0: America and the Dilemmas of Liberal World Order." *Perspectives on Politics* 7:71–87.

Ikenberry, G. John. 2018. "The end of liberal international order?" *International Affairs* 94:7–23.

Jee, Haemin, Hans Lueders and Rachel Myrick. 2022. "Towards a unified approach to research on democratic backsliding." *Democratization* 29:754–767.

Johnston, Alastair Iain. 2007. *Social States: China in International Institutions, 1980-2000.* Princeton University Press. Johnston, Alastair Iain. Social States: China in International Institutions, 1980-2000. Princeton, NJ: Princeton University Press, 2007.

Johnstone, Ian and Michael Snyder. 2016. Democracy promotion. In *The Oxford handbook of international organizations.*

Kelemen, R. Daniel. 2020. "The European Union's authoritarian equilibrium." *Journal of European Public Policy* 27:481–499.

Kim, Yooneui. 2023. "A network analysis of naming and shaming in the universal periodic review." *International Interactions* 0(0):1–28.

Kneuer, Marianne, Thomas Demmelhuber, Raphael Peresson and Tobias Zumbrägel. 2019. "Playing the regional card: why and how authoritarian gravity centres exploit regional organisations." *Third World Quarterly* 40:451–470.

Koliev, Faradj. 2020. "Shaming and democracy: Explaining inter-state shaming in international organizations." *International Political Science Review* 41:538–553.

Krain, Matthew. 2012. "J'accuse! Does Naming and Shaming Perpetrators Reduce the Severity of Genocides or Politicides?" *International Studies Quarterly* 56:574–589.

Lakatos, István. 2022. The United Nations Human Rights Council: A Platform for the Clash of Civilizations or a Forum of Cross-Regional Cooperation? In *Comparative Human Rights Diplomacy.* Springer International Publishing pp. 335–363.

Landman, Todd. 2005. "The Political Science of Human Rights." *British Journal of Political Science* 35(3):549–572.

Lebovic, James H. and Erik Voeten. 2006. "The politics of shame: The condemnation of country human rights practices in the UNCHR." *International Studies Quarterly* 50:861–888.

Lebovic, James H. and Erik Voeten. 2009. "The cost of shame: International organizations and foreign Aid in the punishing of human rights violators." *Journal of Peace Research* 46:79–97.

Levitsky, Steven and Lucan A. Way. 2010. *Competitive Authoritarianism: Hybrid Regimes After the Cold War.* Cambridge University Press.

Libman, Alexander and Anastassia V. Obydenkova. 2018. "Regional international organizations as a strategy of autocracy: The Eurasian Economic Union and Russian foreign policy." *International Affairs* 94:1037–1058.

Linz, Juan and Alfred Stepan. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe.* Baltimore: Johns Hopkins University Press.

Lipps, Jana and Marc S. Jacob. 2023. "Undermining Liberal International Organizations from Within: Evidence from the Parliamentary Assembly of the Council of Europe.". Working paper.

Luhrmann, Anna and Staffan I. Lindberg. 2019. "A third wave of autocratization is here: what is new about it?" *Democratization* 26:1095–1113.

McMahon, Edward and Marta Ascherio. 2012. "A Step Ahead in Promoting Human Rights? The Universal Periodic Review of the UN Human Rights Council." *Global Governance* 18:231–248.

Meyerrose, Anna M. 2018. "It is all about value: How domestic party brands influence voting patterns in the European Parliament." *Governance* 31:625–642.

Meyerrose, Anna M. 2020. "The Unintended Consequences of Democracy Promotion: International Organizations and Democratic Backsliding." *Comparative Political Studies* 53.

Meyerrose, Anna M. 2021. International Sources of Democratic Backsliding. In *Routledge Handbook of Illiberalism*, ed. András Sajó, Renáta Uitz and Stephen Holmes. Routledge.

Meyerrose, Anna M., Thomas Edward Flores and Irfan Nooruddin. 2019. From Elections to Democracy in Hard Times. In *Oxford Research Encyclopedia of Politics*. Oxford University Press.

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

Mudde, Cas and Cristóbal Rovira Kaltwasser. 2017. *Populism: A Very Short Introduction.* Oxford University Press.

Neumayer, Eric. 2005. "Do International Human Rights Treaties Improve Respect for Human Rights?" *Journal of Conflict Resolution* 49(6):925–953.

Pahre, Robert David. 1995. Wider and deeper: The links between expansion and integration in the European Union. In *Towards a new Europe: Stops and starts in regional integration.* Praeger pp. 111–136.

Pevehouse, Jon. 2005. *Democracy from above : regional organizations and democratization.* Cambridge University Press.

Poast, Paul and Johannes Urpelainen. 2018. *Organizing democracy: How international organizations assist new democracies.* University of Chicago Press.

Risse, Thomas and Kathryn Sikkink. 1999. *The Socialization of International Human rights Norms into Domestic Practices: Introduction.* Cambridge University Press.

Seligman, Steven. 2011. "Politics and principle at the UN human rights commission and council (1992-2008)." *Israel Affairs* 17:520–541.

Smith, Alastair. 2016. "Leader Turnover, Institutions, and Voting at the UN General Assembly." *Journal of Conflict Resolution* 60(1):143–163.

Teets, Jessica C. 2014. *Civil Society under Authoritarianism: The China Model.* Cambridge University Press.

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

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

Tsutsui, Kiyoteru and Christine Min Wotipka. 2004. "Global Civil Society and the International Human Rights Movement: Citizen Participation in Human Rights International Nongovernmental Organizations." *Social Forces* 83(2):587–620.

von Soest, Christian. 2015. "Democracy prevention: The international collaboration of authoritarian regimes." *European Journal of Political Research* 54:623–638.

Voss, M. Joel. 2019. "The Use (or Misuse) of Amendments to Contest Human Rights Norms at the UN Human Rights Council." *Human Rights Review* 20:397–422.

Vreeland, James Raymond. 2008. "Political Institutions and Human Rights: Why Dictatorships Enter into the United Nations Convention Against Torture." *International Organization* 62(1):65–101.

Vreeland, James Raymond. 2019. "Corrupting International Organizations." *Annual Review of Political Science* 22:205–222.

Waldner, David and Ellen Lust. 2018. "Unwelcome Change: Coming to Terms with Democratic Backsliding." *Annual Review of Political Science* 21(5):1–21.

Walker, Christopher. 2016. "The Authoritarian Threat: The Hijacking of "Soft Power"." *Journal of Democracy* 27:49–63.

Way, Lucan Ahmad. 2022. "The Rebirth of the Liberal World Order?" *Journal of Democracy* 33:5–17.

Weyland, Kurt. 2017. "Autocratic diffusion and cooperation: the impact of interests vs. ideology." *Democratization* 24:1235–1252.

Winzen, Thomas. 2023. "Walking the tightrope: How backsliding governments keep the EU hospitable for autocracy.". Working paper.

Woo, Byungwon and Amanda Murdie. 2017. "International organizations and naming and shaming: does the International Monetary Fund care about the human rights reputation of its client?" *Political Studies* 65(4):767–785.

Ziblatt, Daniel and Steven Levitsky. 2018. *How Democracies Die.* New York: Crown.

Trojan Horse Autocracies, International Organizations, and the Liberal International Order

Online Appendices

Contents

1 Operationalizing Democratic Backsliding

As discussed in the manuscript, we operationalize backsliding as any country-year observation for which a UNHRC member state's polyarchy index from the V-Dem dataset has declined by 0.1 (e.g. 10%) or more since 1995. Our main independent variable is coded as 1 for all country-year observations where the voting state's polyarchy score was above 0.5 in 1995 and has decreased by 0.1 or more since 1995; otherwise, it takes a value of 0. With this coding, we identify 14 UNHRC member states between 2006 and 2021 as cases of backsliding.

While our rationale for using 1995 as a baseline to identify these cases of backsliding members of the UNHRC is related to the end of the third wave of democracy and the start of the subsequent wave of autocratization, there is no single year that clearly marks the end of the former and beginning of the latter. To explore the effect of using 1995 as our baseline year, in Figure A1 we show the democratic trajectory of the states we identify as cases of backsliding among UNHRC members, finding that overall these states' levels of democracy increased in the 10 years prior to 1995. Furthermore, with a few exceptions, this states experienced steady, monotonic declines in their levels of democracy sometime after 1995.

Furthermore,in Table A1 we show what our universe of cases of backsliding members of the UNHRC would be if we set the baseline years as 2000 and 2006, the year in which the UNHRC was founded. Here, we find that most of the cases of backsliding we identify with the 1995 baseline also are coded as backsliding with these alternative baseline years. This gives us further confidence in our choice to use 1995 as the baseline.

Figure A1: Democratic trajectories of backsliding members of the UNHRC, 1985–2021

Table A1: Universe of backsliding states in UNHRC with different baseline years

1995 baseline	2000 baseline	2006 baseline
Ecuador	Ecuador	Ecuador
Nicaragua	Nicaragua	Nicaragua
Brazil	Brazil	Brazil
Hungary	Hungary	Hungary
India	India	India
Poland	Poland	Poland
Bolivia	Bolivia	Bolivia
Ukraine		Ukraine
Philippines	Philippines	
Bangladesh	Bangladesh	
Venezuela	Venezuela	
Sri Lanka	Sri Lanka	
Madagascar	Madagascar	
Fiji		
	Indonesia	Indonesia
	South Korea	South Korea
	Bulgaria	
		Macedonia

2 UNHRC Model: Descriptive Statistics

Table A2: Descriptive Statistics: Voting in the UNHRC, 2006–2021

Statistic	N	Mean	St. Dev.	Min	Max
Vote choice	13,721	2.68	0.60	1	3
Backslide since 1995	13,721	0.07	0.25	0	1
Current EDI score	13,626	0.57	0.27	0.02	0.92
Target state = Israel	13,721	0.27	0.44	0	1
Target EDI score	13,721	0.38	0.25	0.07	0.76
Voting and target in same region	13,626	0.07	0.25	0	1
Voting UNGA US agree	12,716	0.38	0.15	0.00	0.81
Target UNGA US agree	12,879	0.44	0.29	0.04	0.96
Target state HR score	11,993	-1.27	0.69	-2.30	1.04

3 UNHRC Voting: Autocratization

We define backsliding as a process unique to democracies in that the starting point for any case of backsliding is within a democratic state. However, the literature recently has focused more broadly on autocratization, which is defined as any move away from democracy in both democracies *and* autocracies (Luhrmann and Lindberg, 2019). When we operationalize backsliding as a 10% or greater shift away from their 1995 EDI score in any state, we find similar results, as reported in Table A3.

Table A3: Autocratization and vote choice in the UNHRC, 2006–2021

	Dependent variable:	
	Abstain	No
	(1)	(2)
Autocratize since 1995	−0.13	0.75***
	(0.10)	(0.11)
Current EDI score	−0.35***	−2.55***
	(0.14)	(0.21)
Target state = Israel	−0.11	0.05
	(0.19)	(0.31)
Target EDI score	1.82***	2.12***
	(0.32)	(0.48)
Voting and target in same region	−0.90***	−0.63***
	(0.13)	(0.16)
Voting UNGA US agree	−1.09***	1.10***
	(0.26)	(0.41)
Target UNGA US agree	−1.49***	−2.79***
	(0.35)	(0.59)
Target state HR score	0.25***	0.05
	(0.04)	(0.07)
Year fixed effects	Yes	Yes
Clustered standard errors	Yes	Yes
Model type	Multinomial logit	Multinomial logit
Observations	11,760	11,760
K	3	3
Akaike Inf. Crit.	13,190.25	13,190.25

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

In Table A3 we find evidence that autocratizing states are more likely to vote against targeted UNHRC resolutions. Specifically, we find that, for a state that has autocratized since 1995, its log-odds of voting "no" relative to "yes" increases by 0.75. To facilitate interpretation, we also calculate the relative risk of a backsliding state voting "no" rather than "yes" by exponentiating the reported coefficient: a backsliding state is roughly 2.1 times more likely to vote "no" as opposed to "yes" on a targeted UNHRC resolution. The negative coefficient for the current democracy score variable in the "no" column of Table A3

also suggests that as a state's level of democracy decreases, it is more likely to vote "no," rather than "yes," on targeted UNHRC resolutions.

4 UNHRC Voting: Continuous Measure of Backsliding

While the main models in the manuscript operationalize backsliding as a binary variable, we re-estimated the same models using a continuous measure of backsliding in Table A4. In this model the main independent variable captures the change in a state's level of electoral democracy since 1995, with negative changes indicating a state has backslid since 1995. Our results are robust to this alternative operationalization: states that have backslid since 1995 are less likely to abstain, relative to voting yes, but are more likely to vote "no," rather than "yes."

Table A4: Democratic backsliding and vote choice in the UNHRC, 2006–2021, Continuous Measure of Democratic Backsliding

	Dependent variable:	
	Abstain	No
	(1)	(2)
EDI score change since 1995	0.63***	−2.67***
	(0.18)	(0.28)
Current EDI score	−0.46***	−2.19***
	(0.14)	(0.21)
Target state = Israel	−0.11	0.04
	(0.19)	(0.32)
Target EDI score	1.81***	2.15***
	(0.32)	(0.49)
Same region	−0.93***	−0.56***
	(0.13)	(0.16)
Voting UNGA US agree	−0.87***	0.47
	(0.27)	(0.42)
Target UNGA US agree	−1.49***	−2.82***
	(0.35)	(0.59)
Target state HR score	0.26***	0.05
	(0.04)	(0.07)
Year fixed effects	Yes	Yes
Clustered standard errors	Yes	Yes
Model type	Multinomial logit	Multinomial logit
Observations	11,760	11,760
K	3	3
Akaike Inf. Crit.	13,116.49	13,116.49

Note:	*p<0.1; **p<0.05; ***p<0.01

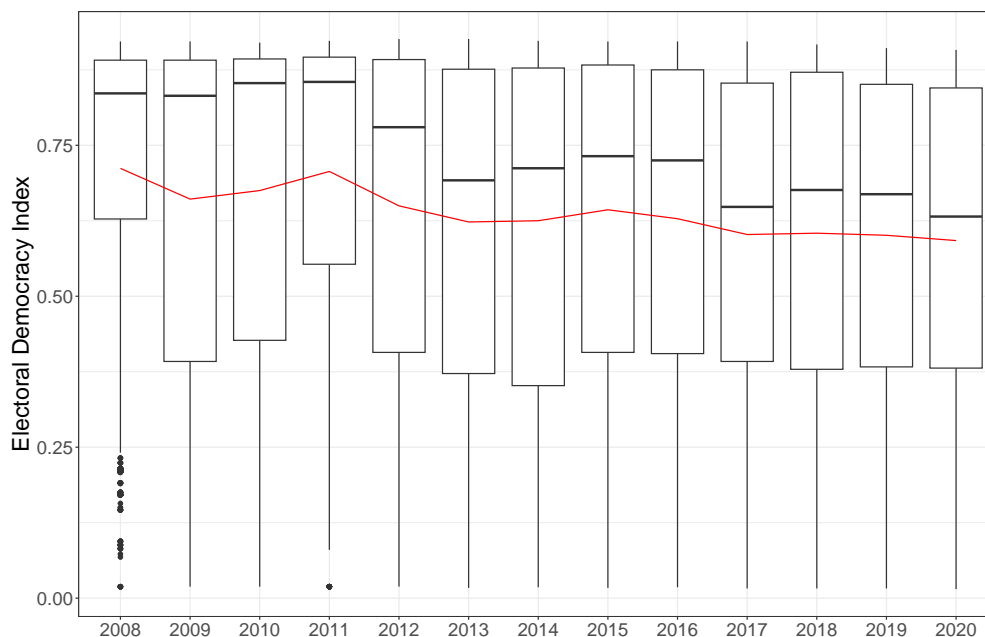# 5   Democracy Levels of UPR Recommending States

Figure A2: The distribution of electoral democracy index scores of UPR recommending states by year. The red line shows the average democracy score by year for all states that wrote UPR reviews in that year. Data source: Varieties of Democracy dataset (Coppedge et al., 2021).

6   UPR Issues: Descriptive Statistics

Table A5: Descriptive Statistics: UPR Reports of Advanced Democracies

Statistic	N	Mean	St. Dev.	Min	Max
Number of issues per review	7,777	3.49	1.99	0	19
Backslide since 1995	7,777	0.12	0.33	0	1
Current EDI score	7,645	0.52	0.25	0.01	0.92
State under review EDI score	7,777	0.88	0.03	0.81	0.92
Same region	7,777	0.10	0.30	0	1
Recommend UNGA US agree	7,710	0.38	0.16	0.00	0.94
State under review UNGA US agree	6,956	0.61	0.09	0.44	0.85
State under review HR score	6,931	2.86	1.12	0.31	5.33
Recommending state under review	7,777	0.19	0.39	0	1

7 UPR Issues: Continuous Measure of Backsliding

As in Table A4, in the models in Table A6 the main independent variable is a continuous one that captures the change in a state's level of electoral democracy since 1995, with negative changes indicating a backsliding state. Our results are robust to this alternative operationalization of backsliding.

Table A6: UPR Issues Identified against Advanced Democracies, 2008–2020, Continuous Measure of Democratic Backsliding

	Dependent variable:	
	Count of issues identified in country report	
	All Reports	Reports by UNHRC Members
	(1)	(2)
Demo score change since 1995	−0.46***	−0.50***
	(0.03)	(0.04)
Current EDI score	0.11***	0.39***
	(0.03)	(0.04)
State under review EDI score	−0.54*	−0.03
	(0.31)	(0.50)
Same region	0.22***	0.29***
	(0.02)	(0.03)
Recommend UNGA US agree	−0.64***	−0.90***
	(0.05)	(0.09)
State under review UNGA US agree	0.58***	0.44***
	(0.10)	(0.15)
State under review HR score	0.02***	0.02*
	(0.01)	(0.01)
Recommending state under review	−0.07***	−0.15***
	(0.01)	(0.02)
Year fixed effects	Yes	Yes
Clustered standard errors	Yes	Yes
Model type	Quasi-Poisson	Quasi-Poisson
Observations	6,256	2,639

Note: *p<0.1; **p<0.05; ***p<0.01
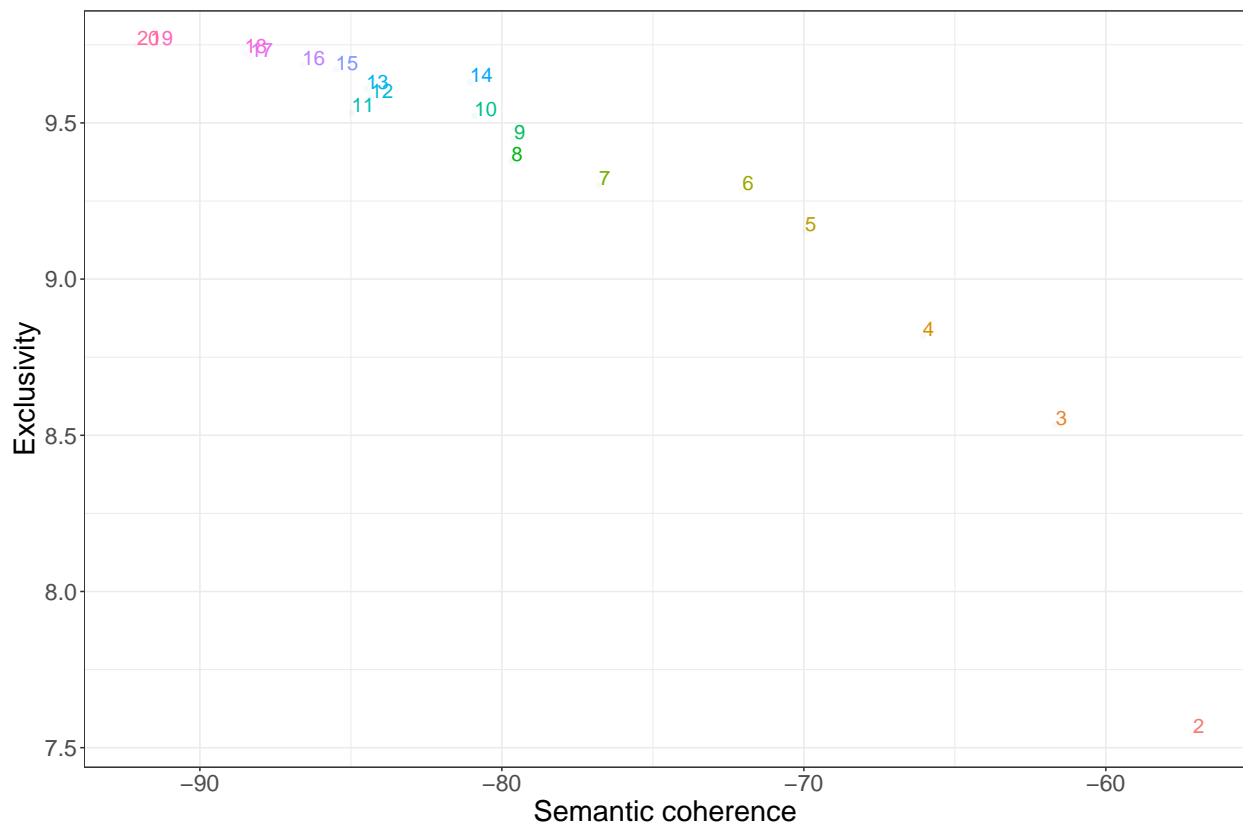
8   STM Diagnostic Plot

Figure A3: **STM Topic Number Selection.** This plot shows semantic coherence and exclusivity for STM models fitted to 2–20 topics on our UPR report corpus. We select 6 topics to jointly maximize the two statistics.

9   STM Sample Excerpts

Below we include representative sample excerpts from each of the 6 topics we identify in the STM model estimated using our UPR dataset.

Topic 1: Human Development:

- "Continue efforts to improve the quality of education, especially in rural and mountainous regions, Continue efforts to implement Sustainable Development Goal 9 on building resilient infrastructure, promoting inclusive and sustainable industrialization and fostering innovation in order to secure and promote access to and the enjoyment of economic development progress for everyone"

- "Continue to support children's enrolment in school to develop skills in various fields through sound education in order to help poor and other vulnerable groups, including in rural society, to improve their skills and have better access to economic opportunities

and basic services,Continue its efforts to further enhance gender equality and women's empowerment in both public and private administration"

- "Consolidate the implementation of the Growth and Transformation Plan 2016–2020, which is decisive in fighting poverty and social exclusion,Continue to strengthen the Productive Safety Net Programme, which supports about 8 million people affected by a chronic food shortage"

Topic 2: Ratify international human rights conventions

- "Ratify the Convention on the Rights of Persons with Disabilities and its Optional Protocol, Ratify without reservations the International Convention for the Protection of All Persons from Enforced Disappearance,Ratify without reservations the Agreement on Privileges and Immunities of the International Criminal Court,Ratify the Optional Protocol to the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment,Ratify the Second Optional Protocol to the International Covenant on Civil and Political Rights with a view to abolishing the death penalty"

- "Accede to the Rome Statute of the International Criminal Court and to the Agreement on Privileges and Immunities of the Court,Enforce the International Covenant on Civil and Political Rights fully and accede to its Second Optional Protocol,Ratify the Optional Protocol to the Convention on the Rights of the Child on the involvement of children in armed conflict,Ratify the Optional Protocol to the Convention against Torture"

- "Extend a moratorium on the death penalty and abolish the death penalty, and ratify the Second Optional Protocol to ICCPR, Ratify the Rome Statute of the International Criminal Court,Ratify the Convention on the Elimination of All Forms of Discrimination against Women, the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment and its Optional Protocol"

Topic 3: Establish national human rights institutions

- "Consider the establishment of a national human rights institution accredited by the International Coordinating Committee of the National Institutions for the Promotion and Protection of Human Rights"

- "Consider developing Human Rights Indicators as suggested by OHCHR as an instrument that allows for a more precise and coherent evaluation of national human rights policies,Conclude, as a matter of priority, the process leading to the establishment of a national human rights institution, with a broad human rights mandate"

- "Recommended that a national commission on human rights be established in accordance with the Paris Principles in fulfilment of its voluntary pledges."

Topic 4: Civil and political liberties

- "Observe constitutional guarantees of freedom of religion and ensure religious groups are not subject to discrimination"

- "Put an end to all practices that threaten the right to freedom of expression, including threats against journalists and media outlets"

- "Release political detainees, cease arbitrary detentions and take other measures to ensure that those arrested and detained are given the right to a fair trial"

- "Create the conditions for genuine, multiparty democracy in the country and for free and fair elections to be held"

Topic 5: Women and LGBTQ+ rights

- "Adopt legal provisions specifically criminalizing female genital mutilation, breast ironing and discriminatory widowhood rites"

- "Accelerate the adoption of a general law on violence against women and girls criminalizing all forms of violence, including domestic violence and marital rape"

- "Ensure availability of safe abortions, at a minimum in cases where the pregnancy is the result of rape or incest and in cases of severe and fatal foetal impairment"

- "Decriminalize sexual activity between consenting adults of the same sex, and address hate crimes on the grounds of sexual orientation and gender identity"

Topic 6: Protect minorities and vulnerable groups

- "Adopt effective measures to improve conditions in prisons, reduce overcrowding and eliminate ill- treatment and forced labour of persons deprived of liberty"

- "Take effective measures to combat and punish hate speech"

- "Take more effective measures in order to fight, and even to eradicate, crimes and hate speech against minorities"

- "Effectively combat discrimination against refugees, migrants and ethnic minority in order to ensure their rights"