

Human Rights Backsliding

Andrew T. Guzman^{*} & Katerina Linos^{**}
Forthcoming, California Law Review

ABSTRACT^{***}

Human rights advocates and international lawyers view international agreements and other international norms as important tools to improve human rights around the world. This Article explains that, contrary to widely held beliefs, international human rights norms are not a one-way street. Norms capable of generating improved behavior in poorly performing states will also tend to exert a downward pull on high-performing states. This downward pull leads to what we term “human rights backsliding”—a tendency for high-performing states to weaken their domestic human rights regimes relative to prior behavior or relative to what they would otherwise have done.

The theory of backsliding is a novel one, and so we introduce it with several real-world examples. In order to make the theory, its assumptions, and its consequences as explicit as possible, we also provide a formal model of backsliding. We then explain how an understanding of human rights backsliding helps explain state behavior that is otherwise puzzling. We explore some of the implications of backsliding for the design of international agreements and we consider strategies for advocates seeking to advance the cause of human rights internationally.

^{*} Jackson H. Ralston Professor of Law and Associate Dean for International and Advanced Degree Programs, Berkeley Law School. Harvard Law School, J.D.; Harvard University, Ph.D.

^{**} Assistant Professor of Law, Berkeley Law School. Harvard Law School, J.D.; Harvard University, Ph.D.

^{***} We are grateful to Roxanna Altholz, Robert Bartlett, Anu Bradford, David Caron, Robert Cooter, Justin McCrary, Jeffrey Dunoff, Stavros Gadinis, David Gartner, Jack Goldsmith, Ryan Goodman, Larry Helfer, Saira Mohamed, Mark Pollack, Alison Post, Greg Shaffer, David Sklansky, David Sloss, Mila Versteeg and participants at the 2012 Berkeley Law Faculty Retreat, the Northern California International Law Workshop, the Junior International Law Scholars Association Meeting, the Cornell Law faculty workshop, and the Temple International Law Workshop for their helpful comments. Sylvia DeTar, Helena Jung, Jerome Hsiang and May Whitaker provided excellent research assistance.

Table of Contents

Introduction	2
Part I. A Puzzle: How Does Backsliding Happen?	9
Part II. Three Examples of Backsliding	13
A. Defendants' Rights in the United Kingdom.....	13
B. Same Sex Marriage in the United Kingdom and the United States.....	15
C. Women's Rights in Sweden	18
Part III. The Model	22
A. International Norm Higher than Domestic Equilibrium ($k > S^*$)	25
B. International Norm Lower than Domestic Equilibrium ($k < S^*$).....	29
Part IV. Extending the Theory of Backsliding	34
A. Domestic Factors Influencing Backsliding	34
B. How the Content and Form of the Norm Matters.....	36
1. Floors, Ceilings, and Prohibitions	36
2. Hard Law versus Soft Law.....	39
3. Multiple Standards	39
4. Vague Norms and Uncertainty	42
C. Other Theories of International Law's Influence	42
Part V. Implications	44
A. Some Empirical Puzzles Explained	44
1. The Behavior of High Performing States	44
2. Regionalism	46
3. Aspirational Norms.....	48
B. Treaty Membership	49
Part VI. Conclusion	52

Introduction

Human rights practices have improved dramatically in many parts of the world over the last century, in no small part because of the tireless efforts of transnational advocacy groups using international legal

instruments.¹ A recent award-winning book by Beth Simmons reports that after ratifying international human rights treaties, governments in many countries detain and torture fewer people,² allow for greater religious freedom,³ reduce child labor rates,⁴ and improve women's access to contraception.⁵ However, this same research notes that these effects are concentrated among moderately democratic countries and countries transitioning towards democracy.⁶ Treaty ratification often does not have these salutary effects in stable democracies.⁷ Most importantly for this Article, ratification of international treaties is sometimes correlated with worse practices by such states.⁸ These findings invite us to think carefully about the varied effects human rights treaties may have.

In this article, we question an implicit assumption held by many human rights scholars and advocates: that international human rights standards are a one-way street that can only lead states to expand

¹ See generally BETH A. SIMMONS, *MOBILIZING FOR HUMAN RIGHTS: INTERNATIONAL LAW IN DOMESTIC POLITICS* (2009) [hereinafter SIMMONS, *MOBILIZING*]; Ryan Goodman & Derek Jinks, *How to Influence States: Socialization and International Human Rights Law*, 54 DUKE L.J. 983 (2004); David Gartner, *Transnational Rights Enforcement*, BERKELEY J. INT'L L. (forthcoming 2013), available at <http://law.asu.edu/LinkClick.aspx?fileticket=tl-Vp0ejZiM%3D&tabid=2927>; Katerina Linos, *Diffusion through Democracy*, 55 AM. J. POL. SCI. 678 (2011) [hereinafter Linos, *Diffusion*]; KATERINA LINOS, *THE DEMOCRATIC FOUNDATIONS OF POLICY DIFFUSION: HOW HEALTH, FAMILY AND EMPLOYMENT LAWS SPREAD ACROSS COUNTRIES*, (Oxford University Press) (forthcoming 2013) [hereinafter LINOS, *DEMOCRATIC FOUNDATIONS*]; MARGARET E. KECK & KATHRYN SIKKINK, *ACTIVISTS BEYOND BORDERS: ADVOCACY NETWORKS IN INTERNATIONAL POLITICS* (1998) [hereinafter KECK & SIKKINK, *ACTIVISTS BEYOND BORDERS*]; Margaret E. Keck & Kathryn Sikkink, *Transnational Advocacy Networks in International and Regional Politics*, 51 INT'L SOC. SCI. J. 89 (1999) [hereinafter Keck & Sikkink, *Transnational Advocacy Networks*].

² See SIMMONS, *MOBILIZING*, *supra* note 1, at 305.

³ *Id.* at 200.

⁴ *Id.* at 328.

⁵ *Id.* at 254.

⁶ *Id.* at 200, 328, 254, 305.

⁷ *Id.* at 200, 328. See also Emilie M. Hafner-Burton & Kiyoteru Tsutsui, *Human Rights in a Globalizing World: The Paradox of Empty Promises*, 110 AM. J. SOC. 1373, 1374-78 (2005) (suggesting that countries that ratify human rights agreements often do not improve their behavior); Oona A. Hathaway, *Do Human Rights Treaties Make a Difference?*, 111 YALE L.J. 1935, 1940-41 (2002) (same).

⁸ See SIMMONS, *MOBILIZING*, *supra* note 1, at 281-82 (suggesting that among high-rule-of-law states, countries that have ratified the Convention Against Torture report a higher prevalence of torture than countries that have not ratified the Convention); Eric Neumayer, *Do International Human Rights Treaties Improve Respect for Human Rights?*, 49 J. CONFLICT RESOL. 925, 928 (2005) (exploring severe compliance challenges); Linda Camp Keith, *The United Nations International Covenant on Civil and Political Rights: Does It Make a Difference in Human Rights Behavior?* 36 J. OF PEACE RESEARCH 95, 95 (1999) (finding that formal acceptance of the Covenant and optional protocol had no observable impact on state behavior).

domestic protections.⁹ Though international law and international norms can, indeed, be useful tools to improve human rights performance in poorly performing states, we argue that they can also undermine efforts to adopt or maintain high levels of protection in countries that would otherwise offer protections above the international norm. We call this phenomenon “human rights backsliding.”

We define human rights backsliding as a process in which governments react to international standards by providing fewer or weaker human rights protections. Our definition includes the withdrawal of previously available rights as well as stasis or stagnation where we would otherwise observe an expansion of rights (although it may be more difficult to clearly establish causation in the latter situation).

The notion of backsliding is novel and so there is some burden on us not only to explain that it is possible as a matter of theory (which we do), but also to show that it is plausible in fact. To this end, consider the observation made by Professors Zachary Elkins, Thomas Ginsburg, and Beth Simmons regarding the Universal Declaration of Human Rights (UDHR).¹⁰ Following the adoption of the UDHR, some rights began to appear more frequently in national constitutions, including the right to life, the prohibition on ex post facto punishment, the right to join a trade union, the presumption of innocence in trial, the right to free movement, and the prohibition on cruel and inhuman treatment.¹¹ All of this is consistent with conventional views on how international human rights law can lead to the diffusion and expansion of rights domestically. Less consistent is the fact that certain other rights “fell out of fashion especially quickly due to their exclusion from the UDHR.”¹² The authors note, for example, that the right to a jury trial, the prohibition on censorship, the right to petition, certain intellectual property rights, prohibitions on child employment and the right to a free press were far less common in national constitutions after the UDHR was adopted than we would otherwise expect.¹³ These rights not only failed to advance following the UDHR, they became *less* popular than they would have been absent the Declaration. This is an example of human rights backsliding at work.

⁹ For related criticisms, see Eran Shor, *Conflict, Terrorism, and the Socialization of Human Rights Norms: The Spiral Model Revisited*, 55 SOCIAL PROBLEMS 117, 118 (2008) (critiquing prominent human rights scholars for their assumption that “once states adopt the rhetoric of human rights and begin to move toward norm compliance, there is no turning back . . . [they] move forward uniformly towards norm compliance, or alternatively remain stagnant.”). See also David Rieff, *The Precarious Triumph of Human Rights*, N.Y. TIMES MAG., Aug. 8, 1999, at 37. (stating that “the human rights movement has assumed that establishing norms will lead to a better world”).

¹⁰ See Zachary Elkins, Tom Ginsburg & Beth A. Simmons, *Getting to Rights: Treaty Ratification, Constitutional Convergence and Human Rights Practice*, 54 HARV. INT’L L.J. 201, 220-21 (2013).

¹¹ *Id.* at 10.

¹² *Id.*

¹³ *Id.*

In the pages that follow, we discuss additional examples and focus on specific rights in particular countries to clarify the mechanism at work.¹⁴ We describe how Britain reduced the scope of criminal defendants' rights to exclude hearsay evidence, influenced in part by decisions from the European Court of Human Rights (ECtHR) that failed to include the common law hearsay rule among the minimum protections required in all European countries.¹⁵ We also show how opponents of same-sex marriage in Britain used ECtHR decisions in their lobbying efforts. The ECtHR held that the European Convention does not require states to legalize same-sex marriage, and conservative groups used these decisions to fight proposals to legalize same-sex marriage in Britain.¹⁶ We explain how Sweden responded to a European directive setting minimum standards relating to maternity leave benefits by limiting the choices and benefits available to women. Swedish feminists, joined by advocates for children, the elderly, and the disabled have expressed a concern that the European Union (EU) represents a threat to Sweden's generous welfare state. They fear human rights backsliding in Sweden.¹⁷

Developing a coherent theory of backsliding requires that we think seriously about how international norms affect domestic practices and requires that we make some assumptions about how those norms are transmitted to domestic policy makers. For most of this Article, we adopt a theory of domestic politics that is consistent with prevailing perspectives on how international human rights norms come to affect domestic policies.

According to major strands of the human rights literature, international standards can be effective because they focus attention on particular issues and place these issues on national agendas.¹⁸ In addition, international standards can influence the views of domestic publics, interest groups, and decision-makers.¹⁹ When an international organization promotes a certain policy proposal as an

¹⁴ See discussion *infra* Part II.

¹⁵ See discussion *infra* Part II.A.

¹⁶ See discussion *infra* Part II.B.

¹⁷ See discussion *infra* Part II.C.

¹⁸ See SIMMONS, MOBILIZING, *supra* note 1, at 128-29.

¹⁹ See SIMMONS, MOBILIZING, *supra* note 1, at 14-15; Goodman & Jinks, *supra* note 1, at 654-55; Keck & Sikkink, *Transnational Advocacy Networks*, *supra* note 1, at 89-90; Thomas Risse & Kathryn Sikkink, *The Socialization of International Human Rights Norms into Domestic Practices: Introduction*, in *THE POWER OF HUMAN RIGHTS: INTERNATIONAL NORMS AND DOMESTIC CHANGE* 1, 4-5, 16-17 (Thomas Risse, Stephen C. Ropp & Kathryn Sikkink eds., 1999). National standards may also serve to formalize emerging international human rights norms and correct perception biases regarding actual human rights behavior in a country. It may be that states are more likely to conform to prevailing norms when they have accurate information regarding other states' human rights behavior. See Robert Cooter, Michal Feldman & Yuval Feldman, *The Misperception of Norms: The Psychology of Bias and the Economics of Equilibrium*, 4 REV. L. & ECON. 889, 891 (2008) for a discussion on the psychology of misperceiving a norm and its effects on incentives to comply with that norm.

international human right, this endorsement from a credible and disinterested outsider makes the proposal seem less radical and strengthens the rhetorical position of advocates for the position.²⁰ Rights that previously seemed undesirable or unattainable begin to appear feasible.²¹ At the margin, this “nudge” from an international institution can persuade citizens to support the policy, potentially tipping the balance in favor of the introduction of this right into the national legal system.²²

At the core of our argument is the observation that, under reasonable assumptions, this same mechanism can also operate to lower human rights’ protections in high performing states—to trigger human rights backsliding.

Protecting human rights inevitably involves some form of trade-off. Granting criminal defendants more rights may limit victims’ rights; wider access to food and water may require higher taxes; increased protection of free speech may require a relaxation of hate speech codes. Domestic interest groups and political leaders line up on both sides of such debates. Statements from an international body can serve as a thumb on the scale and strengthen the position of both conservative and progressive groups. In a country debating an expansion of domestic human rights protection above the international standard, opponents of the expansion can point to the international standard and argue that it reflects the right balance between diverse concerns. Protecting human rights any further, they argue, would represent a radical and ill-thought-out experiment, compromise competing values (such as economic growth, law and order, national security, liberty, or any number of other priorities) and make their country worse off. Thus, a relatively low international standard arms opponents of the expansion with an apparently neutral external benchmark, and strengthens the persuasiveness of their arguments vis-à-vis undecided citizens in just the same way that international norms can lead to the expansion of rights in low-performing countries.²³

The above theory of human rights, in which domestic constituencies are critical to the transmission of international standards, is an important and conventional one, but it is not the only one. Toward the end of the Article we consider other familiar theories of human rights transmission from international norm

²⁰ See Xinyuan Dai, *Why Comply? The Domestic Constituency Mechanism*, 59 INT’L ORG. 363, 363, 366, 384-85 (2005); Linos, *Diffusion*, *supra* note 1, at 679-80, 692; LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, at 13-14, 28.

²¹ See LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, 96-98.

²² *Id* at 22-23.

²³ We do not claim that increases in human rights performance in low-performing countries and backsliding in high-performing countries are of the same magnitude. Our theory does, however, have a certain symmetry inasmuch as influences that increase (or decrease) the impact of international norms on low performers will also increase (or decrease) the impact on high performers, and vice versa.

to domestic policy and demonstrate that backsliding can result from any of these established mechanisms as well.²⁴

If we are correct that international human rights instruments not only expand human rights protections in poorly performing states, but also limit human rights practices among top performers, major implications follow for the design of international human rights regimes. There is some (imperfect) evidence that drafters of major human rights instruments are worried about the possibility of backsliding. Many international agreements clearly specify that they set minimum standards and that countries are free to set much higher goals. For example, the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) concludes by stating that: “Nothing in the present Convention shall affect any provisions that are more conducive to the achievement of equality between men and women which may be contained [in other national and international instruments].”²⁵ Both the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR) use substantively similar language to set forth that there shall be no restriction in national protections “on the pretext that the present Covenant does not recognize such rights or that it recognizes them to a lesser extent.”²⁶ Similarly, the European Convention on Human Rights concludes by specifying that: “Nothing in this Convention shall be construed as limiting or derogating from any of the human rights and fundamental freedoms which may be ensured under [other national and international instruments].”²⁷ The American Convention on Human Rights highlights that it should not be read to “preclud[e] other rights or guarantees that are inherent in the human personality” or “restric[t] the enjoyment or exercise of any right or freedom recognized by virtue of the laws of any State Party.”²⁸ Indeed, this type of language is common in many treaties, and can be found both in general provisions like the ones just mentioned²⁹ and in the specification of the scope of

²⁴ See discussion *infra* Part IV.C.

²⁵ Convention on the Elimination of All Forms of Discrimination against Women art. 23, Dec. 18, 1979, 1249 U.N.T.S. 13; 19 I.L.M. 33 (1980).

²⁶ International Covenant on Civil and Political Rights art. 5(2), Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976); See International Covenant on Economic, Social and Cultural Rights art. 5(2), Dec. 16, 1966, 993 U.N.T.S. 3 (1967) (stating that: “No restriction upon or derogation from any of the fundamental human rights recognized or existing in any country in virtue of law, conventions, regulations or custom shall be admitted on the pretext that the present Covenant does not recognize such rights or that it recognizes them to a lesser extent.”).

²⁷ Convention for the Protection of Human Rights and Fundamental Freedoms art. 53, Nov. 4, 1950, 213 U.N.T.S. 221, § 6(3)(d) [hereinafter ECHR].

²⁸ American Convention on Human Rights art. 29, Nov. 22, 1969, O.A.S.T.S. No. 36, 1144 U.N.T.S. 123 (entered into force July 18, 1978).

²⁹ See also, e.g., International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families art. 81, G.A. res. 45/158, annex, 45 U.N. GAOR Supp. (No. 49A) at 262, U.N. Doc. A/45/49 (1990) (entered into force 1 July 2003) (declaring that “[n]othing in the present

particular rights.³⁰ Such articles indicating that international agreements are floors and not ceilings are more than a simple statement of the legal reality. The wording of the obligations themselves makes it clear that states are free to provide protections in excess of those in the agreement. The main function of the additional language is to provide emphasis, something that would not be relevant if it was just a matter of making the legal requirements of the agreement clear. It is useful language, however, if the drafters are concerned that domestic political actors will attempt to use the agreement to argue against higher standards. In other words, it is useful as an attempt to guard against backsliding.

While these drafting strategies are helpful, we do not believe that they are enough to fully prevent backsliding. The fact that a norm is expressed as a floor might be central to litigation surrounding human rights instruments, but it would not prevent backsliding as a result of the political mechanisms we consider. This is so because backsliding is not mandated by the formal legal rules. It is, instead, an unintended consequence of those rules. In the discussion that follows we consider this drafting question in detail along with related issues concerning the formal legal status of the norm, the possibility of a norm with multiple standards at different levels, and the consequences of vague norms.³¹

Understanding the risk of backsliding helps explain several patterns in international human rights law that have puzzled other theorists. One such puzzle is the reluctance of top-performing states to join

Convention shall affect more favourable rights or freedoms granted to migrant workers and members of their families by virtue of: a) The law or practice of a State Party; or b) Any bilateral or multilateral treaty in force for the State Party concerned.”).

³⁰ See *e.g.*, International Covenant on Civil and Political Rights art. 14(3), Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976) (outlining protections states must afford criminal defendants and explicitly providing that these are “minimum guarantees”); International Covenant on Economic, Social and Cultural Rights art. 7(a), Dec. 16, 1966, 993 U.N.T.S. 3 (1967) (describing how states must define, at a minimum, fair remuneration for workers); Optional Protocol to the Convention on the Rights of the Child on the Sale of Children, Child Prostitution and Child Pornography art. 3, G.A. Res. 54/263, Annex II, 54 U.N. GAOR Supp. (No. 49) at 6, U.N. Doc. A/54/49, Vol. III (2000) (entered into force Jan. 18, 2002) (calling on states to criminalize, at minimum, a list of offenses against children); American Convention on Human Rights, *supra* note 28, at art. 8 (guaranteeing minimum rights guarantees for the criminally accused); Convention for the Protection of Human Rights and Fundamental Freedoms, *supra* note 27, at art. 6 (specifically setting forth a list of minimum rights applicable to the accused); International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families, *supra* note 29, at art. 8 (articulating a set of minimum rights guarantees extended to migrant workers and family members when accused of crimes); International Convention for the Protection of All Persons from Enforced Disappearance, G.A. res. 61/177, U.N. Doc. A/RES/61/177 (2006) (entered into force Dec. 23, 2010) (detailing the minimum scope of information that states must produce on persons deprived of liberty).

³¹ See discussion *infra* Part IV.B.

international treaties whose standards they already meet or exceed. Another puzzle concerns the proliferation of regional human rights regimes. Regional regimes are surprising because human rights derive a good deal of their moral force from claims to universality—from the notion that all human beings have fundamental rights by virtue of being human, not by virtue of residing in particular parts of the world. Finally, the risk of backsliding can help explain why human rights standards are often set at extremely high, and even unrealistic, levels, unlike other international treaty commitments that only call for modest reforms.³²

In sum, our contribution is threefold. First, we introduce the important possibility of backsliding to debates in international human rights law and challenge the implicit assumption that international agreements that specify floors serve only to improve human rights performance. We introduce our theoretical contribution in Part I and present some examples and illustrations of backsliding in Part II. Second, we develop a simple formal model to clearly explain the mechanisms through which backsliding happens and the circumstances in which it is most likely to occur. Part III of our paper presents this formal model, while Part IV extends this model to discuss what changes when we vary some of the model's assumptions. The main contribution of the formal model is to clarify the logic of our argument and show precisely where we depart from existing theoretical writings. Third, we explain in some depth how the possibility of backsliding can explain major puzzles in international human rights law and outline important practical implications for states, NGOs, and others involved in the design of human rights instruments. These implications are presented in Part V. Part VI concludes.

Part I. A Puzzle: How Does Backsliding Happen?

We begin by noting that recent evidence supports the view that human rights treaties can influence state behavior for the better, at least in some circumstances, and lead to an expansion of domestic protections.³³ We highlight this point because our theory assumes that human rights instruments can matter. Theories suggesting that human rights treaties have very little influence on state behavior rule out the possibility of backsliding because they exclude both positive and negative influences from abroad.³⁴

The empirical evidence that human rights agreements can have a positive effect cries out for a theoretical explanation. The results are surprising, at least initially, because some of the most prominent theories of international law's influence do not work well in the human rights field. One such theory, institutionalism, treats states as rational and unitary actors that seek to maximize their own gains from

³² See discussion *infra* Part V.

³³ See SIMMONS, *MOBILIZING*, *supra* note 1, at 14-15.

³⁴ See Jack Goldsmith & Eric Posner, *The Limits of International Law* 108 (2006).

the international system.³⁵ While this theory is helpful in explaining what we observe in many areas of international law, including the environment, trade, security, and more, it does not work well in the human rights field. Institutionalism performs poorly in this area because its assumption of selfish states cannot easily accommodate efforts to improve the well-being of human beings in other states. It cannot explain why an improvement in human rights abroad is valuable to the state.³⁶

Furthermore, even if one accepts that states wish to influence human rights policies abroad, why would the resulting treaties be effective? The familiar incentives to comply, known as the Three Rs of Compliance, are reciprocity, retaliation, and reputation. These do not predict much compliance-pull in the human rights area.³⁷ When a state deprives its citizens of fundamental rights, other states are unlikely to reciprocate or retaliate by violating their own citizens' rights.³⁸ Moreover, a state that engages in human rights violations may be able, by continuing to comply with its trade, military, and other commitments, to develop separate reputations in each issue area, and thus continue to enjoy a strong reputation in the areas it cares about.³⁹

If institutionalism fails us, a different approach must be used. In particular, at least one of the assumptions of institutionalism must be relaxed. We choose to relax the unitary state assumption. This has the obvious appeal of eliminating one of the least plausible assumptions of the classic rational choice model. Opening the black box of the state, however poses myriad challenges. It is one thing to observe that domestic politics matters, but quite another to describe a model of domestic politics that is both realistic and tractable.

In the paragraphs that follow, we outline such a model with a focus on how government leaders respond to pressures from voters and interest groups. The claim that government leaders respond to domestic political pressure is, of course, not novel. Indeed, we believe it is the most plausible

³⁵ See ANDREW T. GUZMAN, *HOW INTERNATIONAL LAW WORKS: A RATIONAL CHOICE THEORY* 15-24 (2008) (discussing the relative advantages and disadvantages of analyzing international law through the lens of institutionalism).

³⁶ See *id.* at 33.

³⁷ See *id.*; see also Andrew T. Guzman, *Formation of Human Rights Treaties, A Response to Alex Geisinger*, OPINIO JURIS (Feb. 11, 2008), <http://opiniojuris.org/2008/02/11/formation-of-human-rights-treaties-a-response-to-alex-geisinger/>; Andrew T. Guzman, *The Puzzle of Human Rights Treaties*, OPINIO JURIS (Feb. 12, 2008, 5:13 pm), <http://opiniojuris.org/2008/02/12/the-puzzle-of-human-rights-treaties/> (highlighting methodological difficulties with explaining human rights treaties).

³⁸ There are some exceptions. If a particular ethnic group controls a state, for example, it may respond to foreign conduct that harms members of the same ethnic group abroad with retaliatory actions against a minority ethnic group within its borders.

³⁹ See GUZMAN, *supra* note 35, at 100-106; Rachel Brewster, *The Limits of Reputation on Compliance*, 1 INT'L THEORY 323, 327-28 (2009).

mechanism available in the literature to explain why international human rights instruments often improve state behavior. We take this well-established theory and develop an important implication that has not been highlighted to date: the possibility of backsliding.

International agreements can influence domestic human rights debates because human rights policies involve important trade-offs and place different societal groups in opposing positions. For example, an investigation into allegations of police misconduct can please human rights advocates seeking justice and accountability, but displease police forces and hobble security efforts. A law increasing minimum labor standards can pit labor unions seeking improved working conditions against employer associations concerned about competitiveness and economic growth. Expanded access to water and sanitation can help poor communities enjoy fundamental social and economic rights, but can involve cuts in other government programs and trigger significant criticism from these programs' beneficiaries.

In making these trade-offs, politicians seek to maintain the support of the public at large.⁴⁰ Popular support is critical to democratic leaders concerned about the next election,⁴¹ and social unrest is highly problematic, even for autocrats.⁴² Politicians who take radical positions that please special interest groups alone can lose voter support. Politicians thus try to adopt relatively mainstream policies, consistent with voters' values and likely to benefit the country at large. Relatedly, both politicians and interest groups often try to present their positions in ways that appear mainstream and beneficial not only to them, but also to the public at large.

International law, international norms, and the practices of other countries can influence this effort and serve as benchmarks against which citizens evaluate government performance. In deciding between a new, untested policy, and a policy that is endorsed by international organizations and widely adopted around the world, governments may often select the latter, as it is easier to justify to domestic audiences.⁴³

For example, imagine that labor unions in a developing country seek to eliminate child labor, in order to increase wages for adult workers and reduce accidents. If international law prohibits child labor, and many neighboring developing countries have restricted the practice, labor unions will have an easier time mobilizing widespread support for their position. As Beth Simmons puts it: "local agents have the

⁴⁰ See Emilie M. Hafner-Burton, Laurence R. Helfer & Christopher Farris, *Emergency and Escape: Explaining Derogations from Human Rights Treaties*, 65 INT'L ORG. 673 (2011) (arguing that governments use derogations clauses in human rights treaties to signal to domestic audiences that the emergency rights restrictions are necessary and temporary deviations from normal levels of rights protections).

⁴¹ See LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, at pp. 1-12, 19, 30; see also Donald Wittman, *Candidate Motivation: A Synthesis*, 77 AM. POL. SCI. REV. 142, 142 (1983).

⁴² See Jessica L. Weeks, *Autocratic Audience Costs: Regime Type and Signaling Resolve*, 62 INT'L ORG., 35, 62 (2008).

⁴³ See Linos, *Diffusion*, *supra* note 1; LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, at 2-6.

motive to use whatever tools may be available and potentially effective to further rights from which they think they may benefit.”⁴⁴ International law, international norms and foreign states’ practices can be powerful tools, and are often referenced in pursuit of diverse domestic objectives.⁴⁵

Conformity with an international norm, all else being equal, gives a government greater political support at home and, in this sense, is valuable to leaders. The heart of the theory, then, is that an international signal regarding “proper” or “expected” human rights conduct will empower local interest groups, giving them a domestic political advantage and, therefore, draw local human rights policy closer to the standard specified in the international agreement or other instrument. The international focal point generates a gravitational pull on policy. This theory explains why an international agreement without strong enforcement provisions can impact human rights outcomes.⁴⁶

We demonstrate that this same mechanism also limits the scope of particular rights. Imagine that instead of banning child labor, international agreements take a more nuanced position and argue that “children’s or adolescents’ participation in work that does not affect their health and personal development or interfere with their schooling, is generally regarded as being something positive.”⁴⁷ This is in fact the International Labour Organization’s (ILO) current position.⁴⁸ In this situation, employer associations and conservative politicians can point to international norms in their efforts to allow certain types of child labor and fight labor unions’ efforts to completely ban the practice.

The argument we propose—that international norms can lead some countries to offer lower protections than they otherwise would—works best under certain conditions. First, it works best in cases in which domestic leaders are responsive to the public at large. This is not a major scope limitation for our argument, because the states that offer high levels of human rights protections, and thus could potentially be dragged down by international human rights standards, are very likely to be representative democracies. Second, we assume, as most of the human rights literature does, that states are responsive to human rights norms. Again, this assumption is particularly plausible for the

⁴⁴ SIMMONS, *MOBILIZING*, *supra* note 1, at 373.

⁴⁵ See generally KECK & SIKKINK, *ACTIVISTS BEYOND BORDERS*, *supra* note 1; Keck & Sikkink, *Transnational Advocacy Networks*, *supra* note 1; Risse & Sikkink, *supra* note 19.

⁴⁶ This argument draws most directly on the work of Katerina Linos but is consistent with arguments made by many other authors to explain why human rights instruments influence state behavior. See generally Linos, *Diffusion*, *supra* note 1; LINOS, *DEMOCRATIC FOUNDATIONS*, *supra* note 1, at 175-85; KECK & SIKKINK, *ACTIVISTS BEYOND BORDERS*, *supra* note 1; Keck & Sikkink, *Transnational Advocacy Networks*, *supra* note 1; Risse & Sikkink, *supra* note 19; Goodman & Jinks, *supra* note 1; Dai, *supra* note 20; SIMMONS, *MOBILIZING*, *supra* note 1.

⁴⁷ INTERNATIONAL LABOUR ORGANIZATION, *WHAT IS CHILD LABOUR*, <http://www.ilo.org/ipec/facts/lang--en/index.htm> (last visited Jan. 20, 2013).

⁴⁸ *Id.*

subset of countries at risk of backsliding. Democracies that offer unusually high levels of human rights protections tend to be among the most fervent supporters of international law and international norms, and are often highly integrated in transnational networks. Third, we assume that in many areas, international standards will be set at moderate levels. If international agreements instead set maximal standards that no government is likely to fully meet, this would eliminate the risk of backsliding. However, as we explain below, this would also limit the possibility of positively influencing poor performers.⁴⁹ Before presenting our formal model in Part III, we offer three extensive examples of backsliding at work.

Part II. Three Examples of Backsliding

This section provides some examples of backsliding in action. We explore how defendants' rights declined in the UK in part because of minimum European standards, how same-sex marriage opponents used European Court decisions to fight the legalization of same-sex marriage, and why women's rights advocates in Sweden saw European maternity and parental leave minimums as a threat. Note that each of these examples involves clear trade-offs. Some human rights, such as the prohibition on torture, slavery, or genocide are often perceived as absolute prohibitions. As we explain in Part IV below, prohibitions on such grave violations also involve trade-offs, as it is possible to define each of these practices broadly or narrowly. In addition, we highlight that our cases come disproportionately from Europe. As we explain in Part IV below, this is not a coincidence—the risk of backsliding is highest in democratic states that offer high levels of rights protections and are very integrated in the international community.⁵⁰ In contrast, countries that are less integrated in the international community and resist increasing rights protections as a result of international standards, such as the United States, may face lower risks of backsliding.⁵¹

A. Defendants' Rights in the United Kingdom

In the United States, the common-law rule against hearsay evidence—in conjunction with a constitutional right to confrontation—offers criminal defendants important protections by disallowing the admission of out-of-court statements from persons who are not testifying at trial.⁵² But the UK, the country that first developed the common-law hearsay rule, has drastically reduced the rights of criminal

⁴⁹ Such agreements exist, and we suggest in Part V.A.3 that concerns about backsliding may explain why they are written with such unrealistically high requirements.

⁵⁰ See discussion *infra* Part IV.A.

⁵¹ See discussion *infra* Part IV.A.

⁵² See Note, *Hearsay and Confrontation: Can the Criminal Defendant's Rights be Preserved Under a Bifurcated Standard?*, 32 WASH. & LEE L. REV. 243, 243 (1975).

defendants to question the prosecution's witnesses.⁵³ Specifically, reforms introduced in the 2003 Criminal Justice Act allow the prosecution to introduce hearsay evidence in the form of pre-trial statements from persons unavailable to testify at trial.⁵⁴

The European Convention on Human Rights sets minimum human rights standards for dozens of European states, including the UK.⁵⁵ Article 6(3)(d) of the Convention states that every person charged with a crime has the right "to examine or have examined witnesses against him."⁵⁶ However, "as construed by the [ECtHR] the opportunity to cross-examine adverse witnesses is not invariably mandated by Article 6."⁵⁷ For example, in *Isgrò v. Italy*, the ECtHR unanimously upheld Mr. Isgrò's conviction, even though Mr. Isgrò's attorney was not able to cross-examine Mr. Isgrò's primary accuser.⁵⁸

The Criminal Justice Act was largely based on a 1997 Law Commission Report titled "Evidence in Criminal Proceedings: Hearsay and Related Topics."⁵⁹ This Report offers several reasons about why The UK should limit the scope of its hearsay rule, and discusses the European Convention, and the Jurisprudence of the European Court, and the practices of continental European countries extensively.⁶⁰ The report finds it "significant that in many European countries what would be called hearsay in England and Wales is admissible and does not appear to be in contravention of the [European] Convention."⁶¹ The Commission concludes "that the Convention does not require direct supporting evidence where it is

⁵³ See David A. Sklansky, *Hearsay's Last Hurrah*, SUP. CT. REV., 1, 2 (2010) (noting that the UK, along with several other common law countries, have reduced the scope of hearsay protections).

⁵⁴ See Criminal Justice Act, 2003, c. 44, § 116 (1)-(2) (U.K.) (allowing the testimony of witnesses unavailable at trial due to death, physical or mental unfitness, absence from the United Kingdom when a return is not practical, disappearance despite reasonable efforts to locate the witness, or fear that the witness testimony will lead to recrimination); see also *id.* at 29.

⁵⁵ See Luzius Wildhaber, *The European Court of Human Rights: The Past, the Present, the Future*, 22 AM. U. INT'L L. REV. 521, 537-538 (essay by the President of the European Court of Human Rights from 1997 to 2007, explaining the role of the Court).

⁵⁶ See ECHR, *supra* note 27, at § 6(3)(d).

⁵⁷ Todd E. Pettys, *Counsel and Confrontation*, 94 MINN. L. REV. 201, 247 (2009).

⁵⁸ *Isgrò v. Italy*, 194 Eur. Ct. H.R. (ser. A) at 2 (1991).

⁵⁹ See Adrian Keane, *THE MODERN LAW OF EVIDENCE*, 7th EDITION, 270 (2008); see also *R v. Horncastle* [2009] UKSC 14 (appeal taken from [2009] EWCA Crim 964) at ¶ 29.

⁶⁰ The Law Commission, *Evidence in Criminal Proceedings: Hearsay and Related Topics*, L.C. 245, 1997, at ¶15.1- ¶15.40.

⁶¹ *Id.* at ¶15.34; see also *id.* at ¶15.7 (explaining that practices that "fall foul of the hearsay rule in England and Wales . . . would be considered unobjectionable in most Continental systems" and might not violate the European convention, "depending on all the circumstances taken together.").

⁶¹ *Id.*

sought to prove a particular element of the offence by hearsay. Adequate protection for the accused will be provided by the safeguards we propose.”⁶² In short, while The UK likely had multiple reasons to reduce the rights of criminal defendants, the fact that it was out of line with Europe appears to have been part of the decision process.⁶³

Neither the European Convention nor the resulting jurisprudence was ever intended to reduce criminal defendants’ rights (or any other rights). Rather, its goal is to secure a minimum set of shared protections across Europe.⁶⁴ Nevertheless, influenced in part by these minimum standards, The UK reduced its criminal protections. It experienced human rights backsliding.

B. Same Sex Marriage in the United Kingdom and the United States

Our definition of backsliding includes not only a withdrawal of previously available rights, but also use of international norms to limit or delay domestic efforts to expand rights. In The UK, the introduction of same-sex marriage is likely to happen soon, but it is important to note that ECtHR decisions were used to bolster opposition to this reform. A relatively low international standard armed opponents of the expansion with an apparently neutral external benchmark, and strengthened the persuasiveness of their arguments vis-à-vis undecided citizens.

Since 1981, when an ECtHR decision, *Dudgeon v. UK*,⁶⁵ decriminalized sodomy, some regional and international bodies have offered greater protection to gays and lesbians than many national legal systems. Advocates seeking to improve the status of gays and lesbians within domestic systems have responded by adopting the language of international human rights to persuade national decision-makers and national public that their country is out of step with the world.⁶⁶ This framing based on

⁶² *Id.* at ¶15.41.

⁶³ There has been a largely concurrent trend throughout the common-law world to weaken the hearsay rule. Moreover, most of these jurisdictions, including Canada, Australia, and New Zealand, are outside the formal purview of the ECtHR, suggesting that common-law hearsay reform may have occurred for a variety of reasons. Nonetheless we believe that, at least in the case of the United Kingdom, the well documented conversation between the ECtHR and both the British courts and legal commentators provides evidence that the ECtHR provided a significant “nudge” that allowed the UK to proceed with hearsay reform.

⁶⁴ See Wildhaber, *supra* note 55.

⁶⁵ *Dudgeon v. The United Kingdom*, 45 Eur. Ct. H.R. (ser. A) (1981).

⁶⁶ See Larry R. Helfer & Eric Voeten, *International Courts as Agents of Legal Change: Evidence from LGBT Rights in Europe*, INT’L ORG. (forthcoming 2013), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1850526##. Gay rights advocates point to ECtHR

international human rights has, to some extent, replaced alternative framings, such as “national civil rights conceptions” and “framings based on gay and lesbian liberation and emancipation.”⁶⁷

ECtHR decisions, however, do not always go as far as progressives might like. One consequence is that (relatively) conservative decisions can undermine efforts to expand human rights in countries that would otherwise offer protections beyond the international norm. In two recent decisions, *Schalk and Kopf v. Austria* and *Gas and Dubois v. France*, the ECtHR held that the European Convention on Human Rights does not require member states’ governments to grant same-sex couples access to marriage.⁶⁸ These decisions were influential in the UK, a country that appears headed toward legalizing the practice. The UK is a high performer in many areas of human rights. In addition, domestic support for same-sex marriage seemed widespread, as both the opposition Labour Party, and the Conservative government of David Cameron, in Coalition with the Liberal Democrats, had argued for this reform.⁶⁹ Opponents of same-sex marriage, however, have taken a page from the international human rights strategy of gay rights advocates and have used the ECtHR rulings as a sword to strengthen their arguments.⁷⁰ The Church of England, for example, emphasized the ECtHR’s rulings in explaining its opposition to same-sex marriage proposals.⁷¹ Similarly, conservative advocacy groups focused on the ECtHR rulings. For

decisions such as *Dudgeon v. UK*, to UN Human Rights Committee Decisions such as *Toonen v. Australia*, to UN Human Rights Council resolutions condemning discrimination and violence against LGBT people, and even to statements by UN Secretaries-General. *Toonen v. Australia*, Views of the Human Rights Committee, U.N. Doc CCPR/C/50/D/488/1992 (1994), available at <http://www1.umn.edu/humanrts/undocs/html/vws488.htm>; Human Rights Council Res., Human Rights Council, 17th Sess., May 30-June 17, 2011, 17/19, A/HRC/17/L.9/Rev.1 (June. 17, 2011); Ronald Holzacker, “*Gay Rights are Human Rights*” *The Framing of New Interpretations of International Human Rights Norms* 1, 13-20 (presented at APSA 2011 Annual Meeting), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1902165.

⁶⁷ Holzacker, *supra* note 66, at 1.

⁶⁸ See *Schalk and Kopf v. Austria*, App. No. 30141/04 Eur. Ct. H.R. (2010); *Gas and Dubois v. France*, App. No. 25951/07 Eur. Ct. H.R. (2012).

⁶⁹ See generally *David Cameron’s Conservative Party Conference Speech in Full*, THE GUARDIAN (Oct. 5, 2011), <http://www.guardian.co.uk/politics/2011/oct/05/david-cameron-conservative-party-speech> (statement of United Kingdom Prime Minister David Cameron) (“So I don’t support gay marriage despite being a Conservative. I support gay marriage because I’m a Conservative.”).

⁷⁰ Brief for International Jurists and Academics as Amicus Curiae Supporting Petitioner Hollingsworth and Respondent Bipartisan Legal Advisory Group at 24-26, *Hollingsworth v. Perry*, No. 12-144 (Supreme Court argued Mar. 26, 2013), *U.S. v. Windsor*, No. 12-307 (Supreme Court argued Mar. 27, 2013), 2013 WL 1780813.

⁷¹ See *A Response to the Government Equalities Office Consultation – “Equal Civil Marriage” – From the Church of England*, Church of England (2012), <http://www.churchofengland.org/media/1475149/s-%20marriage.pdf>.

example, Norman Wells from the Family Education Trust, said: “The ruling from the ECtHR will embolden those whose concerns about same-sex marriage and adoption are not inspired by personal hatred and animosity, but by a genuine concern for the well-being of children and the welfare of society.”⁷²

Indeed, both proponents and opponents of same-sex marriage in the United States are taking a page from the same playbook. International and comparative law experts recently filed a Supreme Court amicus brief for *Hollingsworth v. Perry* in support of respondents seeking marriage equality. They provided examples of how international and foreign courts have upheld fundamental notions of equal protection, liberty, and dignity, and have not contravened religious freedom.⁷³ But a less expected move, prominent conservative opponents of same-sex marriage are also now invoking international law to buttress their argument that same-sex marriage is not an international norm. For example, a number of conservative-leaning legal commentators also filed a Supreme Court amicus brief for *Hollingsworth v. Perry* as well as *United States v. Windsor*, arguing that the ECtHR “found no such consensus [regarding same-sex marriage] in Europe.”⁷⁴ Furthermore, they cite the ECtHR in *Schalk and Kopf v. Austria* to show that Article 12 of the European Convention on Human Rights “does not impose an obligation on the respondent Government to grant a same-sex couple like the applicants access to marriage.”⁷⁵ In his dissent joined by Justice Thomas in *United States v. Windsor*, Justice Alito argued that “the right to same-sex marriage [is not] deeply rooted in the traditions of other nations. No country allowed same-

⁷² Steve Doughty, *Gay Marriage is not a “Human Right”*: European Ruling Torpedoes Coalition Stance, THE DAILY MAIL (Mar. 20, 2012), <http://www.dailymail.co.uk/news/article-2117920/Gay-marriage-human-right-European-ruling-torpedoes-Coalition-stance.html>; See also *European Court Rules Same-Sex Marriage is Not a Human Right*, WOMEN OF GRACE (Mar. 27, 2012), <http://www.womenofgrace.com/blog/?p=13371>. However, same-sex marriage proponents countered references to *Schalk and Kopf v. Austria* and *Gas and Dubois v. France* by using examples from other European countries that had previously recognized same-sex unions. See also LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, at 33 (explaining how international and comparative benchmarks may pull in different directions, and why a single model may be helpful in defining a European norm).

⁷³ Brief for Foreign and Comparative Law Experts Harold Hongju Koh, Sarah H. Cleveland, Laurence R. Helfer, and Ryan Goodman as Amici Curiae Supporting Respondents at 5, 12, 38, *Hollingsworth v. Perry*, No. 12-144 (Supreme Court argued Mar. 26, 2013), 2013 WL 795544.

⁷⁴ Brief for International Jurists and Academics as Amicus Curiae Supporting Petitioner *Hollingsworth* and Respondent Bipartisan Legal Advisory Group at 24-26, *Hollingsworth v. Perry*, No. 12-144 (Supreme Court argued Mar. 26, 2013); *U.S. v. Windsor*, No. 12-307 (Supreme Court argued Mar. 27, 2013), 2013 WL 1780813.

⁷⁵ Brief for International Jurists and Academics as Amicus Curiae Supporting Petitioner *Hollingsworth* and Respondent Bipartisan Legal Advisory Group at 24-26, *Hollingsworth v. Perry*, No. 12-144 (Supreme Court argued Mar. 26, 2013); *U.S. v. Windsor*, No. 12-307 (Supreme Court argued Mar. 27, 2013), 2013 WL 1780813.

sex couples to marry until the Netherlands did so in 2000.”⁷⁶ Despite historical discomfort with United States invocations of international law among conservatives, Alito was the only Justice to cite to foreign practice in both *Hollingsworth v. Perry* and *United States v. Windsor*, and he did so with the purpose of countering the Supreme Court’s tentative support of same-sex equality.

The ECtHR decisions are also finding their way into the debate through the media.⁷⁷ The rulings themselves are nuanced and leave room for a diversity of national positions on the issue, but they are often transmitted to a popular audience in much simpler and more conclusory terms. Many press accounts have simply reported that the Court had rejected the notion of same-sex marriage as a human right. *The Telegraph* headline was “Gay marriage is not a human right, according to European ruling.”⁷⁸ The *Daily Mail* declared that “Gay marriage is not a 'human right': European ruling torpedoes Coalition stance.”⁷⁹ It is self-evident that gay rights advocates would prefer that the public not be exposed to such headlines. Their concerns are that political support for same-sex marriage may be undercut by the relatively low international human rights norm resulting in backsliding. In short, the exclusion of same-sex marriage from a transnational set of human rights standards can make domestic advocacy for same-sex marriage harder and can also lower the chances that legislation for same-sex marriage will pass.

C. Women’s Rights in Sweden

⁷⁶ *United States v. Windsor*, 133 S.Ct. 2675, 2715 (2013) (Alito, J., dissenting) (citing CURRY–SUMNER, A PATCHWORK OF PARTNERSHIPS: COMPARATIVE OVERVIEW OF REGISTRATION SCHEMES IN EUROPE, IN LEGAL RECOGNITION OF SAME–SEX PARTNERSHIPS 71, 72 (K. Boele–Woelki & A. Fuchs eds., rev. 2d ed., 2012)). Justice Alito also notes “virtually every culture, including many not influenced by the Abrahamic religions, has limited marriage to people of the opposite sex.” *United States v. Windsor*, 133 S. Ct. 2675, 2718 (2013) (citing Brief on the Merits for Respondent the Bipartisan Legal Advisory Group of the U.S. House of Representatives at 2, *United States v. Windsor*, No. 12-307 (Supreme Court argued Mar. 27, 2013), 2013 WL 267026 (citing *Hernandez v. Robles*, 7 N.Y.3d 338, 361 (2006) (“Until a few decades ago, it was an accepted truth for almost everyone who ever lived, in any society in which marriage existed, that there could be marriages only between participants of different sex”))).

⁷⁷ 10 Dec. 2012, PARL. DEB., H.C. (2012) 23-25 (U.K.); 11 Dec. 2012, PARL. DEB., H.C. (2012) 156-57; 164 (U.K.).

⁷⁸ Donna Bowater, *Gay Marriage is Not a Human Right, According to European Court*, THE TELEGRAPH (Mar. 21, 2012), <http://www.telegraph.co.uk/news/religion/9157029/Gay-marriage-is-not-a-human-right-according-to-European-ruling.html>.

⁷⁹ Doughty, *supra* note 72.

Our third example of backsliding illustrates a reduction in parental leave benefits in Sweden following a European directive intended to establish minimum standards in this area. In this particular case, concerns about backsliding were strong enough to provoke opposition to the entire project of European integration from some advocates of the Swedish welfare state.

Many people value parental leave as an important vehicle to allow working parents, and especially working women, to continue with their careers. By some estimates, maternity-leave laws can increase the employment rate of women of childbearing age by seven percent to nine percent.⁸⁰ On the other hand, parental leave imposes regulatory burdens on employers and restricts the freedom of business to operate as they wish. Furthermore, a more generous leave policy confers benefits only on parents and only on parents who chose to work rather than stay at home.⁸¹ In short, parental-leave policies, like all policies, are subject to political contestation.

In the course of domestic debates on parental leave, international-leave standards can be used both by progressives, if the international norm is higher than existing or proposed domestic norms and by conservatives, if it is set at a lower level.

Sweden is a global leader with respect to expanded parental leave. Starting in 1974, Sweden started offering six months of leave to either parent, compensated at 90 percent of their prior salary.⁸² Of note for this Article, Sweden cut back its generous benefits and limited women's choices following the introduction of EU-wide minimum standards.⁸³

⁸⁰ See Christopher J. Ruhm, *The Economic Consequences of Parental Leave Mandates: Lessons from Europe*, 113 Q. J. ECON. 285, 310-11 (1998). See also Christine Jolls, *Accommodation Mandates*, 53 STAN. L. REV. 223, 290-300 (2000) (discussing the effects of the FMLA); Gillian Lester, *A Defense of Paid Family Leave*, 28 HARV. J.L. & GENDER 1 (2005) (reviewing empirical work concluding that paid leave increases women's labor force participation, and arguing that this is normatively desirable); CATHERINE R. ALBISTON, *RIGHTS ON LEAVE: INSTITUTIONAL INEQUALITY AND THE MOBILIZATION OF THE FAMILY AND MEDICAL LEAVE ACT* (2010) (explaining how lack of information and uneven enforcement of leave laws influences take-up).

⁸¹ See generally Gillian Lester, *supra* note 72 (reviewing the arguments in favor and against maternity leave); Linos, *Diffusion*, *supra* note 1 (explaining the political economy of maternity leave debates).

⁸² See Ann-Zofie Duvander, Tommy Ferrarini, & Sara Thalberg, *Swedish Parental Leave and Gender Equality Achievements and Reform Challenges in a European Perspective*, 11 INSTITUTE FOR FUTURES STUDIES 3, available at

http://www.politiquessociales.net/IMG/pdf/20051201_134956_fil_U8YIJLRaA7u4FV7gUmy.pdf.

⁸³ The discussion that follows focused on reforms in the early and mid-1990s, a period during which the EU promulgated a major directive on minimum maternity leave standards, and a period during which Sweden joined the EU. There have since been additional reforms to Swedish parental leave benefits. For an overview of current benefits, see Socialförsäkringsbalk, SFS 2010:110, 13 kap. § 10 – 14, available at <http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Socialforsakringsbalk->

Sweden started seriously debating entering the EU in 1989, after the fall of the Berlin Wall.⁸⁴ Soon thereafter, several defenders of the generous Swedish welfare state expressed concerns about whether Sweden's entry into the EU would lessen protections.⁸⁵ Among their worries was a European directive setting minimum maternity leave standards throughout the EU.⁸⁶ The 1992 EU Pregnant Workers directive specified that all EU member states must offer a minimum of fourteen weeks of maternity leave.⁸⁷ Moreover, the directive specified that this leave should be adequately compensated,⁸⁸ and that compensation comparable to that provided for sick leave is considered adequate.⁸⁹ In addition, the directive specified that all women should be obligated to take at least two weeks of leave.⁹⁰ This mandatory two-week minimum reflected concerns some women might face employer pressure to turn down the leave. Both the preamble of the directive and its text clearly indicate that it is intended to set minimum requirements, and that EU member states should not reduce existing protections if they are higher than the directive's minimum requirements.⁹¹

At first, Swedish objections focused on the two weeks of mandatory leave. Swedish feminists saw this element as a stereotypical allocation of parental responsibilities to the female parent, and as a restriction on women's choices. The Swedish government took these concerns very seriously. Whereas

201011_sfs-2010-110/?bet=2010:110#K13". We are very grateful to Helena Jung for locating and translating this and other Swedish sources for us.

⁸⁴ Government Offices of Sweden, Sweden's road to EU membership, *available at* <http://www.government.se/sb/d/3470/a/20685>.

⁸⁵ For many examples and a summary of these concerns, see MARIKA EHRENKRONA, GRANSKNING AV EU-KRITIKEN 45-51 (2001). To translate one example of the many examples listed in Ehrenkrona: "The handicapped, children, the old and the sick will be abandoned. The nurturing model has come to an end What our fathers toiled to create, the EU will take from us." *Id.* at 50-51, citing Sven Svensson Pukavik, *Handikappade inom EU fruktar för sina liv*, Blekinge Läns Tidning, July 8 (1994).

⁸⁶ See Milena Sunnus, *EU Challenges to the Pioneer in Gender Equality: The Case of Sweden*, in GENDERING EUROPEANISATION 239 (Liebert Ulrike ed., 2003). See also Ehrenkrona, *supra* note 77 (giving overview of what the EU-critical voices in Sweden had to say when Sweden was about to have a referendum on EU membership).

⁸⁷ See Pregnant Workers Directive 92/85, art. 8, 1992 O.J. (L348/1) (EEC), *available at* <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31992L0085:EN:NOT>.

⁸⁸ *Id.* at art. 11, § 2.

⁸⁹ *Id.* at art. 11, § 3.

⁹⁰ *Id.* at art. 8.

⁹¹ See, e.g., *id.* at pmb1. ("Whereas Article 118a of the Treaty provides that the Council shall adopt, by means of directives, minimum requirements for encouraging improvements, especially in the working environment, to protect the safety and health of workers; Whereas this Directive does not justify any reduction in levels of protection already achieved in individual Member States"); see also *id.* at art. 1 § 3 ("This Directive may not have the effect of reducing the level of protection afforded to pregnant workers . . . as compared with the situation which exists in each Member State on the date on which this Directive is adopted").

Sweden typically transposes EU directives very promptly, Sweden delayed in the introduction of this provision of the directive, and only transposed years later, after the EU Commission had started infringement proceedings.⁹² A larger concern arose from the fact that Swedish legislation transposing the EU directive also reduced the compensation levels for maternity leave to the levels of compensation typical of sick leave. Parental leave had been compensated at 90 percent of a worker's earlier salary, but the Swedish legislation transposing the EU directive reduced compensation to 80 percent and later legislation reduced this further to 75 percent.⁹³ The Swedish government's report justifying this reduction explicitly referenced the European standards, and stated that maternity leave payments that were in line with sick leave payments were adequate.⁹⁴

To be clear, the policy choices made by Sweden were the result of many factors, as is always the case with significant political decisions. That said, the evidence suggests that European minimum standards put some downward pressure on domestic policies. Opponents of the policy change certainly seem to believe as much: "[A]fter Sweden's entry in the EU, women experienced reductions of parental leave payments from 90 p.c. to 75 p.c. of their income."⁹⁵

Concern about backsliding was not limited to parental leave. Concerns were also expressed about how Europe might influence Sweden's generous benefits to children, the elderly, and the disabled.⁹⁶ Some of these concerns are likely overblown, but they illustrate that domestic interest groups perceived backsliding as a real danger. Furthermore, in some areas, such as parental leave, fears that benefits might be reduced proved justified.

⁹² See Gerda Falkner et al., Non- Compliance with EU Directives in the Member States: Opposition through the Backdoor?, 27 W. EUR. POL. 452, 458-59 (2004).

⁹³ See Socialförsäkringsbalk, 28 kap. § 7, 12 kap. § 25 & 12 kap § 18, *available at* http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Socialforsakringsbalk-201011_sfs-2010-110/?bet=2010:110. (translated from Swedish by Helena Jung).

⁹⁴ Regeringens Proposition [Prop.] 1994/95:207 Ny föräldraledighetslag m.m., [New Act on Parental Leave, etc.] (Swed.), at chp. 4.8, *available at* http://www.riksdagen.se/sv/Dokument-Lagar/Forslag/Propositioner-och-skrivelser/Ny-foraldraledighetslag-mm_GI03207/?html=true (translation from the Swedish by Helena Jung).

⁹⁵ See Sunnus, *supra* note 78, at 239 (describing how feminists have been critical of EU measures, emphasizing that "after Sweden's entry in the EU, women experienced reductions of parental leave payments from 90 p.c. to 75 p.c. of their income.").

⁹⁶ For an overview of these concerns, see EHRENKRONA, *supra* note 77, at 45-51. See also LJUSNANDE FRAMTID ELLER ETT LÅNGT FARVÄL? DEN SVENSKA VÄLFÄRDSSTATEN I JÄMFÖRANDE BELYSNING 175-76 (Agneta Stark ed., 1997) (stating that "Sweden is the archetype for the social democratic welfare model, and appears as the country that, because of the generous integration policies of the EU will lose most of its generous benefits and social services if the country is pressured—directly or indirectly—to adopt the minimum standard prevalent in countries with low taxes and social welfare costs"). See also *id.* at 211-12 for similar concerns. For other interesting examples of backsliding and backlash arguments, see CLIFFORD BOB, THE GLOBAL RIGHT WING AND THE CLASH OF WORLD POLITICS 2012.

The above examples obviously do not constitute a rigorous empirical test of our theory. We offer them, instead, instead to help explain the underlying mechanism. To these illustrations might be added previously mentioned patterns that have puzzled observers: that rights omitted from the UDHR became less popular in national constitutions than they had been before;⁹⁷ and that rich democracies sometimes worsen their behavior following treaty ratification.⁹⁸ In the next Part, we present a simple formal model to demonstrate that backsliding is theoretically coherent and plausible.

Part III. The Model

The merit of our argument rests heavily on our claim that explanations of how human rights norms are transmitted to domestic policy predict both an improvement in the practices of low-performing states and a reduction in human rights protections in high-performing states. The model that follows allows us to make this point in a formal way and makes transparent the assumptions embedded in our analysis.

For purposes of illustration we discuss freedom of speech, but the spirit of our argument applies to any human rights issue, including economic and social rights as well as civil and political rights. The magnitude of the effects we describe is almost certainly affected by a variety of factors, including the issue area at hand, the nature of the international norm, and the domestic politics of the receiving country, among other factors. For clarity of exposition, we first present a simplified model in this Part, and discuss these and other extensions and variations in the next Part.

We begin by imagining how states would adopt a human rights policy in the absence of any international legal obligation. As with any policy, domestic decision-makers must weigh the benefits and costs of their choice.⁹⁹ We focus on the payoffs to the decision-makers themselves because our model of domestic politics turns on how those political leaders' choices are influenced.¹⁰⁰

The trade-offs associated with speech rights are familiar. Granting citizens greater freedom of speech provides benefits to political leaders inasmuch as it offers an outlet for criticism and promotes the

⁹⁷ See Elkins, Ginsburg & Simmons, *supra* note 10.

⁹⁸ See SIMMONS, *MOBILIZING*, *supra* note 1, at 281-82 (suggesting that among high rule of law states, countries that have ratified the Convention Against Torture report a higher prevalence of torture than countries that have not ratified the Convention).

⁹⁹ While the cost-benefit analysis framework may not appear immediately applicable to rights that are sometimes understood as absolute, such as freedom from torture, genocide, or slavery, we find that states think very carefully about the trade-offs involved in defining these prohibitions broadly or narrowly. See discussion *infra* IV.B.1.

¹⁰⁰ This assumption does not limit the applicability of our analysis. We require only that policy decisions are influenced by public perceptions and preferences.

exchange of ideas, hopefully leading to better policy making. It also tends to reduce political unrest or dissatisfaction with the governing regime. Of course, government leaders may also support more liberalized speech rights as a fundamental human right.

On the other hand, there are costs, especially from the perspective of the government currently in power. Allowing more free speech exposes the government to more criticism, may make it more difficult to keep important information confidential, and may discourage leaders from extracting private benefits from their time in office. A more liberal speech regime also necessarily allows more hateful speech, which might be viewed as an infringement on the rights of others. This basic balancing of priorities takes place in every country and political leaders make decisions about what policy to adopt.¹⁰¹ For our purposes the specifics of this balancing are not critical. We need only assume that leaders benefit from providing some protection of the relevant human right, but that at some point the benefits of further protection is outweighed by the costs. This is true for virtually every right discussed at the international level, including economic and social rights.¹⁰²

We index each country's freedom of speech with the variable S , where $S > 0$. A higher value of S represents a more liberal free speech regime. We define a function, B_i , that represents the benefits of greater free speech for country i . We assume B is a simple linear function of S :

$$B_i = 2S_i, S > 0$$

We also define a function, C_i , that represents the costs of greater free speech. Like B_i , C_i is increasing in S . We assume that costs increase quadratically with S :¹⁰³

¹⁰¹ See generally Jeremy Waldron, *THE HARM IN HATE SPEECH* (2012).

¹⁰² The rhetoric surrounding rights (including speech rights) is often absolutist, resisting the notion that a particular standard is the result of competing priorities. This is obviously an incorrect description of the vast majority of rights. The rights of criminal defendants, for example, are inevitably balanced against a societal interest in public safety; the rights of citizens to participate in the democratic process are balanced against concerns about the integrity of the voting system and administrability concerns; the rights of minority groups to be free from discrimination is balanced against a desire to avoid providing those same groups with unjustified advantages. Another way to make the same point is to observe that a human rights issue comes to be of interest only when there is some disagreement about the appropriate standard. There can only be disagreement when there are competing priorities. See Part IV.B.1 for a discussion of human rights standards that are implemented as "absolutist" norms.

¹⁰³ Several factors could contribute to a steep increase in the costs of greater free speech. If we arrange free speech topics from mundane matters that are completely non-threatening to the regime, to highly political matters such as calls for the violent overthrow of the regime, it seems plausible that a regime will find it especially difficult to grant free speech protections to a small final set of comments concerning regime overthrow.

$$C_i = \lambda_i S_i^2$$

where $\lambda_i > 0$ reflects the idiosyncratic cost of human rights to country i . A higher λ_i signals that the country's leaders view human rights as more costly to provide. This may be, for example, because they have a strong preference for corruption and graft that freedom of speech will make more difficult, because they believe suppressing speech will allow them to remain in power longer, because the particular history of the country leads them to be particularly hostile to hate speech, or for some other reason.

The λ_i variable, then, captures the different preferences among countries and leaders with respect to human rights. Simply put, as λ_i increases, the cost of any particular free speech regime also increases, making it less attractive.

Our results would be largely the same with many other functional forms for both costs and benefits. The key is that there be some balancing of costs and benefits that leads to an interior solution with neither an infinite provision of the right nor a complete denial of it. We could have, for example, a convex cost curve (as we do) and a concave benefits curve (rather than our present linear one). This would complicate the arithmetic, but would not change the fundamental results. Indeed, the qualitative results would remain the same for any set of continuous, monotonic curves where curve B crosses curve C only once, from above.

In the absence of international law, each country chooses S_i , the level of freedom of speech it will provide, so as to maximize its resulting gains. We define a leader's welfare function, W_i , as the difference between the benefits and costs of any particular S :

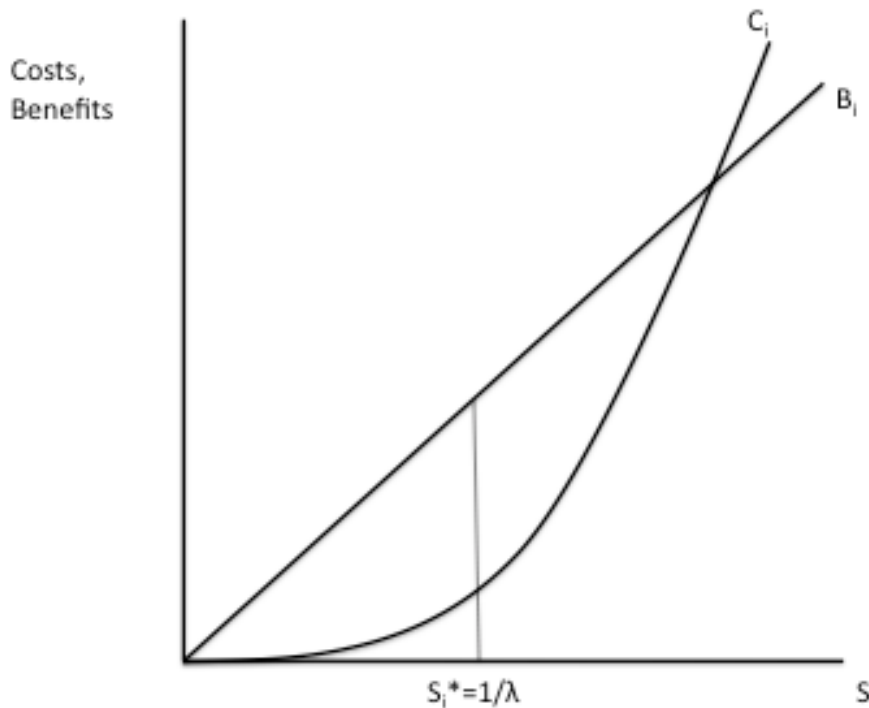
$$\begin{aligned} W_i &= B_i - C_i \\ &= 2S_i - \lambda_i S_i^2 \end{aligned}$$

Differentiating with respect to S to identify the first-order conditions and simplifying yields the optimal speech level:

$$S_i^* = 1/\lambda$$

This result, as one would expect, indicates that as λ increases, the level of human rights chosen by the state decreases. It is illustrated in Figure 1, with the optimal policy, S_i^* located at the point where the costs and benefits curves are furthest apart.

Figure 1: Domestic Human Rights Policy in the Absence of International Law



Now consider how an international human rights agreement might impact this outcome. When a relevant international human rights norm, k , emerges, how is the domestic policy decision impacted?¹⁰⁴ We need to consider two cases. In the first, the international norm is higher than the equilibrium norm, S^* , that a country would otherwise choose. The second case is the opposite situation, where the international norm is lower than the domestic choice in the absence of an international influence.

A. International Norm Higher than Domestic Equilibrium ($k > S^*$)

Consider first the case in which the international norm demands a higher level of human rights protection—greater protection of freedom of speech in our illustration. As already mentioned, we assume this norm matters because it bolsters the political power of interest groups within the country

¹⁰⁴ We put to one side the important question of what it takes for such a norm to exist. It might require a treaty to which the state is a party, or perhaps it is enough that the treaty be widely signed, even if the state in question does not join. It could be that no formal treaty is required—a norm promulgated by an international organization might be sufficient.

seeking an increase in human rights. In our case, that means that those who support greater freedom of speech gain support.

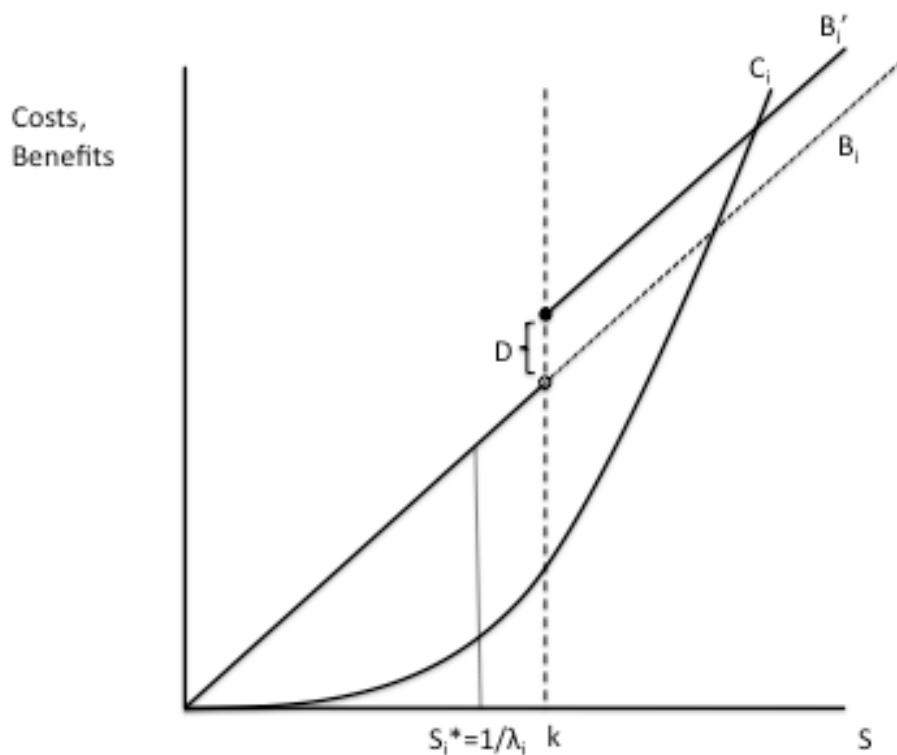
Put another way, if the country's leaders adopt the international standard, they will enjoy an increased level of political support from those who consider free speech an important issue. We can think of this, for example, as putting them in a better position for future elections because they can point to the international standard and argue that they have embraced an internationally accepted human rights position.

We model this as a discontinuity in the benefits curve facing the state. If the international threshold is satisfied, political leaders gain a discrete increase in support, which we call D . This can be represented with a modified benefits curve, B'_i :

$$B'_i = \begin{cases} 2S_i & , \text{ if } S < k \\ 2S_i + D & , \text{ if } S \geq k \end{cases}$$

The cost and benefits curves then look as follows:

Figure 2: Human Rights Behavior with International Law



Notice that the international norm is to the right of the state's ex ante preferred position. This reflects the fact that the international norm is more protective of speech than is S_i^* . The arrival of this norm presents the state with a new possibility. By moving from S_i^* to k , the state incurs greater costs, but it also benefits from the discrete increase in benefits that occurs at k .

Because S_i^* is the best outcome for the state in the absence of international law, and because the international norm has not affected the state's payoff anywhere to the left of k , we know that the state prefers to remain at k rather than move to any point between S_i^* and k . Similarly, because the slope of C_i increases as we move to the right, we know that choosing a position to the right of k , $k+\epsilon$, for any $\epsilon > 0$ is worse for the state than choosing k .¹⁰⁵

¹⁰⁵ We know this because the slope of C at the pre-international law optimum, S^* , is equal to the slope of B (this is necessarily true if S^* is the optimum). Moving to the right from there, the slope of C increases while the slope of B remains constant and the slope of B' is the same as that of B . It follows that the net benefits decrease as we move to the right.

The state, then, will maximize its welfare, W_i , by choosing either S_i^* or k . The relevant welfare gains are represented as follows:

$$\begin{aligned} W_i(S_i^*) &= 2S_i - \lambda S_i^2 \\ &= 1/\lambda \end{aligned}$$

$$\begin{aligned} W_i(k) &= 2S_i + D - \lambda_i S_i^2, \text{ where } S_i=k \\ &= 2k + D - \lambda_i k^2 \end{aligned}$$

Comparing these two outcomes, the state will choose the larger one. In other words, if:

$$1/\lambda_i > 2k + D - \lambda_i k^2$$

then the state remains at S_i^* and is unaffected by international law. If the opposite is true, the human rights norm is effective and the state moves to k . We are interested, therefore, in the relative values of k and λ at which the state is indifferent between improving its conduct and remaining at S_i^* .

The critical value of k at which a state with costs λ_i is indifferent is given by solving:

$$1/\lambda_i = 2k + D - \lambda_i k^2$$

$$\lambda_i k^2 - 2k + (1/\lambda_i - D) = 0$$

Solving this for k yields:¹⁰⁶ $\bar{k} = 1/\lambda_i + D^{1/2}/\lambda_i^{3/2}$

This relationship between k and λ_i identifies how states will react to the international norm. Specifically, a state with cost λ_i will respond to the norm by improving its human rights conduct (moving to k) as long as the chosen k is less than \bar{k} :

$$k < 1/\lambda_i + D^{1/2}/\lambda_i^{3/2}$$

If, on the other hand, the international norm is greater than \bar{k} , it will have no effect on the country. This means that a human rights norm that is too ambitious will leave states with poor human rights records unaffected.

If this is correct, then, for example, article 5(a) of CEDAW, which requires that “State Parties shall take all appropriate measures to modify the social and cultural patterns of conduct of men and women, with a view to achieving the elimination of prejudices and customary and other practices which are based on the idea of the inferiority or the superiority of either of the sexes or on stereotyped roles for men and

¹⁰⁶ To arrive at this result we use the quadratic formula to solve $ax^2 + bx + c$: $(-b \pm (b^2 - 4ac)^{1/2})/2a$. This yields (omitting subscripts) $k = 1/\lambda \pm D^{1/2}/\lambda^{3/2}$. We know, however, that $k > S_i^*$ and $S_i^* = 1/\lambda_i$. Therefore, $k > 1/\lambda_i$ and so we must add the second term in the expression for k rather than subtract it.

women,”¹⁰⁷ is unlikely to have much impact on states with the worst records on women’s rights. This relationship between the content of the norm and the set of countries impacted suggests that there is some reason to produce treaties that are less ambitious in the hope of influencing states with poor human rights records.¹⁰⁸ Thus, for example, in the ongoing negotiations on a potential multilateral treaty that would regulate trade in conventional weapons, many European governments are criticizing the United States for watering down human rights standards.¹⁰⁹ Our theory suggests that lower standards may in fact have a chance to generate an improvement in human rights that would not be possible with a tougher treaty.¹¹⁰

Among the other implications of this model is that if we are able to reduce the costs of adopting better human rights, more states will be responsive to an international norm. In particular, if the cost of better human rights conduct is reduced (i.e., smaller λ) then more states will move to satisfy the norm and states further away from the norm will do so as well.

Similarly, an increase in the benefits received by states that satisfy the norm (D) increases the range of states that are responsive and does so by making states farther away from the norm move to k .

Notice that this result accords well with our intuitions. An increase in D makes more states responsive to any particular choice of k ,¹¹¹ and an increase in the cost of granting human rights (i.e., higher λ) makes it less likely that a state will respond to an international norm.

B. International Norm Lower than Domestic Equilibrium ($k < S^*$)

The discussion above considered the case in which the chosen international norm was more protective of human rights than the national policy in the absence of international law. We now turn to consider the alternative situation in which the international norm is less protective than the ex ante domestic policy. In most human rights discussions, it is implicitly assumed that an international norm, even if it is

¹⁰⁷ Convention on the Elimination of All Forms of Discrimination Against Women art. 5(a), *opened for signature* Dec. 18, 1979, 1249 U.N.T.S. 13. (entered into force Sep. 3, 1981).

¹⁰⁸ An important caveat to this point is developed below, where we explain that states may wish to include high standards to prevent backsliding.

¹⁰⁹ See Nick Hopkins, *UK Presses US on Human Rights Clause in Arms Trade Treaty*, THE GUARDIAN (July 1, 2012), <http://www.guardian.co.uk/world/2012/jul/02/human-rights-arms-trade-treaty>.

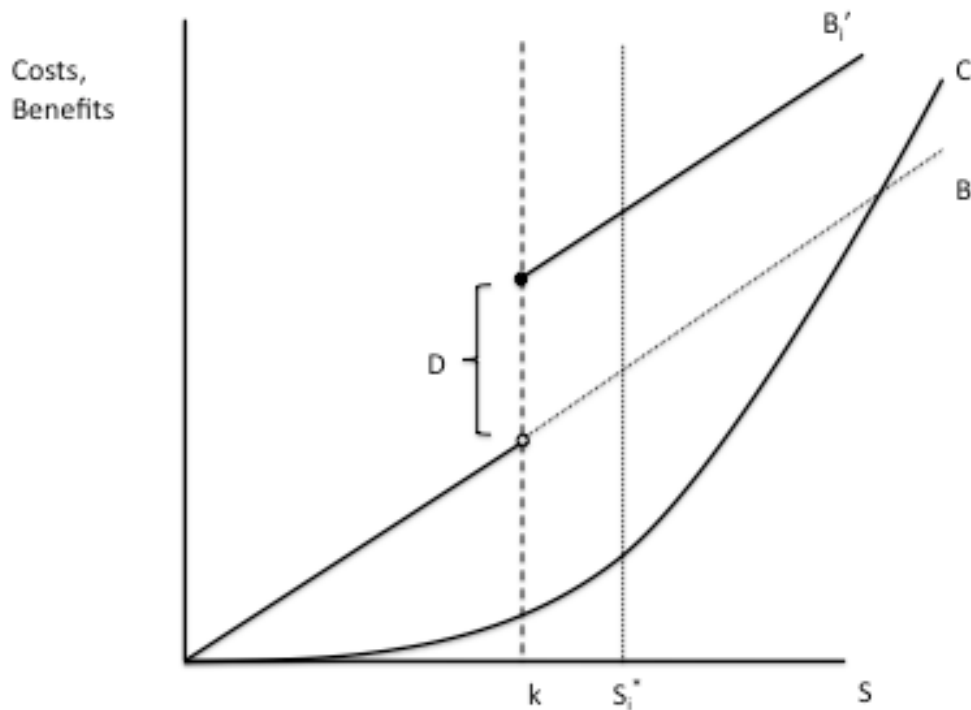
¹¹⁰ It is sometimes observed that treaties must sometimes be weakened in order to get more states to join the treaty regime. Our point is somewhat different. A weaker treaty regime may affect different states even if all the same states are part of the regime. By weakening a regime, then, we may be able to influence those with poor human rights records more than could be done with a tougher treaty, even if those states were willing to join either regime.

¹¹¹ Notice also that if $D=0$, $k=1/\lambda=S^*$, meaning that we are back in the base case without international law. This would also be the case if international law had no effect.

effective, will have no impact on states that provide a higher level of human rights protections. In what follows, we first consider the assumptions necessary to generate this result, and then turn to show how a change in those assumptions might lead to a different outcome.

Figure 3 illustrates the case in which the state's preferred position in the absence of international law, S_i^* , is more protective of speech rights than the international norm, k . The presence of the norm increases the benefits received by the state's leaders because they enjoy the discrete gain, D , associated with meeting the norm. It does not, however, change a state's human rights conduct. Graphically, this is evident because the point at which the state maximizes its gains is the point at which the slope of B_i (or B_i'), which reflects the marginal benefits from increase human rights protection, is equal to the slope of C_i , which is the marginal cost of increased protections. The discontinuity in the B_i curve at the point k does not affect the slope of either the B_i or the C_i curve, so the optimum is unchanged.

Figure 3: Human Rights Behavior with International Law II



This representation of a human rights norm is consistent with the dominant view that establishing a norm will help to improve the performance of some states that previously fell short of the norm, but will have no impact on those that exceed the norm at the time of adoption.

Given our theory of human rights policy formation, however, there is reason to wonder if this result is correct. Our theory assumes that an international norm affects domestic politics by providing support to domestic groups that favor an outcome similar to that specified in the international norm. There is, however, no reason to think that the focal point created by the international norm will only influence states with weaker human rights standards. The increase in political influence enjoyed by groups that support a domestic policy similar to the international norm must come at the expense of other groups, possibly including groups that prefer a higher level of human rights.

This point is illustrated in the examples given in Part II. Consider how it would come about in the context of free speech. For countries with limited protections for free speech, our theory suggests that an international norm specifying some required (or perhaps only recommended) level of protection for speech can increase the influence of local interest groups that seek domestic policies with similar protections. This increase in influence can prompt political leaders to make different policy choices. We model this domestic politics as a one-time gain to countries that satisfy the international norm.

Now consider a state with existing protection in excess of that required by the international norm. This country also has interest groups engaged in a debate about the merits of increased freedom of speech. When an international norm emerges, one can easily imagine groups that favor a less speech-protective and more hate-speech focused domestic policy will gain influence. These groups would point to the international norm to support their claim that the appropriate balance is less speech-protective than the status quo.

For example, the ICCPR provides protections for freedom of expression in Article 19.¹¹² Yet Article 20 of this instrument can be interpreted to require that hate speech be regulated in a way that would restrict speech more than is done in the United States.¹¹³ Note that the United States has made a reservation to this article, and thus does not face a legal obligation to restrict hate speech.¹¹⁴ However, the logic we propose here is political; advocates of hate speech restrictions can use Article 20 of the ICCPR to suggest that the United States is out of line with the rest of the world.¹¹⁵

¹¹² See International Covenant on Civil and Political Rights art. 19, *opened for signature* Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976) [hereinafter ICCPR].

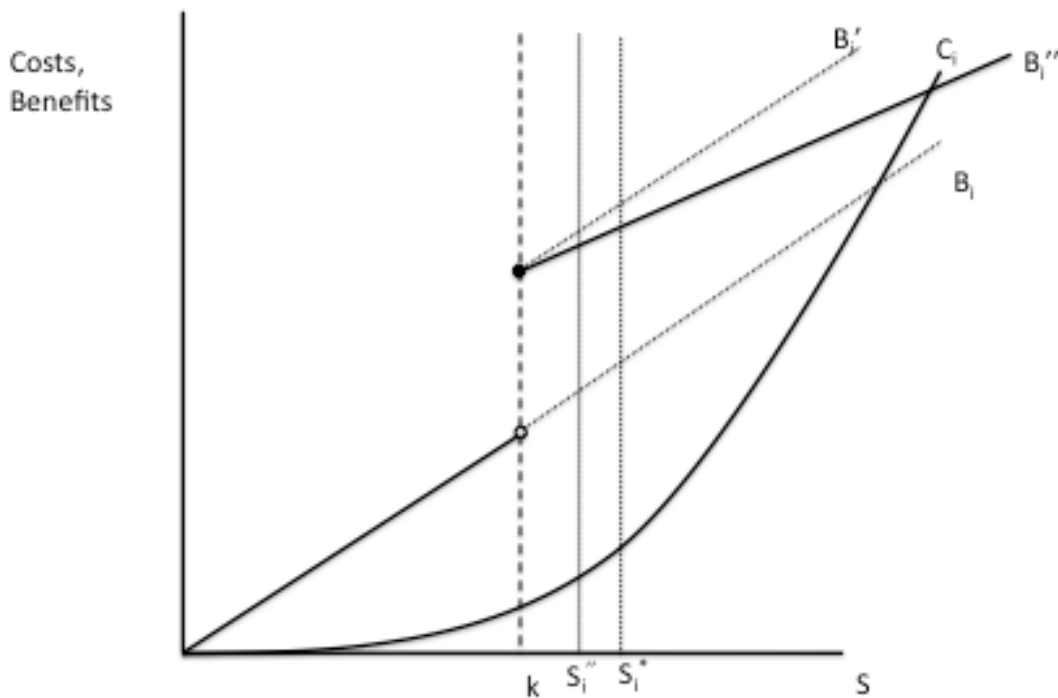
¹¹³ See *id.* at art. 20.

¹¹⁴ See U.S. reservations, declarations, and understandings, International Covenant on Civil and Political Rights, 138 Cong. Rec. S4781-01 (daily ed., April 2, 1992).

¹¹⁵ For example, the description of Professor Jeremy Waldron's prominent attack on the United States' unusually broad free speech protections begins with the sentence "Every democracy has laws or codes

Figure 4 shows how this backsliding could happen. Once the international norm, k , is established, the perceived benefits of marginal movements to the right of k are reduced. That is, the international norm signals that the balance between the human right in question (e.g., speech) and other concerns (e.g., hate speech) recommend a policy located at k . That information filters through the domestic political process in a way that reduces the marginal benefits enjoyed by political leaders when they adopt policies to the right of k .¹¹⁶ We represent this change in benefits as a change in the slope of the B_i' curve to the right of k , shown by the line B_i'' .

Figure 4: Human Rights Behavior with International Law III



against hate speech—except the United States.” *The Harm in Hate Speech*, HARVARD UNIVERSITY PRESS (last visited Mar. 5, 2013), <http://www.hup.harvard.edu/catalog.php?isbn=9780674065895>.

¹¹⁶ A similar result could be achieved by modeling the change as an increase in costs.

The state reacts to this new benefits curve by adjusting its provision of human rights so as to maximize the net gains from its policy. This entails a move to the left, to S_i'' .

The same result can be shown more formally:

$$C_i = \lambda_i S_i^2$$

$$B_i = \begin{cases} 2S_i, & \text{if } S_i < k \\ 2k + D + (S_i - k), & \text{if } S_i \geq k \end{cases}$$

The benefits curve includes $2k + D$, the benefit enjoyed by the state if it adopts policy k , plus the additional benefits of moving to the right beyond k . Note that as drawn the slope of the B_i'' curve is one, less than the slope of the B_i curve, which is two. This captures the notion that the international norm may provide a kind of anchor below that country's status quo ante.

Any country that would choose a level of human rights less than k in the absence of international law will be affected in the same way as already discussed in Part III.A. That is, states located to the left of k , but sufficiently close to it, will migrate to k . Those located farther to the left will not be affected.

Because the slope of B_i'' has changed, there is now an effect felt by states to the right of k . In the absence of international law, these states choose $S_i^* = 1/\lambda_i$, as shown in Figure 4.

With the international norm in place, they maximize the difference between benefits and costs:

$$[A] \quad W_{Li} = 2k + D + S_i - k - \lambda_i S_i^2$$

Differentiating with respect to S allows us to identify the first-order conditions for this problem and yields:

$$S_{Li}^* = 1/2\lambda_i$$

Notice that this is to the left of the prior outcome of $S_i^* = 1/\lambda_i$.

S_{Li}^* may be the optimal policy for the state's leaders, but we must consider one additional constraint. Given that states started off to the right of k , we know that they will not respond to the international norm by moving to the left of k . Doing so would sacrifice the benefit D and place them on the original B_i curve. We know this is inferior to an outcome in which the state chooses to locate itself at k .

They will, however, move to k if the welfare from being at k exceeds that of being at S_{Li}^* . From the first order conditions, however, we know that if $1/2\lambda_i$ is larger than k , the state is better off at $1/2\lambda_i$. If the opposite is true, and $k > 1/2\lambda_i$, then the state will always be better off at k , which we know because prior to the arrival of international law the state preferred $S^* > k$, which implies that moving to the left from point k reduces the payoff to the state.

All of this means that the state will move to the larger of $1/2\lambda_i$ and k . In either case we get backsliding. To summarize: it is the flattening of the B curve that is doing the work in our backsliding theory. A

completely flat B curve would lead every country that would otherwise provide higher protection than the international standard to backslide all the way back to the international standard.¹¹⁷

Comparing the two cases, it is worth noting that the positive effects of the international norm are felt only by states that are below but close to the norm while the backsliding effects are felt by all states that are above the norm. This does not necessarily mean that the norm does more harm than good, of course, because we would need additional assumptions in order to make a prediction about the number of countries in each category and the magnitude of the movements, not to mention how to aggregate outcomes across states.

Part IV. Extending the Theory of Backsliding

For clarity of exposition, we intentionally kept the backsliding model presented above simple. In this section, we offer some further analysis of the theory and consider several extensions. We first outline how domestic factors, such as democracy and the role of the judiciary, influence the impact of backsliding. We then examine how changes to the form and content of a norm might matter. In this second category we consider: the use of minimum standards (floors), maximum standards (ceilings), and prohibitions; the legal form of the norm (hard law versus soft law); norms that identify multiple standards; and vague norms and uncertainty with respect to national compliance. Finally, we discuss the extent to which alternative theories of how international norms affect human rights outcomes can lead to backsliding.

A. Domestic Factors Influencing Backsliding

A broad implication of our theory is that once an international standard comes into place, state behavior should tend toward this standard as poor performers are pulled upwards and top performers are pulled down. It does not predict, however, that states will fully converge on the international norm or that all similarly situated countries will necessarily move in the same fashion.

¹¹⁷ This extreme case of a B' curve that is flat to the right of K is worth considering, not because it is realistic, but because it can clarify this logic. A B' curve that is flat to the right of k reflects a norm strong enough to eliminate any governmental benefits from moving farther to the right. It is difficult to imagine such a norm in practice, but it might be one that is not merely a floor, but also a ceiling, meaning it specifies a precise level of treatment that is required. So, for example, if (contrary to fact) the international human rights law specified a specific balance between speech rights and hate speech, the B' curve might be flat. If the international norm causes the B curve to become flat to the right of k, then every country that would otherwise provide a higher level of protection will backslide all the way to k.

Different countries are more or less receptive to international norms. We argue here that factors that make some countries especially receptive to international norms, and thus especially likely to improve their behavior when an international norm is set at a high level, also contribute to a high risk of backsliding. Countries that are highly integrated into the international system, and whose governments regularly participate in international meetings and take on international obligations are more likely to be influenced by international norms than countries that are more isolated from the international system, both positively and negatively.¹¹⁸ For example, because the United States resists international norms more strongly than do many other rich democracies, the risk of backsliding is reduced, as is the potential for an upward pull.¹¹⁹ We also expect democratic countries to be more receptive to international norms than authoritarian ones, because the domestic politics mechanism we just described relies on pressures from voters, interest groups, and other key constituencies.¹²⁰ The heightened receptivity of democracies to international norms should influence countries both when the international standard is set high relative to their current level of domestic protections, and when it is set relatively low.

It follows that democracies that are highly integrated in international and regional systems face the greatest risks of backsliding. For example, countries like the UK and Sweden, which are highly integrated in European regional systems, face high risks of backsliding on issues where Anglo-Saxon and Scandinavian legal traditions offer very different (and more robust) protections from those prevalent among continental European states. Similarly, strong regional human rights ties within the Americas can place top performers in this region at risk of backsliding.

We note one curious feature of this prediction. If one assumes that democracies have, in general, better human rights records in general than autocracies, the asymmetry between the response of these two groups to a norm may at times frustrate human rights efforts. A norm aimed at improving the conduct of low performers (by assumption, predominantly autocracies) will have a relatively small positive influence on those same low performers but will have a relatively large backsliding impact on high-performing democracies. One can debate whether this compromise position—modest upward movement among low performers and more significant downward movement among high-performers—is desirable. Moreover other mechanisms may also be at work, mechanisms that operate asymmetrically

¹¹⁸ Other causes of rights stagnation and diminutions include changing views of the impacts of rights, economic and political shocks, treaty fatigue, and a desire to avoid creating new international institutions. See e.g. Benedikt Goderis & Mila Versteeg, *Human Rights After 9/11 and the Role of Constitutional Constraints*, 41 J. LEGAL STUD. 131 (2012) (arguing that after 9/11, the human rights records of the United States and its allies deteriorated due to negative norm diffusion).

¹¹⁹ See Chicago Council on Foreign Relations, *Global Views 2004: American Public Opinion and Foreign Policy* 50–53 (2004) (explaining that Americans are less supportive of a variety of international efforts than are citizens of other countries).

¹²⁰ See Linos, *Diffusion*, *supra* note 1, at 679; LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, at 18-26.

and are designed to work only in the direction of expanding rights protections. But at a minimum, the risk of backsliding should give one pause before advocating for such a norm.

The above discussion identifies democracies as susceptible to backsliding because we expect democracies to be more responsive to domestic interest groups and, therefore, to international norms. This point can be generalized. Any feature of the domestic political system that affects the force with which international norms are transmitted to domestic policy will operate in a similar way.

Consider, for example, the role of courts within a system. In our formal model, we use freedom of speech as an illustration. The content of that freedom in the United States is heavily influenced—arguably determined—by the judiciary rather than the legislature or the executive. Where this is true—where the content of a right is determined by courts rather than other actors, domestic policy is less subject to political influences.¹²¹ It follows that international human rights norms will have less impact, with the possible exception of legally binding norms that domestic courts are required to respect and enforce. It follows that backsliding may be less of a concern in these circumstances.¹²²

For our purposes the key point here is that the risk of backsliding goes hand-in-hand with the potential for international norms to improve outcomes. The politics that leads to backsliding is fundamentally the same as the politics that allows international norms to enhance domestic policies. Any influence that causes international norms to have a stronger effect on domestic policies will do so by promoting human rights policies in some countries and by increasing the risk of backsliding in others.

B. How the Content and Form of the Norm Matters

1. Floors, Ceilings, and Prohibitions

Our discussion so far has assumed that the norms at issue are floors and not ceilings. That is, we have assumed that the relevant international norm requires states to meet some minimum threshold but allows them to exceed it if they wish. This is a fair characterization of most key international human rights obligations.¹²³ The ICCPR provides a right to freedom of speech, for example, but countries are at liberty to provide more expansive rights, as do the United States and many others.¹²⁴ Similarly, there is a

¹²¹ See Linos, *Diffusion*, *supra* note 1; LINOS, DEMOCRATIC FOUNDATIONS, *supra* note 1, at 27-29.

¹²² However, especially in Europe, there is a growing trend for national judges to interpret domestic constitutional rights in line with international human rights standards. As a result, a national court may serve as an additional mechanism for backsliding rather than creating a buffer against it. See initial case studies in this paper, *supra* Part II.A-C.

¹²³ See *supra*, notes 25-30.

¹²⁴ See ICCPR, *supra* note 112, at art. 19.

right for men and women of marriageable age to marry, but this does not prohibit states from allowing same sex couples to marry.¹²⁵

Our theory of norm transmission from the international to the domestic level is such that backsliding can take place even though no state is obligated to reduce the level of human rights it provides. We acknowledge that when international norms require (even if in a non-binding fashion) a floor, the upward pull of those norms may be stronger than the downward pull we label backsliding. The magnitude of the difference depends on the particulars of how the norm affects domestic decisions. To the extent one believes that the formal legal commitment to meet the specified floor is what causes a change in behavior, backsliding is a relatively small problem.¹²⁶ To the extent one believes (as we do) that the signal provided by the international norm also provides political cover for those opposed to a particular expansion of rights, backsliding should be taken seriously.¹²⁷

A closely related issue is whether it is possible to minimize backsliding through careful drafting. Our sense is that while it may be possible to mitigate the backsliding effect, alternative forms of drafting that do not change the substantive content of the norm are unlikely to make much of a difference. Because our transmission mechanism relies on a political dynamic more than a legal one, technical and legal changes to a treaty or other instrument are not likely to have a large impact on how the content of the treaty is used in domestic political debates. One could, for example, make explicit that the requirements of a treaty constitute a floor and that states are free to exceed that floor. Indeed, this is often done, as noted above. We are not persuaded, however, that this has much impact on the ways the treaty is used by opponents of the right at issue.¹²⁸ They will still be able to point to the treaty as evidence that the international community does not believe that a particular behavior is required.¹²⁹

¹²⁵ See ICCPR, *supra* note 112, at art. 23.

¹²⁶ A weakness with such a position is that the theoretical explanations for why the formality of the legal commitment does the work are not especially satisfactory.

¹²⁷ In terms of the diagrams provided in Part III, the pull of the minimum requirement corresponds to the upward shift in the B curve while the empowerment of opposition to enhanced rights corresponds to the flattening of the slope of the same curve.

¹²⁸ Other mechanisms, such as the creation of optional protocols, the delegation of treaty interpretation to courts or other bodies that can increase standards over time, and the creation of regional agreements can all help reduce the risk of backsliding by creating multiple standards. We elaborate on these points in Parts IV.B.3 and V.A.2 below.

¹²⁹ Our view here reflects the political mechanism adopted throughout the paper. Alternative assumptions about how international norms influence domestic actions might lead to a different view of whether drafting is a promising tool to minimize backsliding.

Clearly, a norm that was both a floor and a ceiling would pull in both directions, but in that situation both forms of movement would be consistent with the conventional understanding of international human rights affecting domestic policy, and we would not term any of it backsliding.

Some human rights norms are better characterized as absolute prohibitions rather than a floor or a ceiling. The prohibition on genocide, for example, might fit this description, as might the prohibition on torture.¹³⁰ In both cases the international norm imposes a complete ban on the activity. At first glance it may seem that these norms are not subject to backsliding because there is no sense in which a country could be provide a higher level of protection than that required by a complete prohibition.

If, however, one views torture (or genocide) as simply a label that we assign to that which is prohibited, then things are more complicated. Leaving a particular practice off the list of actions that constitute torture, for instance, would open the door to backsliding. Although there is no way for a country to find itself with a stance toward torture that is stricter than the complete ban imposed by the Convention Against Torture or customary international law, it is possible for a country to define torture more broadly than these sources or to prohibit actions that go beyond the formal definition of torture.¹³¹ For example, while the United Nations Convention Against Torture applies only to torture by or with the consent or acquiescence of a public official,¹³² the Inter-American Convention to Prevent and Punish Torture also proscribes torture by private actors.¹³³

If the international norm fails to capture, for example, a particular interrogation technique, domestic actors arguing in favor of using that technique may be able to point to international standards in support of their claim that the appropriate trade-off between humane treatment of detainees and the need to carry out effective interrogations leaves room for the method in question. Indeed, when the

¹³⁰ Convention on the Prevention and Punishment of the Crime of Genocide, Dec. 9, 1948, U.N.T.S. 277, S. Exec. Doc. O, 81-1 [hereinafter Genocide Convention]; David Luban, *Calling Genocide by Its Rightful Name: Lemkin's Word, Darfur, and the UN Report*, 7 CHI. J. INT'L L. 303, 306 (2006) (arguing that the overly legalistic reading of genocide by international courts is entirely divorced from the way most people think of genocide as the "crime of crimes" deserving absolute prohibition).

¹³¹ Oona A. Hathaway et al., *Tortured Reasoning: The Intent to Torture Under International and Domestic Law*, 53 VA. J. OF INT'L L. 791, 791 (2012).

¹³² United Nations Convention Against Torture and other Cruel, Inhuman or Degrading Treatment or Punishment art. 1, Dec. 10 1984, 1465 U.N.T.S. 85 (entered into force June 26, 1987).

¹³³ Inter-American Convention to Prevent and Punish Torture provides an example of a treaty that expands rights perceived as absolute. Inter-American Convention to Prevent and Punish Torture art. 1, Dec. 9 1985, O.A.S.T.S. No. 67 (entered into force Feb. 28, 1987).

United States was debating this exact topic during the first decade of the 21st century, both sides argued that their preferred position was consistent with international law and norms.¹³⁴

Notice that this is not a point about the vagueness of the norm. It focuses instead on the boundaries of a prohibition, however clear they are. Genocide, for example, is defined in the Genocide Convention to include only actions targeted at “national, ethnical, racial, or religious groups.”¹³⁵ To the extent one believes that this convention has an upward pull on the behavior of states considering acts against such groups, it will also have a downward pull (backsliding) on states considering such acts against groups that are not listed, such as groups targeted because of their political ideology or their wealth.

2. Hard Law versus Soft Law

Our theory of backsliding rests on a particular view of the transmission mechanism through which international norms affect domestic policy. Specifically, international norms enhance the political power and influence of some domestic actors at the expense of others. This political transmission mechanism need not depend on the formal legal status of an international norm (i.e., binding international law, or hard law versus non-binding commitments, or soft law), implying that both soft law and hard law norms can have an impact on policy and lead to backsliding. Among other consequences, this means that state consent is not a necessary condition for a state to run the risk of backsliding.

To say that legal form is not determinative is not to say that it is irrelevant. The particular form that a norm takes may affect its influence over domestic actors, so formal legal status (among other things) may still be relevant. To the extent that formal international law is transmitted to the domestic sphere more powerfully, it will have a larger impact than other norms that do not have this legal status.

The general lesson, then, is that the impact of norms and the risk of backsliding depend on the force with which international policies or commitments get translated into domestic political influence. The formal legal status of a norm may matter, as may the issue area, the source of the norm, the enthusiasm of the domestic political system for foreign models, and so on.

3. Multiple Standards

One can imagine a situation in which there are multiple international norms at play in a single issue area. The simplest such case would feature several threshold levels of performance ranked from the easiest to the hardest to achieve.

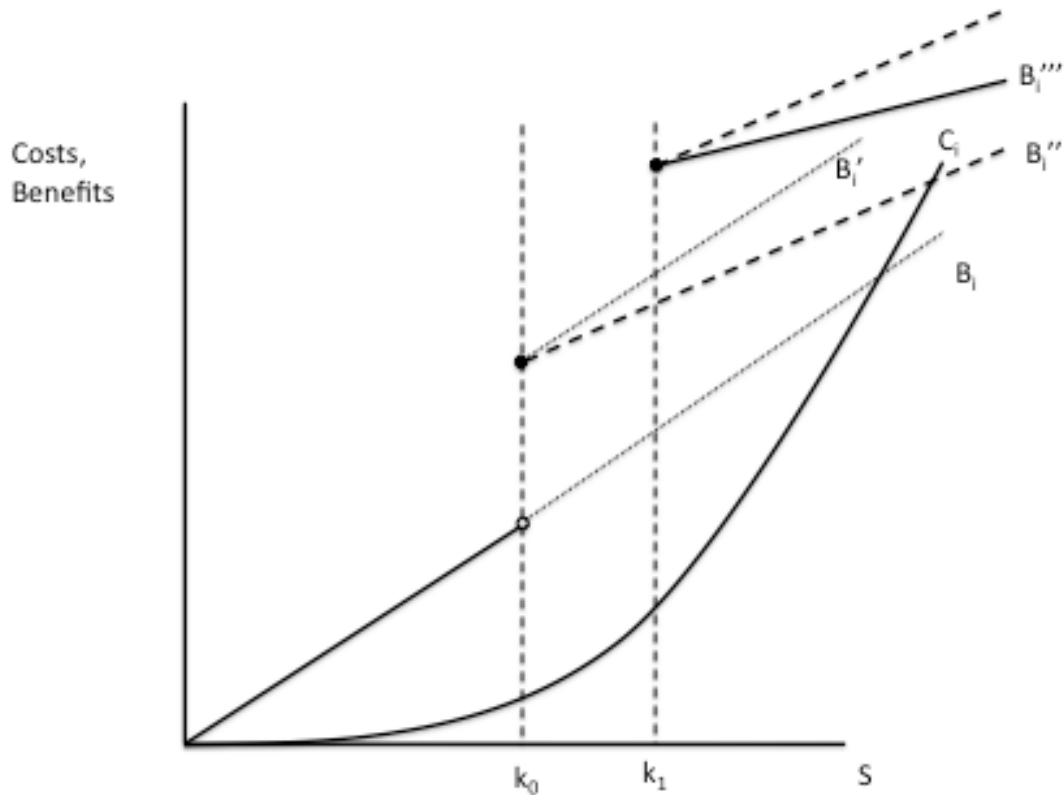
This corresponds to a variation of our model in which there are several discontinuities in the B curve, rather than just one. Instead of a single point, k , there would be several, which could be labeled k_0 , k_1 ,

¹³⁴ See generally Jinee Lokaneeta, *Torture Debates in the post-9/11 United States: Law, Violence, and Governmentality*, 13 THEORY & EVENT 1 (2010); see also Zachary R. Calo, *Torture, Necessity, and Supreme Emergency: Law and Mortality at the End of Law*, 43 VAL. U. L. REV. 1591, 1594-97 (2008-2009).

¹³⁵ Genocide Convention, *supra* note 130, at art. II.

k_2 , and so on. A country that satisfied the first threshold, k_0 would enjoy a discontinuous jump in benefits, just as in our base model. A country that also met the next threshold would see a further jump.

Figure 5: Human Rights Behavior with Multiple Standards



The results here are not so different from the case of a single threshold. To the left of the first threshold and the right of the last one, states will behave just as they do in our base model. States to the left of the first threshold will improve their performance if they are already sufficiently close to k_0 . If not, the norm will have no impact on them. All states to the right of the last threshold will backslide to some degree, as in the basic model.

The more interesting cases are states that are “in between” two thresholds, including states that were pulled up to a particular threshold from below. These states may face an incentive to move up to the next higher threshold in order to benefit from the discrete jump in benefits. They will also, however, experience some backsliding as a result of the lower threshold’s tendency to flatten the benefits curve, B .

The dynamic involved is clearest if we distinguish the two forces at work. First, the existence of a threshold to the left of where a country would otherwise be leads to backsliding, and that is what we observe if all thresholds are to the left of the country. The extent of the backsliding depends on the extent to which the slope of B is flattened, meaning the impact of the norm of on the state's benefits from providing higher levels of protection. Second, having a threshold to the right means that the state may wish to improve its performance enough to enjoy the discrete gain in benefits by meeting that threshold. The state's choice, then, is between moving to the higher threshold and backsliding in the direction of the lower one. By comparing these two options the state can determine which is more favorable. The closer it is to the higher threshold, the more likely it is to improve its performance. It is conceivable, then, that having two thresholds will cause some relatively high performers to improve while causing some relatively poor performers to backslide.

Several tools help generate multiple standards in the human rights field. One such tool is the optional protocol. In many human rights areas, the main treaty sets a relatively low and uncontroversial standard, while one or more optional protocols articulate higher and more controversial standards. For example, the ICCPR forbids the death penalty for juveniles and pregnant women specifically,¹³⁶ while Covenant's Second Optional Protocol abolishes the death penalty more generally.¹³⁷ Another way to create multiple standards involves delegation of the interpretation of key provisions to a treaty body or Court that can alter a standard's meaning over time.¹³⁸ For example, the Human Rights Commission interpreted the prohibition on sex discrimination in Articles 2 and 26 of the ICCPR to include sexual orientation, and concluded that an Australian province's prohibition on consensual same-sex activity violated the ICCPR.¹³⁹ Still another way to generate multiple standards involves regional agreements that define fundamental rights differently from universal agreements. We examine this point in detail in Part V.A.2 below.

¹³⁶ International Covenant on Civil and Political Rights art. 6, Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976).

¹³⁷ Second Optional Protocol to the International Covenant on Civil and Political Rights, Aiming at the Abolition of the Death Penalty, Dec. 15, 1989, 1642 U.N.T.S. 414 (entered into force Jul. 11 1991).

¹³⁸ Cite to Guzman article on Tribunals

¹³⁹ International Covenant on Civil and Political Rights arts. 17, 26, Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976). *Toonen v. Australia*, Views of the Human Rights Committee, U.N. Doc CCPR/C/50/D/488/1992 (1994), *available at* <http://www1.umn.edu/humanrts/undocs/html/vws488.htm>; Human Rights Council Res., Human Rights Council, 17th Sess., May 30-June 17, 2011, 17/19, A/HRC/17/L.9/Rev.1 (June. 17, 2011). Treaty negotiators who draft human rights as minimum standards may thus anticipate the expansion of rights over time, and the increased difficulty of backsliding. *But see* Laurence R. Helfer, *Overlegalizing Human Rights: International Relations Theory and the Commonwealth Caribbean Backlash Against Human Rights Regimes*, 102 COLUM. L. REV. 1832, 1851-60 (2002) (detailing the increased likelihood of backlash as human rights expand).

4. Vague Norms and Uncertainty

Up to this point the Article has assumed that both the international norm and the actions of states are observable. In practice, of course, the international norm itself may be contested or uncertain and policy implementation by a state may be difficult for others to observe. At one extreme, a norm that is devoid of content, meaning there is no common understanding of what is required, will have no effect on state behavior. Similarly, if the conduct of the state is entirely unobservable, including by its own citizens, then the norm will not affect behavior.

Whether uncertainty stems from a vague norm or difficult-to-observe government policies, the effects on the model are the same. Rather than a discrete jump upwards at k , the benefits curve has a more gradual increase in slope as it approaches k and then flattens once it passes to the right of k . This more continuous change in the curve reflects the fact that a state with policies close to that required by an international norm may be able to persuade some observers (including perhaps local citizens) that it is actually in compliance with the norm. Similarly, if the norm is uncertain, a state may be able to persuade some observers that its conduct is sufficient to satisfy the norm.

Because there is no discrete jump up in the benefits curve, some states may improve their performance and yet not move all the way to k . As one would expect, vague norms or difficult-to-observe policies generate less upward pull than clear norms and easy to observe domestic conduct. For high-performing states there can be a similar reduction in the risk of backsliding. If domestic policy is difficult to observe, a state that is in compliance, but close to the international norm may enjoy only a portion of the benefits of compliance. Graphically this would be represented by a benefits curve that is still flattening to the immediate right of the international norm. This has the effect of reducing the risk of backsliding for states close to the international norm because the cost of doing so is greater.

C. Other Theories of International Law's Influence

Several theories aim to explain how and why international law can induce governments to better protect human rights on the ground, despite a lack of coercive enforcement. Up to this point, we have focused on a mechanism we find particularly plausible. In our account, international norms shape the views of domestic interest groups, especially undecided swing voters. These international norms can be used to support a domestic policy proposal, and can help persuade ordinary citizens that a policy proposal is not a radical experiment but rather a mainstream, tried-and-true solution.¹⁴⁰ In this section, we explain why other prominent theories of international law's influence might also predict that backsliding can occur.

A caveat: there are some scholars who reject the possibility that international law can have any positive effects on state behavior under most circumstances. These scholars argue that strategic concerns and domestic pressures entirely unrelated to international law dominate leaders' calculations.¹⁴¹ If this view

¹⁴⁰ See *infra* Parts I and III.

¹⁴¹ See GOLDSMITH & POSNER, *supra* note 34, at 13.

is correct, and international law is too weak to influence state behavior for the better, then it is quite unlikely that international law could influence state behavior for the worse. We set aside this viewpoint as we believe the data show that international law and international norms can influence state behavior in a variety of circumstances.

One prominent theory of international law's influence, developed by scholars such as Abram and Antonia Chayes and Thomas Franck, states that treaties exert a broad "compliance pull" on states because states have a preference for compliance with international law.¹⁴² These leading international law scholars describe the conditions under which international law is most likely to exert its influence, but do not provide micro-foundations for their accounts. That said, if an international instrument has a sort of gravitational pull, it is plausible that this could operate in both directions: just as states that would otherwise perform below the international standard can be drawn upwards, states that would otherwise perform above this standard can be pulled down.

More recent scholarship builds on these theories and develops the micro-foundations of international law's influence. Drawing on Abram and Antonia Chayes and on Thomas Franck, Harold Koh argues that treaties and non-binding international instruments can, under the right conditions, change the views of national government officials. These officials interact with the international system, interpret it, and, ultimately, internalize its rules.¹⁴³ In political science, related viewpoints are most closely associated with the constructivist school of thought. Prominent constructivist theorists argue that following the adoption of an international human rights instrument, government leaders may be persuaded that their prior views were wrong, and their preferences may change as a result.¹⁴⁴ Alternatively, leaders may be socialized to view certain actions as unacceptable to the international community and face peer pressure to modify their behavior.¹⁴⁵

If leaders are persuaded or socialized to expand protection for a particular right when international norms call for extensive protections, we believe state leaders can be similarly persuaded or socialized to limit their protections when the international norm calls for more limited protections. Leaders might be persuaded, for example, that a lower level of protection for criminal defendants does not significantly increase the risk of placing innocent persons in jail while significantly reducing the costs of the

¹⁴² See Abram Chayes & Antonia H. Chayes, *On Compliance*, 47 INT'L ORG. 175, 176, 178 (1993); THOMAS M. FRANCK, *THE POWER OF LEGITIMACY AMONG NATIONS* (1990).

¹⁴³ See Harold H. Koh, *Why Do Nations Obey International Law?*, 106 YALE L.J. 2599, 2602, 2646 (1997) (reviewing Abram Chayes & Antonia H. Chayes, *THE NEW SOVEREIGNTY: COMPLIANCE WITH INTERNATIONAL REGULATORY AGREEMENTS* (1995); Thomas M. Franck, *FAIRNESS IN INTERNATIONAL LAW AND INSTITUTIONS* (1995)).

¹⁴⁴ See Jeffrey T. Checkel, *Why Comply? Social Learning and European Identity Change*, 55 INT'L ORG. 553, 577 (2003).

¹⁴⁵ See Ryan Goodman & Derek Jinks, *International Law and State Socialization: Conceptual, Empirical, and Normative Challenges*, 54 DUKE L.J. 983, 984-85 (2005).

administration of justice. Or even if leaders are not persuaded about the merits of this lower level of protection, they might simply follow their peers towards this lower level because they now consider it the appropriate level of protection for a modern state. Indeed, in their forthcoming book, Ryan Goodman and Derek Jinks recognize this risk. They write: “Emulation and mimicry associated with acculturation can also produce a ‘race to the middle.’ The concern here is that states that would otherwise aspire to heightened levels of human rights protection gravitate to lower expectations or standards of success.”¹⁴⁶

In short: a variety of theoretical mechanisms could lead to backsliding. We propose a particular theoretical model to clarify the logic of the argument, but alternative accounts of why states respond to international norms could also lead to similar reductions in rights.

Part V. Implications

Our theory of backsliding has important implications for states and others involved in designing human rights agreements. We first describe how backsliding clarifies some prominent yet confusing features of international human rights agreements and then apply the theory to highlight some practical implications for treaty membership.

A. Some Empirical Puzzles Explained

Three particularly puzzling features of international human rights regimes become clear through our theory: the reluctance of top performers to join international agreements, the proliferation of regional human rights systems, and the extremely high standards that characterize many human rights agreements. Our theory provides an explanation for each of these observed patterns.

1. The Behavior of High Performing States

We start with the observation, made in several empirical studies of human rights agreements, that top human rights performers are often reluctant to join international agreements.¹⁴⁷ Professors Emily Hafner Burton, Kiyoteru Tsutsui, and John Meyer note that while “it seems obvious that states with

¹⁴⁶ RYAN GOODMAN & DEREK JINKS, *SOCIALIZING STATES: PROMOTING HUMAN RIGHTS THROUGH INTERNATIONAL LAW* 99 (forthcoming 2013).

¹⁴⁷ See Andrew Moravcsik, *The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe*, 54 INT’L ORG. 217, 228-229 (2000); SIMMONS, *MOBILIZING supra* note 1, at 67-77 (arguing that governments that already respect rights can still face substantial domestic political costs when considering whether to join an international treaty).

negative human rights records might be somewhat more reluctant than others to ratify treaties that subject them to intensified internal and external criticism[,] the empirical data contradict this intuitive expectation.”¹⁴⁸ Similarly, Professor Oona Hathaway reports that “the states with the best human rights practices (and hence the best reputations) are often more reluctant to join human rights treaties than those with worse practices (and hence worse reputations).”¹⁴⁹ This pattern is puzzling because under existing theories of how human rights instruments interact with domestic policy making, the cost of joining international agreements that a country already complies with should be very low.¹⁵⁰

Our theory of backsliding, however, is consistent with this pattern and would predict it.¹⁵¹ That said, it is not the only explanation: several other factors could also account for the reluctance of high performing countries to join international agreements, such as the United States’ reluctance to join treaties more generally.¹⁵² Still, “top performers” might worry that the relevant international standards would create a downward pull for them and that joining a treaty may impede future progress or even cause a reversal of human rights gains. Indeed, sometimes advocates for minority groups oppose the ratification of international treaties intended for their benefit, by making closely related arguments about backsliding. For example, advocates for Canada’s First Nations successfully lobbied against the ratification of ILO Convention 169 on Indigenous People. They worried that Canada would settle for the low standards of rights protected by the Convention, and that the Canadian government would oppose increasing these standards in the future by referencing this Convention.¹⁵³

¹⁴⁸ See Emilie M. Hafner-Burton, Kiyoteru Tsutsui & John W. Meyer, *International Human Rights Law and the Politics of Legitimation: Repressive States and Human Rights Treaties*, 23 INT’L SOC. 115, 117 (2008).

¹⁴⁹ Oona A. Hathaway, *The New Empiricism in Human Rights: Insights and Implications*, 98 PROC. ANN. MEETING (AM. SOC’Y INT’L L.) 206, 209 (2004).

¹⁵⁰ See George W. Downs, David M. Roche & Peter N. Barsoom, *Is the Good News About Compliance Good News About Cooperation?*, 50 INT’L ORG. 379, 393-99 (1996).

¹⁵¹ See discussion *infra* Part V.A.1.

¹⁵² Oona A. Hathaway et al., *The Treaty Power: Its History, Scope, and Limits*, CORNELL L. REV. (forthcoming 2013). See, e.g., DAVID SCHEFFER, *ALL THE MISSING SOULS* 166-67 (2012) (describing the United States’ reluctance to ratify the Rome Statute). Another reason for hesitancy from already-complying countries may be because many of the worst rights violators do ratify. Oona A. Hathaway, *The New Empiricism in Human Rights: Insights and Implications*, 98 PROC. ANN. MEETING (AM. SOC’Y INT’L L.) 206, 209 (2004); Oona A. Hathaway, *Why Do Countries Commit to Human Rights Treaties?* 51 J. OF CONFLICT RESOL. 588, 588 (2007).

¹⁵³ See, for example, the statements of Judith Sayers, a Nuu-Chah-Nulth lawyer from British Columbia, in her article titled “ILO Convention Must Be Stopped” (stating that “If Indians in Canada ask the Canadian government to ratify this convention, it will be seen as a consent to a very low standard of rights. The Canadian government will use this as a knife in our backs as we continue our work internationally. They will maintain they do not have to raise the standard higher, because indigenous people agreed to the

2. Regionalism

A second puzzle in international human rights law concerns the strength and number of regional human systems. Though human rights practices vary from one region to another, supporters often emphasize the universal nature of human rights. Indeed, human rights claims derive much of their moral power from the belief that they are universal: that all humans deserve to enjoy certain fundamental rights because they are human.¹⁵⁴ The assertion of universality is also an essential part of the argument that demanding compliance with human rights norms is not an unjustified interference in the domestic affairs of the state.¹⁵⁵

The rhetoric of universality is somewhat incongruous with the fact that many of the most important human rights institutions are regional rather than universal. The ECtHR, for instance, has been called “the crown jewel of the world’s most advanced international system for protecting civil and political liberties” as it reviews tens of thousands of cases annually, protecting the rights of 800 million people.¹⁵⁶ Similarly, the Inter-American Court of Human Rights has great influence on Latin American governments’

convention. That was the strategy behind revising the ILO convention. Now many states are referring to it, stating this is what a Declaration of Indigenous Rights should look like.”). Judith F. Sayers, *ILO Convention Must be Stopped*, 8 WINDSPEAKER 4 (1990), available at <http://205.186.158.152/publications/windspeaker/ilo-convention-must-be-stopped-0>. This convention was criticized by several other indigenous groups as well, and these passed a resolution opposing its ratification. See also Sharon Venne, *The New Language of Assimilation: A Brief Analysis of ILO Convention 169*, 2 WITHOUT PREJUDICE: THE EAFORD INTERNATIONAL REVIEW OF RACIAL DISCRIMINATION 53, 66-67 (1989); Douglas Sanders, *Developing a Modern International Law on Indigenous People* 14 (1994), available at <http://www.ubcic.bc.ca/files/PDF/Developing.pdf>. Related concerns about assimilation pressures had been raised with respect to ILO Convention 107, the predecessor of ILO Convention 169, and may “have prevented its ratification by some more progressive countries which might otherwise have wished to apply its protective provisions.” Lee Swepston, *A New Step in the International Law on Indigenous and Tribal Peoples: ILO Convention 169 of 1989*, 15 OKLA. CITY U. L. REV. 677, 683 (1990).

¹⁵⁴ See Jack Donnelly, *The Relative Universality of Human Rights*, 29 HUM. RTS. Q. 281, 282 (2007); JACK DONNELLY, *UNIVERSAL HUMAN RIGHTS IN THEORY AND PRACTICE* (2003).

¹⁵⁵ See generally Jack Donnelly, *State Sovereignty and International Intervention: The Case of Human Rights*, in *BEYOND WESTPHALIA? STATE SOVEREIGNTY AND INTERNATIONAL INTERVENTION* (Gene M. Lyons & Michael Mastanduno eds., 1995).

¹⁵⁶ Larry Helfer, *Redesigning the European Court of Human Rights: Embeddedness as a Deep Structural Principle of the European Human Rights Regime*, 19 EUR. J. INT’L L. 125, 125 (2008).

choices.¹⁵⁷ And the more recently established African Court of Human and People’s Rights has a particularly wide-ranging mandate.¹⁵⁸ The importance of regional human rights systems in Europe, the Americas, and even Africa is puzzling as it is hard to develop forceful normative claims that derive from residence in a particular continent. Moreover, a proliferation of human rights tribunals – a symptom of regionalism – carries important risks, such as conflict and contradictions in evolving human rights doctrine, and overlapping jurisdiction for individual cases, as prominent international lawyers have noted.¹⁵⁹

Our theory of backsliding helps explain this puzzle by suggesting that different standards for countries with different capabilities reduce the risk that top performing countries will get stuck to a low universal standard.¹⁶⁰ If one assumes that human rights are universal and that creation or clarification of international norms serves only to improve the behavior of low-performing states, a more sensible policy would be to establish a single set of global norms. If one were concerned that establishment of highly protective norms would fail to influence low-performing states, the solution is not regionalism but tiered human rights norms. One could identify the “true” human rights norm (by assumption a highly protective one) along with a series of lower “intermediate” norms. States that managed to meet the intermediate thresholds would be recognized as falling short of the ideal, but still making good progress.

The theory of backsliding helps explain why states have opted for regional human rights norms rather than limiting themselves to universal norms. To the extent backsliding is a concern, regional norms can be tailored to achieve maximum benefit for the affected countries. The maximization of benefits must consider both the potential to generate upward pull on low performers and the risk of causing backsliding by high performers. A regional norm is crafted for a set of countries that feature less diversity in human rights conduct than would be the case for a global norm. If, for example, countries in the region are high performers from a global perspective, the regional norms can be ambitious. Even demanding norms will be relatively close to the conduct of the relevant states, so there is reason to hope that the norms will exercise some upward pull. The high norms will also reduce the risk of backsliding because they will be more protective than the existing practice of many states. A norm that might be globally appropriate, on the other hand, would impose higher backsliding costs. This description is consistent with what we observe within the European human rights system as well as

¹⁵⁷ See Lea Shaver, *The Inter-American Human Rights System: An Effective Institution for Regional Rights Protection?*, 9 WASH. U. GLOBAL STUD. L. REV. 639 (2010).

¹⁵⁸ See Cesare P.R. Romano, *The Proliferation of International Judicial Bodies: Pieces of the Puzzle*, 31 INT’L L. & POL. 709, 722 (1999).

¹⁵⁹ See Benedict Kingsbury, *Foreword: Is the Proliferation of International Courts and Tribunals a Systemic Problem?*, 31 INT’L L. & POL. 679, 681-684 (1999).

¹⁶⁰ See discussion *infra* Part V.A.2.

aspects of the Americas where “the Inter-American Court of Human Rights [has] perhaps imposed higher standards on states than has the ICJ” on questions of state responsibility to take affirmative measures of protection.¹⁶¹

If a region is populated by relative low-performing states, less ambitious norms will be more effective because they will be more effective at generating upward pull and will have only modest backsliding costs.¹⁶² So, for example, a global standard that is set at a high level will likely be ignored by war-torn and impoverished states whose citizens have a great need for just such protections.

3. Aspirational Norms

A third puzzle concerns the very high standards many human rights treaties sometimes enshrine. Other international agreements, including many military, trade, and environmental agreements, set modest standards that require only small adjustments for participating states.¹⁶³ In the human rights arena, in contrast, international agreements frequently set very high standards, “above a level that many participating countries can or want to comply with immediately or within the foreseeable future.”¹⁶⁴ For example, according to critics, several “CEDAW provisions are plainly unrealistic . . . with the gross burden of providing basic needs to their citizens still unfulfilled, many developing nations hesitate to take on the additional responsibility of un-stereotyping women's and men's roles, as Article 5 of the convention suggests.”¹⁶⁵ International treaties that call for social and economic rights are similarly criticized for setting impossibly high standards. For example, Article 11 of the ICESCR calls for “the right of everyone to an adequate standard of living for himself and his family, including adequate food, clothing and housing,” a right that many developing countries cannot easily fulfill.¹⁶⁶

¹⁶¹ See Kingsbury, *supra* note 139, at 681.

¹⁶² Indeed, the follow-on protocols of the ECtHR, the Inter-American Court of Human Rights, and the African Court of Human Rights have contained more stringent rights protections than their original treaty formations as well as equivalent global conventions. This model of a primary convention with supplementary protocols may function to constrain backsliding as we explain in Part IV.B.3 above. It is less clear whether overlapping international laws on the same subject matter, such as violence against women, children's rights, disappearances, discrimination, and protection of people with disabilities, impede or aid backsliding.

¹⁶³ See Downs, Rocke & Barsoom, *supra* note 150, at 388-392.

¹⁶⁴ Eric Neumayer, *Do International Human Rights Treaties Improve Respect for Human Rights?*, 49 J. CONFLICT RESOL. 925, 928 (2005).

¹⁶⁵ Meghana Shah, *Rights Under Fire: The Inadequacy of International Human Rights Instruments in Combating Dowry Murder in India*, 19 CONN. J. INT'L L. 209, 222 (2003).

¹⁶⁶ International Covenant on Economic, Social and Cultural Rights art. 11, *opened for signature* Dec. 16, 1966, 993 U.N.T.S. 3 (entered into force Jan. 3, 1976).

Critics worry that very high standards dilute their status of human rights agreements as legally binding obligations and limit their potential to influence states. For example, philosopher Onora O’Neill worries that such aspirational standards dilute the value of human rights generally, and lead us to “accept that where human rights are unmet there is no breach of obligation, nobody at fault, nobody who can be held to account, nobody to blame and nobody who owes redress.”¹⁶⁷ It is unremarkable to witness the conclusion of a human rights agreement that everyone involved expects to be routinely violated. It is certainly true that states fail to meet their obligations in other areas, but nowhere is the disconnect between promise and behavior as stark as in the human rights area.

Our backsliding theory helps explain why this curious acceptance of empty promises exists in human rights agreements more than elsewhere. If states are concerned about backsliding, they have an incentive (all else equal) to prefer high standards over low standards in a way that is not explained by existing understandings of international human rights. A demanding norm—perhaps even one that exceeds every country’s existing practice—reduces the associated backsliding costs because it reduces the number of states for which the international norm is less demanding than their *ex ante* conduct. In the most extreme cases, a description that might fit the CEDAW, the number of such countries may be reduced to zero, which also eliminates the risk of backsliding. Very high standards both reduce the upward pull of human rights agreements, and also reduce the risk of backsliding among top performers.

B. Treaty Membership

Backsliding has important implications for the design of international human rights agreements. We have discussed some of these already, including the question of how strong the protections of human rights agreements should be,¹⁶⁸ and the legal form they should take.¹⁶⁹ In this section we explore an additional set of implications, concerning membership in human rights treaties.

Our theory provides one reason why the presence of high-performing states may tend to raise the level of standards in the treaty. These high-performing states presumably have a preference for high human rights standards—otherwise they would not be high performing—and this should have some upward pull on the content of an agreement. That said, it is hard to explain why a high-performing state would choose to expend political capital to increase the standard of a human rights agreement. Absent backsliding, after all, the agreement will have no direct effect on a state that is already in compliance with its terms.

¹⁶⁷ Onora O’Neill, *The Dark Side of Human Rights*, 81 INT’L AFF. 427, 430 (2005).

¹⁶⁸ See *infra*, Part IV.B.1

¹⁶⁹ See *infra*, Part IV.B.2.

It is true that the country may prefer a higher standard because it cares about the welfare of citizens of foreign countries. Even when treaties are effective, however, it is not certain that higher standards lead to better human rights outcomes. In our model, and in many models of the relationship between international norms and domestic policy, only countries close to the international norm are pulled upward. As a result, as we explain above, it may be better to have a lower standard so as to pull up the worst behaving states.¹⁷⁰

Backsliding provides an alternative (or perhaps additional) explanation. High performing states do not seek higher standards out of a concern for the treatment of foreign individuals, or at least not only for this reason. They prefer a higher standard so as to reduce and perhaps eliminate the extent to which they are exposed to backsliding. Even if a state is unable to get the international norm to be at the level it provides to its own citizens, the closer the norm gets, the less backsliding the state will face.

One implication of the above discussion is that having high-performers at the table may distort the resulting agreement toward higher-than-optimal standards. More generally, the identity of the states involved will influence the content of the agreement. This is, of course, both obvious and well understood, but an appreciation of backsliding changes the way we think about the preferred mix of participants.

Once we realize that the identity of states involved in norm-formation matters, it is natural to think about how one might manage participation to get better results. One point has already been mentioned—the appeal of regionalism. If states in a particular region are similar to one another in their human rights conduct, it is easier to create a norm that will have positive effects with a minimum of backsliding. Put another way, a group of similar states committed to an improvement in their human rights regimes may generate a more valuable agreement if higher-performing states are not present or involved in the process.¹⁷¹ This is because high-performing states may push for norms that are too high, in the sense that they do not affect enough low-performing states, in an effort to minimize the consequences of backsliding.

For human rights advocates, however, the absence of high-performing states may present its own disadvantages. For an agreement to generate an upward pull on existing policies requires, of course, that the resulting norm be more protective of human rights than the practices of the states. A group that is more uniform in its behavior may not be enthusiastic about changing that behavior. Without a

¹⁷⁰ See *infra* Part III.A, especially text accompanying notes 94-98.

¹⁷¹ There are many influences here, and we do not attempt to capture them all. The incentive to exclude high-performing states described above, for example, may be partially or fully offset by the fact that the presence of a more diverse set of states might give the norm a strong claim to universality and, therefore, make it more forceful. Because there are many influences that we cannot engage in this Article, the points made in the text should be viewed as additional factors, not the entire story.

(relatively) high-performing state present, one wonders where the pressure for a high standard will come from.

We thus have the curious, and perhaps somewhat uncomfortable result that an agreement capable of exerting an upward pull on states may require the presence of at least some high-performing states, but having those states at the table risk pulling the norm above the level that would yield the greatest improvements to human welfare. One possibly effective strategy, then, is to have high-performing states encourage (or pressure) low-performing states to enter into an agreement that ultimately does not include those same high-performing states.¹⁷² There are obvious reasons to be suspicious of such an approach, including the possibility that the high-performing states will impose requirements that serve their own interests rather than those of the participating states. For this reason, we do not advance this as a normatively desirable proposal. Rather, we raise it as one potential implication to point how an appreciation of backsliding can change our normative conclusions.

The issue of membership is still more complicated once we acknowledge that international norms do not have to be consent-based. Norms may emerge from declarations of international organizations that do not require unanimous support (e.g., United Nations resolutions) or that do not include all states (e.g., an OECD agreement). They also may be the result of agreements that are nonbinding, such as the UDHR. The list of potential sources of an international norm is very long: it includes treaties, non-binding agreements, customary international law, other states' practice, decisions of tribunals, and more. This, of course, is a feature of international human rights norms even in the absence of backsliding, but it plays out differently with backsliding. If one ignores backsliding, non-consensual human rights norms (for better or worse) have the potential to have an upward pull on low-performing states, even if those states are not involved in the generation of the norm or have not consented to it.

When we consider backsliding, the non-consensual aspect of human rights norms works in the other direction. A high-performing state that objects to a norm because it risks causing backsliding cannot avoid it simply by refusing to join a relevant treaty or organization. That state may be influenced by the norm even if it does not consent. If we continue to assume a correlation between powerful states and high-performing states, there is a risk that high-performing states will find ways to undermine efforts to create norms that pose the risk of backsliding.

In this Part we have explored important implications of our theory of backsliding for the design of human rights regimes. Our theory helps explain some puzzling features of human rights regimes, including the reluctance of top performers to join international agreements whose standards they already meet, the proliferation of regional human rights systems, and the very high standards many

¹⁷² We have previously discussed the fact that states may be affected by international norms even if those norms are not legally binding. That reality applies here, so the high-performing states cannot fully escape the risk of backsliding. By refusing to join the agreement, however, they weaken the transmission of the norm to their domestic systems and, therefore, reduce the associated risk.

human rights agreements specify. That said, our theory also opens up new questions about the best design for human rights treaties, including questions about who should be included and who should be excluded, that are not easy to answer.

Part VI. Conclusion

Human rights violations remain a widespread phenomenon in our world. International law in general, and international human rights law in particular, seeks to influence, in a positive manner, the ways in which states treat their citizens.

The study of human rights law is, in part, the study of how international legal rules can be used more effectively to advance this objective. As a normative matter, backsliding may be seen as an undesirable feature of the international system, but that in no way diminishes the importance of understanding how it works. The stakes involved are too great to allow room for complacency with respect to our study. We must seek the best possible understanding of how international norms affect behavior, even if they at times do so in ways we do not like.

This Article makes it clear that we should not simply assume that when human rights standards improve state practices in some parts of the world, they always allow other countries—countries that currently perform above those standards—to continue on their upwards trajectories. As we have shown, there is a risk that low international standards will empower domestic opponents of human rights and lead to a reduction of rights in high-performing countries.

It is not our purpose to undermine any ongoing human rights efforts, and we not believe this Article does so. We believe, rather, that an understanding of the risk of backsliding allows both advocates and policy makers to think more clearly about how human rights norms will affect the outcome that really matters—the way in which humans are treated.

Backsliding is a novel argument in the human rights literature. Therefore, in this Article we have aimed to establish that backsliding is theoretically plausible, explore the conditions that would magnify this problem, and identify some tools that might mitigate its consequences. Much future work remains to be done. Some of this work will hopefully be empirical and will help identify the magnitude of the problem and the issue areas and countries where backsliding poses the greatest risk. Other work, we hope, will be pragmatic and policy-oriented, as there are important ways to design international standards, draft international instruments, and design advocacy campaigns to better balance the possibility of improving human rights in some countries against the risk of backsliding in others.

