

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Baekkwan Park

Date

The Politics of Naming and Shaming:
Informational Determinants of Human Rights Reporting

By

Baekkwon Park
Doctor of Philosophy

Political Science

Jeffrey K. Staton Ph.D.
Advisor

David R. Davis Ph.D.
Advisor

Jennifer Gandhi Ph.D.
Committee Member

Jacob Eisenstein Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

The Politics of Naming and Shaming:
Informational Determinants of Human Rights Reporting

By

Baekkwan Park
Master of Arts, Emory University, 2012
Bachelor of Arts, University of Nebraska-Lincoln, 2006

Advisor: Jeffrey K. Staton Ph.D.

Advisor: David R. Davis Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Political Science
2015

Abstract

The Politics of Naming Shaming: Informational Determinants of Human Rights Reporting By Baekkwon Park

This dissertation focuses on the tasks of information collection and state targeting. It centers on two empirical patterns of HRO advocacy. Why are some states frequently shamed for human rights violations despite having reasonable human rights records while many states are rarely criticized despite being serious violators? Why are states shamed differently by different HROs for precisely the same human rights violations? I argue that the process by which HROs collect information is critical to understanding these patterns. HROs that largely rely on third parties to generate information tend to focus on states with relatively good records; and, they shame these states severely. In contrast, HROs that directly engage states tend to focus on the worst human rights violators but they are less likely to harshly criticize these states. The reason is that HROs that indirectly monitor via third parties receive a relatively large amount of information from states in which there is a well-organized civil society prepared and willing to expose violations, and these states generally have reasonable human rights records. HROs that directly engage states focus on serious abuses; however, in order to gain access, HROs are more likely to negotiate the information that they will release publicly. I conclude my dissertation by explaining why HROs choose particular information collection strategies, focusing on variation in the material costs of investigative missions and the political costs of potentially risking an image of impartiality. To evaluate the empirical implications of my argument, I have compiled a comprehensive dataset on the shaming behavior of over 90 HROs listed in the Yearbook of the International Organizations and the United Nations. Utilizing text analysis techniques based on NLP, I have examined and processed nearly 40,000 human rights reports. I find that by choosing an indirect approach, HROs tend to have far more information about abuses in places with strong civil societies and they are likely to both over-report on states with generally good records and under-report on states with relatively bad records. These HROs criticize states harshly and they do not withhold information. HROs that directly engage states; however, tend to focus on states with the worst human rights conditions, but they are less likely to criticize states harshly.

The Politics of Naming and Shaming:
Informational Determinants of Human Rights Reporting

By

Baekkwan Park
Master of Arts, Emory University, 2012
Bachelor of Arts, University of Nebraska-Lincoln, 2006

Advisor: Jeffrey K. Staton Ph.D.

Advisor: David R. Davis Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Political Science
2015

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my two advisors, Dr. Jeffrey Staton and Dr. David Davis for their continuous guidance, patience, and encouragement. They have given me their unmitigated support even when my paths have seemed blurred. They provided astute and helpful feedback at multiple stages of this project. Although I have not fully absorbed their sterling examples of clear and concise writing, they have all provided worthy standards to aim for. I am very grateful to Dr. Jennifer Gandhi who consistently encouraged my every step and inspired me to continue keeping focused and balanced throughout. Her encouraging attitude kept me going. I would like to offer sincere thanks to Dr. Jacob Eisenstein. Much of what I understand about NLP came from his excellent guidance. He introduced me to NLP and has opened up a world of computational linguistics that will intrigue me for the rest of my career. I also owe a debt of gratitude to Sandeep Soni for his unending patience in answering my infinite questions about computer programming. Without his help, I could not have accomplished so much so quickly. I would like to extend many thanks and gratitude to Dr. Amanda Murdie for her continual faith and encouragement and for being my visionary light at the end of the tunnel. I also offer sincere thanks to Dr. David Forsythe who fostered my nascent interest in human rights. I am also very grateful to the entire faculty of the Department of Political Science at Emory, especially to Dr. Dan Reiter and Dr. Kyle Beardsley. I have been fortunate to have wonderfully supportive colleagues and friends at Emory. I thank Nicole Baerg and Jake Rick for their caring advice and above all, their friendship. I am also genuinely thankful to Denise Brubaker who created a home for me in the department and in Atlanta.

Contents

1	Introduction	1
1.0.1	Patterns of Shaming	3
1.0.2	Argument Summary	8
1.0.3	Outline of the Dissertation	10
2	A Theory of Naming and Shaming	13
2.1	International Human Rights and Human Rights Organizations	13
2.1.1	Human Rights Organizations	14
2.2	Existing Studies on HROs	19
2.2.1	First Generation of HRO Research: Normative Analysis	19
2.2.2	Second Generation of HRO Research: Impact Analysis	22
2.2.3	Third Generation of HRO Research: Shaming Mechanism	24
2.3	A Theory of Naming and Shaming	27
2.3.1	Human Rights Violations and Essential Challenges of Information Gathering	29
2.3.2	Indirect Information Gathering	35
2.3.3	Direct Information Gathering	46
2.4	The Choice of Information Gathering Approach	56
2.4.1	Material Costs	56
2.4.2	Political Costs	58
2.5	Conclusion	60
3	Measuring Naming and Shaming	62
3.1	Introduction	62
3.2	Existing Approaches	63
3.2.1	Obtaining Sources	63
3.2.2	Data Coding: Hand-Coded vs. Machine-Coded Data	67
3.2.3	Critiques	71
3.3	An Alternate Approach	73
3.3.1	Data	74
3.4	Measuring Naming and Shaming	78
3.4.1	Shaming Counts	78
3.4.2	Linear-Chain Conditional Random Fields	79
3.4.3	Application: Identifying HROs' Targets	81

3.4.4	Shaming Intensity	86
3.5	Conclusion	101
4	Naming and Shaming Counts Models	103
4.1	Introduction	103
4.2	HROs and Shaming Targets	103
4.3	Human Rights Shaming: Counts	105
4.3.1	Dependent Variable	106
4.3.2	Independent Variables	109
4.4	Estimation Method	116
4.4.1	Negative Binomial Model	118
4.4.2	Results	122
4.4.3	Alternative Statistical Model	127
4.5	Conclusion	133
5	Naming and Shaming Intensity Models	135
5.1	Introduction	135
5.2	Human Rights Shaming: Intensity	136
5.2.1	Dependent Variable	139
5.2.2	Independent Variables	140
5.3	Estimation Method	143
5.3.1	Tobit Regression Model	145
5.3.2	Results	146
5.4	Conclusion	150
6	Choice of Information Gathering Approach	152
6.1	Introduction	152
6.2	HROs and Two Information Gathering Approach	153
6.2.1	Dependent Variable	154
6.2.2	Independent Variables	155
6.2.3	Control Variables	156
6.3	Estimation and Results	159
6.4	Conclusion	163
7	Conclusion	164
7.1	Summary of The Dissertation	167
7.2	Implications and Future Research Agenda	169
7.2.1	Impact Analysis	169
7.2.2	HROs and Media Interaction	170
A	Appendix	173

List of Tables

1.1	Top 5 Targeted States by Five HROs (1995-2004)	6
2.1	Amnesty Estimates and NHRCK Estimates	42
2.2	Material vs. Political Costs	60
3.1	Classifiers Experiments: Approaches and Accuracy Rates	100
4.1	Descriptive Statistics Summary (Count Model)	117
4.2	Determinants of Shaming Counts	126
4.3	Expected Effect of Approach and Civil Society	127
4.4	Determinants of Shaming Counts (Zero-Inflated NB Estimation) . . .	132
5.1	Descriptive Statistics Summary (Intensity Model)	143
5.2	Determinants of Shaming Intensity	149
6.1	Descriptive Statistics Summary (Approach Model)	158
6.2	Determinants of Information Gathering Approach	160
A.1	Determinants of Shaming Intensity: OLS with RE	180

List of Figures

1.1	Human Rights Conditions and Shaming Counts	5
2.1	Growth in the Presence of Human Rights INGOs	14
2.2	The Boomerang Model	20
2.3	Indirect Information Gathering	37
2.4	Direct Information Gathering	47
3.1	Predicted Shaming Intensity and Manual Ratings	100
3.2	Regularization and Correlation Coefficients	101
4.1	Average HRO Shaming Counts (1991-2011)	108
4.2	HROs Shaming Counts and the Level of Civil Society	123
4.3	HROs Shaming Counts and the Level of Human Rights Conditions .	124
5.1	HROs Shaming Intensity and China	147
6.1	The Effect of Political Cost and Budget on Pr (Indirect Approach) . .	161
6.2	First Difference Measuring the Effect of Political Cost on Pr(Indirect Approach) at All levels of Budgets	162

Chapter 1

Introduction

According to the Korean Institute of National Unification (KINU 2000, 118), there were approximately 200,000 political prisoners held in 10 concentration camps across North Korea in 2000.¹ Under the North Korean policy of “re-education” or “rehabilitation,” every detainee was forced to do heavy labor (primarily coal mining) from 5:30 am to 9:00 pm. They were prohibited from walking in groups of more than two persons. Power supply was restricted to three to five hours a day. A fuel shortage limited heating in their rooms and food rationing was purposefully irregular. Prisoners commonly suffered from pneumonia, tuberculosis, and pellagra due to malnutrition and heavy labor, yet no medical services were made available (KINU 2000, 130). Torture and extrajudicial killings were frequent (KINU 2000, 24).

¹In the Hamkyung-bukdo province at Hweiryong (50,000 prisoners), Chungjin (15,000) Hwasung (20,000); in the Hamkyung-namdo province at Danchon (10,000), Duksong (10, 000), Yoduk (50,000); in the Pyongan-namdo province at Keichon (15,000), Bukchang (5,000), in the Pyongan-bukdo province at Chonma (15,000) and in the Jakangdo province at Dongshin (17,000). In 1998, the US State Department reported that 150,000 – 200,000 political prisoners were being held in concentration camps. South Korea’s National Intelligence Agency estimated 207,000 political prisoners (KINU 2000, 122)

The incarceration of political prisoners in North Korea reflects a clear violation of international human rights norms; however, highly autocratic states are not the only violators. We know that even the most consolidated democracies commonly commit violations of human rights, especially when leaders perceive threat (Conrad and DeMeritt 2011, Conrad and Moore 2010, Davenport, Moore and Armstrong 2007). Human rights abuse is a global phenomenon. It is also an extraordinarily difficult problem to address, precisely because of the secretive nature in which violations are carried out. Abuses are difficult to detect, to observe and, even more difficult to verify (Roth 2004, Simmons 2009). Especially as states democratize, they commonly transform their use of repression from overt to covert strategies (Conrad 2010, Cunningham 2004, Frank 1990, McPhail, Schweingruber and McCarthy 1998). Consequently, although interested publics may have a rough sense of a state's human rights practices, the facts are far from complete. Thus, an essential problem of human rights promotion is informational. If we do not know what is happening on the ground, we cannot possibly hope to hold leaders accountable, much less change state behavior.

Resolving this informational problem is a primary function of international human rights NGOs and IGOs (what I will refer to as HROs). They collect, assess, and publicize information about human rights conditions in states in order to bring pressures and persuade governments to alter their behavior and comply with international standards. They try to “convert moral authority into ideational power” (Keck and Sikkink 1998, Risse 2002). These activities are commonly referred to as

“naming and shaming” (Hawkins 2002, 20).²

My dissertation considers this aspect of the international process by which human rights are promoted. I develop a theory of naming and shaming that accounts for two notable empirical puzzles in the literature and sheds light on the conditions under which we should expect HROs to change the human rights behavior of states by publicly identifying violations.

1.0.1 Patterns of Shaming

The most prominent theory of naming and shaming derives from Keck and Sikkink (1998)’s Transnational Advocacy Networks (TANs), also known as “boomerang model,” which highlights the role of a dense international network of HROs. In this model, domestic human rights activists or groups seek to change repressive state behavior. When they are blocked internally by their government, they seek contacts with outside actors, hoping that international attention will lead to international condemnation of, or action against, the government (Keck and Sikkink 1998, Ramos, Ron and Thoms 2007, Ron, Ramos and Rodgers 2005). With the information that domestic groups provide, international actors are able to coordinate their efforts via a dense network of like-minded activists. They pool their resources and bring coherent pressure on repressive governments, in part via the naming and shaming process. In this way, state efforts to block domestic mobilization produces an international mobilization that can induce the change domestic groups were seeking

²According to *the Yearbook of International Organizations* (2009-2010), there are more than 500 HROs and majority of them (74 percent) publish human rights information regularly, that is, engage in naming and shaming.

in the first place.

Empirical work on naming and shaming has largely focused on evaluating its effectiveness. Human rights practitioners and scholars agree that spotlighting abuses has important consequences for reducing human rights abuses (Brysk 1993, Hawkins 2002, Hertel 2006, Keck and Sikkink 1998, Risse and Sikkink 1999, Roth 2004), but skeptics have pointed out a gap between belief and empirical evidence. In particular, studies that have evaluated the effects of naming and shaming from a statistical perspective find that on the whole, such efforts are not uniformly effective (Hafner-Burton 2008), though the sources of the shaming matters (Franklin 2008) or the presence of domestic NGOs and third party states (Murdie and Davis 2012). This debate over effectiveness is important, but it has obscured empirical patterns that raise critical questions about the process by which human rights information emerges and about a critical feature of the boomerang model itself.

In 2000, recognizing the horrendous human rights violations in the concentration camps in North Korea, Amnesty International (AI) issued reports and press releases about the North Korean government four times (AI 2000). Yet in the very same year, AI criticized the United States sixty-seven times (extensive background reports and press releases), primarily for its super-maximum security prison in Colorado and police brutality. There is little doubt that many practices of the American criminal justice system violate international human rights norms, but it is far from clear that the United States deserved to be criticized seventeen times as much as North Korea. A cursory glance at the global naming and shaming suggests a number of puzzling

patterns.

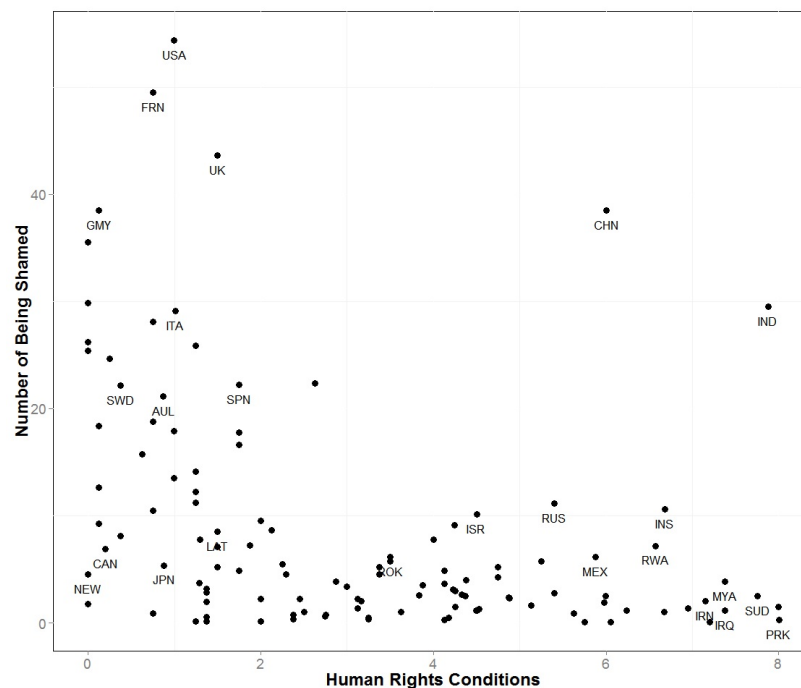


Figure 1.1: Human Rights Conditions and Shaming Counts
Source: Murdie (2009)

Figure 1.1 plots the number of times a state was shamed by HROs on Cingranelli and Richards (2014)'s measure of general human rights practices from 1992 to 2004. While many of the states that were frequently shamed possessed poor human rights records, many were liberal democracies with good, though imperfect human rights records (e.g. U.S., U.K., France, Germany). And many states with terrible records were not shamed at all. Critically, if there is a pattern at all, it would appear that states are more likely to be shamed when their human rights records are good. If the international community is supposed to identify the most serious violators of human rights, these data are troubling.

Further, despite strong incentives, HROs are not always coordinated in their activities. For example, Table 1.1 lists the top 5 most frequently targeted states by 5 different HROs. It is clear that each HRO has a different list.

Table 1.1: Top 5 Targeted States by Five HROs (1995-2004)

Amnesty International	ICRC	Human Rights First	Protection International	International Comm of Jurists
U.S.	Serbia	U.S.	Guatemala	U.S.
Israel	Afghanistan	Egypt	Uganda	Israel
U.K.	Russia	Bahrain	Burundi	Nepal
China	DRC	U.K.	Colombia	Sri Lanka
Iraq	Israel	China	DRC	U.K.

In addition, it is not only the case that international shaming efforts often target states that have good practices and ignore states with bad practices. Even when states with poor records are targeted, they are not necessarily targeted in the same way. For example, the People’s Republic of China (PRC) has faced external criticism of their human rights records, especially its detention practices. Amnesty International (AI)’s criticism of China is direct and intense, as the following example suggests:

“Hundreds of thousands of people are believed to be held in ‘Re-education Through Labour’ facilities across China, many in harsh conditions, also referred to as ‘black jails.’ These include critics of the government or followers of ‘banned beliefs.’ Detainees in detention remained at high risk of torture and other ill-treatment... No one who is detained receives a fair trial in accordance with international human rights standards. Failing include: lack of prompt access to lawyers, lack of presumption of

innocence, political interference in the judiciary and failure to exclude evidence extracted under torture.... Innocent people are punished by the death penalty extensively, arbitrarily, and often as a result of political interference” (AI 2008, 108-109).

Yet AI’s approach is not the only one. Reporting on the conditions of the Chinese detention facilities in 2005, the UN Special Rapporteur (SP) paid more attention to legal and institutional reform. Instead of critically blaming the conditions of detention, SP complimented the Chinese government for having made a progress in order to improve the situation:

“In the opinion of the Special Rapporteur, methods used in the system of ‘Re-education Through Labour’ in China, and similar methods of re-education in prisons, pretrial detention centres, and other institutions often go beyond legitimate rehabilitation measures provided for in article 10 of the International Covenant on Civil and Political Rights (ICCPR)... The Special Rapporteur also observes positive developments at the legislative level, including the planned reform of several laws relevant to the criminal procedure, which he hopes will bring Chinese legislation into greater conformity with international norms, particularly the fair trial standards.... In general, the Special Rapporteur found that although the specific conditions of the facilities varied, in terms of basic conditions, they were generally satisfactory” (Nowak 2009, 22).

After visiting the Chinese detention facilities and meeting with government officials,

the International Committee of the Red Cross (ICRC) did not reveal any information about the condition of detention facilities.

“In China, discussions were held with the Ministries of Justice and Public Security regarding training in aspects of health in detention. It was agreed to organize two seminars and two study tours in 2010, thus ensuring proper follow-up of the work undertaken in 2008” (ICRC 2009, 237).

It is simply not the case that states are shamed in a coordinated fashion. Different HROs report different information about abuses. Some provide highly detailed information. Some are far more vague.

1.0.2 Argument Summary

In order to shame, HROs need information about states’ actual human rights practices. The field of international human rights seems to have largely assumed that this information is exogenous to human rights politics or at least freely available. The production of good information, however, is not without costs, both material and political. I argue that HROs essentially choose over two broad approaches to information gathering – an indirect approach in which they leverage the information provided by domestic communities and a direct approach in which they collect information themselves. I suggest that choices over these approaches require HROs to confront a critical dilemma, the resolution of which ensures that either aggregate information about human rights practices will be biased or the international human

rights network cannot be fully coordinated or both.

HROs have reason to choose an indirect approach because collecting information directly is not only costly but also involves becoming entangled with the governments being monitored. By instead relying on information provided by domestic actors, HROs can be kept informed on human rights conditions around the world without spending limited resources on investigation missions and without raising questions about their objectivity by remaining functionally independent of the states they watch. However, by choosing an indirect approach, HROs tend to get more information from states in which there is well-organized civil society. States with strong civil societies are less likely to seriously, systematically violate human rights, but in the event that they do, violations are likely to be exposed. The result is when HROs rely on an indirect approach, they will have far more information about abuses in places with strong civil societies and for this reason, they are likely to both over-report on states with generally good records and under-report on states with relatively bad records.

Given the possibility that the indirect approach risks poor information about human rights, especially in states with weak civil societies, HROs also have incentives to take a direct approach to information gathering. A direct approach presents HROs with an opportunity to investigate claims on the ground, to question officials and to make more personal links with activists. By communicating with a government, HROs may also have a chance to directly governments them to alter their practices. They may learn better about which practices are more or less difficult to

reform, in light of political realities. And that information may be critical to real policy change. On the other hand, investigation missions are costly and governments clearly have incentives to hide serious violations from view. HROs have to find ways to observe and understand the contexts they monitor properly. To do so, to gain access, they may have to “negotiate” the information that they will release publicly.

In short, I argue that the politics of information gathering creates a dilemma for HROs. HROs can rely on third parties for information, which allows for wider coverage at lower cost yet risks biased information depending on the nature of domestic civil society. Instead, HROs can seek information directly, but at great cost and typically with strings attached.

1.0.3 Outline of the Dissertation

This dissertation provides theoretically grounded empirical tests of why and how HROs name and shame states over human rights abuses.

Chapter 2 sets out the theoretical framework that I test throughout the rest of the dissertation. By examining the ways in which HROs name and shame, I argue that the process by which HROs collect information is critical to understanding these patterns. HROs that largely rely on third parties to generate information tend to focus on states with relatively good records; and, they shame these states severely. In contrast, HROs that directly engage states tend to focus on the worst human rights violators but they are less likely to harshly criticize these states. The

reason is that HROs that indirectly monitor via third parties receive a relatively large amount of information from states in which there is a well-organized civil society prepared and willing to expose violations, and these states generally have reasonable human rights records. HROs that directly engage states focus on serious abuses; however, in order to gain access, HROs are more likely to negotiate the information that they will release publicly.

Chapter 3 provides detailed discussion of measuring naming and shaming on the basis of new data on naming and shaming. There have been a number of shaming data, but they are incomplete in terms of their scope and depth. I introduce new dataset that compile a comprehensive data on the shaming behavior of over 90 HROs listed in *the Yearbook of the International Organizations* and the United Nations. Utilizing text analysis techniques based on Natural Language Processing (NLP), I process nearly 40,000 human rights documents. I characterize the actors and targets of HROs shaming with a Named Entity Recognize (NER). I also measure the aggressiveness of each HROs shaming activity by using text classification.

Chapter 4 and Chapter 5 seeks to test the theory described in Chapter 2 through a series of statistical analyses. Considering the nature of the event data, I use negative-binomial regression models to estimate the count model that covers which states are targeted by HROs. I also use tobit-regression models to estimate how severely each targeted states are shamed. I find that by choosing in indirect approach, HROs are more likely to both over-report on states with generally good records and under-report on states with relatively bad records. Also, these HROs

criticize states harshly and they do not withhold information. HROs that directly engage states, however, tend to focus on states with the worst human rights conditions, but they are less likely to criticize states harshly.

Chapter 6 discuss and test hypotheses described in Chapter 2, why HROs choose over two broad approaches to information gathering—an indirect approach in which they leverage the information provided by domestic communities and a direct approach in which they collect information themselves. I suggest there are two primary reasons—material costs and political costs. Despite its benefits, it is financially very costly for HROs to collect information directly. Most HROs cannot afford to sent staff members on investigative missions around the world. Direct approach also requires the assistance of the government subject to scrutiny. It gives governments “social camouflage” that can lower the expected costs associated with criticism.

Building on the empirical research in the previous chapters, Chapter 7 concludes. It summarizes the dissertation’s findings and implications for broad theories of human rights and human rights organizations.

Chapter 2

A Theory of Naming and Shaming

2.1 International Human Rights and Human Rights Organizations

While many governments have guaranteed rights to their citizens, in a world of sovereign states, human rights were traditionally seen as a matter of domestic jurisdiction. The prevailing doctrine was that foreign governments or other foreign organizations had no legitimate right to interfere with a state's treatment of its citizens. In particular, during the Cold War, security concerns overwhelmed concerns of human rights among states and within the United Nations, and when human rights concerns were cited, they often became a political tool in the ideological conflict between the superpowers (Donnelly 2003, Forsythe 2012*a*).

For the past 60 years, major changes occurred and human rights became an important issue in world politics. As Cmiel (2004, 117) notes, "few political agendas

have seen such a rapid and dramatic growth as that of ‘human rights.’” Human rights language is increasingly used in debates over military intervention, foreign aid, and globalization, and as the volume of human rights talk has grown (Cleary 1997, Forsythe 2012a, Keck and Sikkink 1998). In particular, major changes occurred in the 1970s making human rights an important issue in world politics, that is—the growth of human rights organizations dedicated to the issue of human rights.

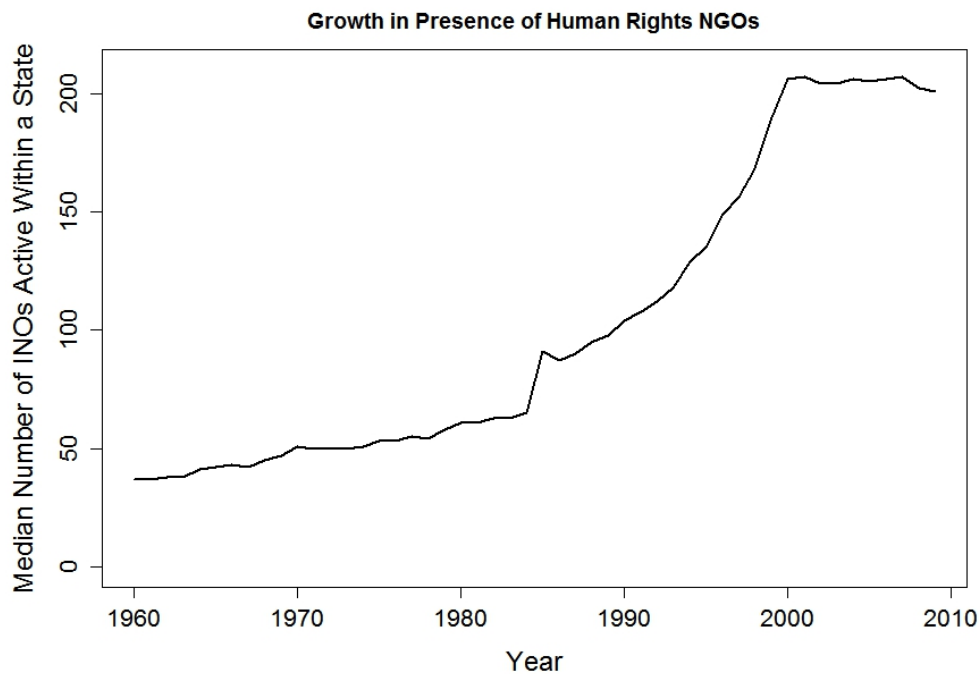


Figure 2.1: Growth in the Presence of Human Rights INGOs
Source: Murdie (2009)

2.1.1 Human Rights Organizations

Before moving to consider naming and shaming, some definitions are called for. There are a considerable number of Human Rights Organizations (HROs), working

at the international and national levels. They come in many shapes and forms. There are a number of ways to define them based on the degree to which their work addresses “human right” issues, how they self-define, and their modes of working or types of intervention. Since human rights are increasingly broadly defined to cover civil, political, social, economic and cultural, and now a range of third generation “solidarity rights” (Bell 2008, Cohen 1995, Donnelly 2003, Wiseberg 1992), large number of groups could be described in some sense as human rights organizations.

For example, in *the Yearbook of International Organizations*, there exist 1,094 organizations under the category of human rights in 2012 (UIA 2012). They cover a broad range of very different groups, from those dealing with campaigns against torture, land mines, or plastic bullets, to those pursuing the rights of smokers, the right to the right to euthanasia, the right to housing, etc. This project focuses on human rights organizations whose core mandate is the promotion and protection of internationally defined human rights. They explicitly pursue long-term social changes based on the internationally recognized human rights listed in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the International Covenant on Economic, Social, and Cultural Rights (Donnelly 2007, 7). This eliminates many organizations that deal with the right of smokers, the right to euthanasia, intellectual property rights, or animal rights as their aims because these rights are not the internationally recognized human rights listed in the International Bill of Human Rights.

HROs engage in various activities in order to achieve their goals of improving and

promoting the respect for human rights such as capacity building, public education (grassroots mobilization, fundraising/grant management, research, and compliance monitoring. HROs' major role can be understood with three broad categories.

First, HROs try to put human rights issues in international community's agenda. HROs were critical in first getting human rights on the international agenda (Korey 1998). Many of the crafters of the UDHR saw HROs as the critical link in making the original UN human rights documents a reality (Korey 1998, 2). This view of HROs as entrepreneurs in the spread of the nascent international human rights regime is much in line with the larger literature on norm diffusion (Finnemore and Sikkink 1998). HROs worked to make human rights, in general, a "standard of appropriate behavior for actors with a given identity" (Finnemore and Sikkink 1998, 891). They were central to the emergence of these standards, working to incorporate norms concerning human rights into a larger set of desirable behaviors by states and nonstate actors (Korey 1998). Very quickly in the human rights movement, HROs became key players in framing discourse around the legitimacy of basic or core human rights (Ropp and Sikkink 1999, 2). These rights, because of larger resonance with social discourse, have remained at the center of the HRO sector's work to improve the governmental practices (Carpenter 2007, Hopgood 2006, Price 2003, Risse 2002).

Second, HROs try to help domestic advocacy groups in their own struggle against repressive governments; HROs, often looking for domestic legitimacy, come to the aid of the domestic population (Bob 2005). HROs provide connections, funds, and

information to support a domestic group in its efforts to pressure the repressive state. Similarly, these organizations can aid in legal and medical aid for victims of repression (Hopgood 2006). Many HROs also focus on educating local populations about human rights in order to increase domestic mobilization for improved human rights practices (Welch 2001). In short, through their work within a state, HROs can augment and heighten pressure on a repressive regime “from below” (Brysk 1993, Simmons 2009).

Third, HROs encourage other states and IGOs to join in advocacy attempts (Keck and Sikkink 1998, Murdie and Davis 2011, Risse and Sikkink 1999, Welch 2001). HROs are critical in increasing world awareness of the human rights situation in the repressive state. As a non-state actor, they do not have sufficient material powers to physically or directly enforce sovereign states to improve the respect for human rights. Thus, HROs have primarily focused on naming and shaming. In other words, The power of HROs comes from their ability to disseminate information about human rights abuses that governments would prefer to remain hidden in the dark, yielding common knowledge about the abuses and undermining the illusion that the abusive governments act unobserved. They collect, assess, and publicize information about human rights conditions in states in order to bring pressures and persuade governments to comply with international standards. They try to “convert moral authority into ideational power” (Hawkins 2002, Keck and Sikkink 1998, Risse 2002). This is commonly called “naming and shaming.” HROs have played in important role in gathering information, framing it in a way that

can gain international attention, pressuring governments to take action, and lobbying for the creation of additional agencies and institutions to enforce human rights standards. For example, in the past twenty years, human rights organizations like Amnesty International and Human Rights Watch have been very active in publicly criticizing states for human rights abuses. The UN Human Rights Council (UN Commission on Human Rights) was given the authority to investigate human rights violations in particular countries, making the UN an important forum for international human rights concerns (Forsythe 2012*a*, Lebovic and Voeten 2006). Likewise, regional intergovernmental organizations, such as the Inter-American Commission on Human Rights, connected with the Organization of American States, became increasingly active on the issue of human rights (Donnelly 2003, Forsythe 2012*a*, Keck and Sikkink 1998).

In sum, HROs are central players for improved human rights. They act as normative entrepreneurs, framing human rights as a norm consistent with existing standards of behavior (Finnemore and Sikkink 1998, Korey 1998). Through their subsequent shaming efforts, HROs both pressure states to change their practices directly and facilitate international public pressure on the targeted state.

2.2 Existing Studies on HROs

2.2.1 First Generation of HRO Research: Normative Analysis

A group of scholars theorized how individual HROs is supposed to work in order to persuade states to adopt human rights norms through shaming activities. Given the fact that HROs do not have sufficient material powers of states, they cannot physically and directly enforce national governments to comply with international human rights laws and standards. HROs' shaming can work to draw states into patterns of discourse where human rights norms are first adopted for instrumental reasons and later become "standard operating procedures" (Risse and Sikkink 1999, 17). HROs use shaming to persuade states to change behavior to reflect their identity as part of the world community (Risse and Sikkink 1999). In this context, Hawkins (2002) defines human rights shaming as "nonviolent activities carried out by transnational networks and states with the primary purpose of improving individual rights by creating economic and political costs for a repressive government" (20). Many scholars argue for the importance of transnational pressure in reducing repression by a foreign government (Brysk 1993, Hawkins 2002, Risse 2002, Sikkink 2007). Probably, one of the most prominent theory is Keck and Sikkink (1998)'s a boomerang pattern of influence for transnational advocacy networks (TANs). In this model, domestic human rights NGOs, facing a recalcitrant government, appeal to international NGOs, which gather information and promote the human rights

concerns with other governments and with intergovernmental organizations, which then pressure the offending government. The boomerang model makes it clear that a variety of actors play a role in applying pressure against repressive governments, but the way the model is depicted (in Figure 2.2) suggests that HROs can bring the greatest degree of pressure to the abusive governments. HROs are critical in increas-

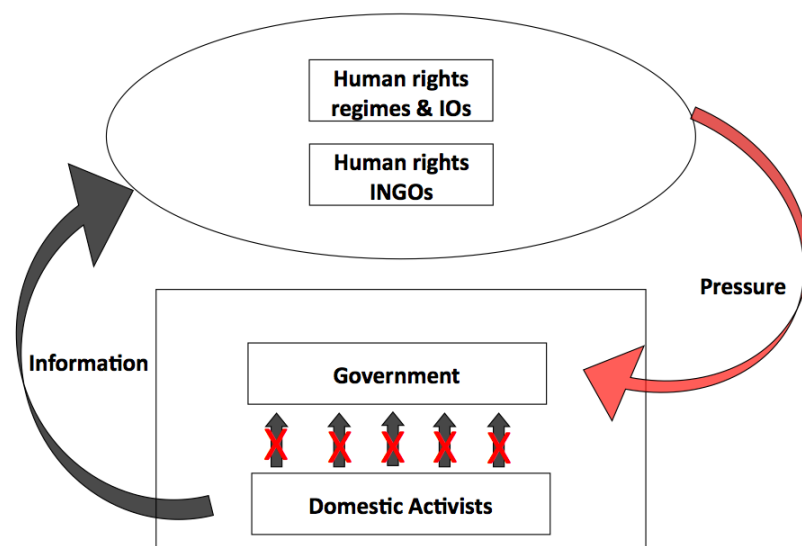


Figure 2.2: The Boomerang Model
Source: Keck and Sikkink (1998)

ing world awareness of the human rights situation in the repressive state. HROs can call on the international community, including third-party states, individuals, and organizations, to help pressure a target-state “from above” (Brysk 1993, Keck and Sikkink 1998). They also raise the “moral consciousness” of third parties (Risse and Sikkink 1999, 244). Thereby, HROs can expect that they could influence aid distribution and foreign direct investment from these third parties, which in turn can cause a targeted government to change its policy relating to a human rights issue

(Richards, Gelleny and Sacko 2001, Risse and Sikkink 1999). Regime leaders care about their identity on the international stage and may be persuaded to change their behavior to stop international attention and maintain their standing as a legitimate regime. Importantly, this model also puts emphasis on the importance of domestic mobilization. Thus, improvements in human rights “can be expected once pressure is exerted on norm-violating governments ‘from above’ and ‘from below’”(Risse and Sikkink 1999, 276).

However, despite its prominence, the boomerang model have not gone through systematic elaboration and testing to move them from the narrative or the metaphor (Bloodgood 2011). In other words, researchers based on this model have largely been descriptive and interpretative in nature.¹ Focusing primarily on how transnational advocacy networks develop (once established) and the impacts of these networks on states, they try to explain the way that the boomerang pattern is “supposed” to work. In particular, they do not properly explain how those multiple actors overcome collective action problem and sustain the network without any structure or organization rules (Prakash and Gugerty 2010).

¹In fact, Keck and Sikkink (1998) do not assert that they propose a theory or a model of NGO behaviors, although it integrates many of the ideas of transnational advocacy networks. They admit that their approach to testing their ideas about transnational advocacy is less rigorous than they might like (Bloodgood 2011, Keck and Sikkink 1998)

2.2.2 Second Generation of HRO Research: Impact Analysis

A growing number of works on HROs have paid much of the attention to evaluating HROs efficacy primarily. The effect of “naming and shaming,” however, remains a topic of intense debate. Whereas quantitative, cross-national studies have found a null, if not negative effect associated with NGO advocacy, qualitative studies continue to assert a positive influence. Human rights practitioners and scholars alike are very firm with the notion that spotlighting abuses has important consequences for reducing human rights abuses (Brysk 1993, Hawkins 2002, Hertel 2006, Keck and Sikkink 1998, Risse and Sikkink 1999, Roth 2004). Built upon the boomerang model, Risse and Sikkink (1999) examined how states responded to international pressure for normative change in Kenya and Uganda. Explicitly elaborating the boomerang effect in a case study of US foreign policy and Helsinki Final Act in the 1980s, Risse (2002) argues that the application of the boomerang effect is not limited to relatively weak states with poor domestic institutions. In the same context, Hawkins (2002) studied the violence occurred shortly after taking power took by military junta in 1973 and argued the human rights pressure helped create space for the domestic political opposition and encouraged the government to allow a relatively fair plebiscite in 1988, which cleared the way for democratic transition (Hawkins 2002, Pieper 1998). On the basis of a few case studies, Brysk (1993) expands the idea to understand how advocacy NGOs help to change human rights practices and policies by putting pressure on states from above and below at the same time.

A number of scholars, however, have pointed out the chasm between belief and empirical evidence. Especially, regardless of support from practitioners (Roth 2004), large-N examinations of naming and shaming often dismiss the notion of the efficacy of naming and shaming efforts across the board. In other words, studies that have evaluated the effects of naming and shaming from a statistical perspective find that on the whole, such efforts are not uniformly effective (Hafner-Burton 2008). Some scholars argue that the sources of the shaming matters (Franklin 2008) or the presence of domestic NGOs matters (Murdie and Davis 2011).

This debate over effectiveness important, but understanding the determinants of HROs' naming and shaming efforts is a methodological precondition for developing accurate assessments of the effectiveness of HROs' shaming activities. Analyses that fail to account for the process by which HROs select their targets are likely to produce biased estimates of the efficacy of the resulting naming and shaming efforts. In other words, if the severe human rights abuses and human suffering generated by a government has little bearing on whether and how HROs engage in naming and shaming strategy, or if some countries that are relatively doing well receive systematically higher levels of blaming than others, then governments would not necessarily feel pressured by international opinion.

2.2.3 Third Generation of HRO Research: Shaming Mechanism

There are some works paying attention to the process of HROs' naming and shaming activities (Bob 2005, Carpenter 2007, Lebovic and Voeten 2006, Meernik, Aloisi, Sowell and Nichols 2012, Risse and Sikkink 1999, Ron, Ramos and Rodgers 2005).

On the basis of "Spiral Model," Risse and Sikkink (1999) analyzed the effect of local protests and rebellions on attracting the scrutiny of international actors investigating human rights violations. A country's human rights issues are more likely to become a part of an HROs' agenda when protests and rebellions undermine a repressive regime whose leaders, in turn, begin responding to public demands for openness and human rights protection. In other words, according to their model, HROs pay attention to states where HROs' activities are more likely to be effective.

Carpenter (2007) uses a domestic-based theoretical framework to explain why certain issues enter the agenda of HROs' while other issues have not. Carpenter suggests a two-step process of issue definition and issue adoption that allows the problem to enter the NGO's agenda. Issue definition implies that a certain state of affairs "is considered neither neutral nor accidental" and "responsible party or parties are identified with credible solutions" (Carpenter 2007, 103). Issue adoption implies that the "the issue is then championed by one major player in the broader network" (Carpenter 2007, 103). Schmitz (2002) suggests that some countries may end up high on an NGO's agenda because the prevalent violations are of particular importance for human rights organizations.

Ron, Ramos and Rodgers (2005) point out that these factors are able to affect the human rights agenda of AI due to the organizational need for exposure. Human rights groups operate with very limited resources. The need for funding and widespread support leads them to tackle those situations that are “high-visibility,” which in turn will help them “gain momentum to protest violations elsewhere” (Ron, Ramos and Rodgers 2005). Sikkink (2007, 17) suggests that we should expect AI to be strategic in the choice of targets given a limited budget, which would direct its agenda “where its limited resources will make the largest difference.” This implies that AI may focus on those places where information is readily available or where a domestic supportive organization (AI’s offices or local activists) facilitates the gathering of information about human rights violations.

Bob (2005) suggests that the drive for media attention, as a factor shaping the agenda of NGOs, is more strictly correlated with the limited resources NGOs have at their disposal. Their ability to create resonance for an issue means more funding, greater impact and sustenance of the organization. On this point, scholars (Cooley and Ron 2002, Mihr and Schmitz 2007, 986) imply that recently the human rights agenda has been dictated by the necessity of organizations to maintain their “market share.” Human rights NGOs have relied on their ability to create publicity that would attract donors’ attention.

Meernik et al. (2012) also argue that while governments’ abusive human rights practices will propel these human rights groups to collect, disseminate, and publicize information pertaining to specific abuses, other permissive conditions must be

present for these groups to mobilize. More specifically, in accord with those who highlight the importance of what we term “openness” (Keck and Sikkink 1998), there should generally be sufficient political freedom, economic development, and exposure to the international community to enable the activation of human rights organizations operating locally.

Despite the importance of the subject, we know little about the determinants of HROs naming and shaming. In addition to the fact that recent work on the topic has largely focused on the efficacy of naming and shaming, much of the literature assumes that HROs are principled or altruistic, motivated solely to help people with their struggles against abusive regimes (Clark 2001, DeMars 2005, Keck and Sikkink 1998, Kelly 2005). Consequently, HROs are assumed to be motivated by “values rather than anything else” (Keck and Sikkink 1998, 2). They have simply assumed that HROs blame the worst human rights violators. Second, human rights scholars have assumed that HROs are very well coordinated so they could work as a single actor to bring coherent pressure on repressive states (Keck and Sikkink 1998, 9). They have assumed that all HROs are homogeneous with the same goal and tend to pay attention to a single group or domestic group. As a result, they focus on descriptive patterns rather than explain why that might be the case. In addition, all the studies are based on Amnesty International (AI) alone. Although AI is one of the leading HROs, it is not the only one and neglects variations within shaming that I have discussed in the previous section.

2.3 A Theory of Naming and Shaming

Contrary to the conventional wisdom that there should be a strong correlation between states' human rights records and the amount of criticism they receive, while many states with good human rights records are frequently shamed, many states with terrible records are rarely shamed. Moreover, even when abusive states are subject to a great deal of criticism, they are not necessarily shamed in the same way. Different HROs report different information about abuses in those states. Some HROs provide highly detailed information. Some are far more vague.

This section outlines the core theory of this dissertation. HROs have choices over their information gathering strategy. They can rely upon an indirect or a direct information gathering and there are real consequences of each approach. These consequences influence the choice that HROs make. First, this chapter specifies that in order to shame, HROs need information about states' actual human rights practices. The field of international human rights has largely assumed that the information is exogenous to human rights politics or at least freely available. The production of good information, however, is not without costs, both material and political. Second, I discuss why some HROs choose an indirect approach. By relying on information provided by domestic actors, HROs can be kept informed on human rights conditions around the world without spending limited resources on investigation missions and without raising questions about their objectivity, remaining functionally independent of the states they watch. However, by choosing an indirect approach, HROs tend to get more information from states in which there is well-

organized civil society. States with strong civil societies are less likely to seriously, systematically violate human rights, but in the event that they do, violations are likely to be exposed. The result is when HROs rely on an indirect approach, they will have far more information about abuses in places with strong civil societies and for this reason, they are likely to both over-report on states with generally good records and under-report on states with relatively bad records. Third, I discuss why some HROs choose to collect information directly. Given the possibility that the indirect approach risks poor information about human rights, especially in states with weak civil societies, HROs have incentives to take a direct approach to information gathering. A direct approach presents HROs with an opportunity to investigate claims on the ground, to question officials and to make more personal links with activists. By communicating with a government, HROs may also have a chance to directly persuade them to alter their practices. They may learn better about which practices are more or less difficult to reform, in light of political realities. And that information may be critical to real policy change. On the other hand, investigation missions are costly and governments clearly have incentives to hide serious violations from view. HROs have to find ways to observe and understand the contexts they monitor properly. To do so, to gain access, they may have to “negotiate” the information that they will release publicly. Fourth, I examine the choice of information gathering strategy that HROs adopt. Given the nature of human rights violations, it could be ideal for HROs to visit a state and observe directly to gather information. However, HROs are constrained by two important costs—material costs and political

costs. It is often very expensive for HROs collect information directly. Most HROs cannot afford to send staff members on investigative missions around the world. There are also political costs associated with information gathering. HROs have to negotiate with governments to make arrangement for visit and to gain access to sensitive facilities. Dealing directly with abusive governments involves some risk to the impartiality and independence of HROs. Not every HROs is willing to pay this price for information; for some HROs, this kind of information is too costly. Thus, interaction between two costs produces the information gathering choices for HROs.

2.3.1 Human Rights Violations and Essential Challenges of Information Gathering

In order to shame, HROs need information about states' actual human rights practices. In order to inspire corrective efforts by governments, HROs must demonstrate that their factual statements are true and thus constitute a reliable basis for remedial governmental policy. Simply "knowing" human rights violations are occurring is not enough; otherwise, HROs will lose adherents and influence. HROs make efforts to find information about human rights practices. Their credibility and legitimacy depend in large measure on the reliability of the facts obtained through such efforts. The field of international human rights seems to have neglected the fact that the production of good information is costly. There are significant challenges of information gathering.

2.3.1.1 The Nature of Human Rights Abuses

First, given the secretive nature of human rights abuses, it is often difficult to detect abuses (Davenport and Armstrong 2004, Dueck, Guzman and Verstapen 2001, Hathaway 2002, Simmons 2009). States' human rights practices take place domestically; thus, despite the widespread development of accountability mechanisms, many human rights violations truly are difficult to detect, observe and verify. As Lutz and Sikkink (2003) argue, "forced disappearances often are difficult to prove because the accuser must show that the victim was deprived of his or her freedom by government agents notwithstanding government claims to the contrary" (635). In particular, human rights violations often occur in situations in which governments destroy evidence, not only including killing the victims and hiding the bodies, as in the cases of disappearances, but also destroying other physical and documentary evidence such as the facilities in which the violations occurred. Very little documentary evidence has been produced, for example, regarding the atrocities committed by the Khmer Rouge in Cambodia, because the regime was able to destroy it (Lupu 2013, Ratner and Abrams 2001). Victims of such violations are often stripped of identification and identifying physical features, thus complicating the ability to use bodies as evidence of governments' abuses. This occurred in Guatemala during the 1970s-1980s, for instance, where only 127 of 3,171 unidentified bodies had been identified as of 2008 (Lupu 2013, Snow et al. 2008). Evidence of torture can also be hidden: "The goal of clean technique is plausible deniability by state executives. Clean techniques permit state agents to shift debate about their treatment of prisoners from blatant

lying to a ‘he said, she said’ context in which uncertainty exists” (Conrad and Moore 2010, 461). Even the most consolidated democracies commonly commit violations of human rights, especially when leaders perceive threat. In particular, as states democratize, they commonly transform their use of repression from overt to covert strategies (Conrad and DeMeritt 2011, Conrad and Moore 2010, Davenport, Moore and Armstrong 2007). Consequently, although interested publics may have a rough sense of a state’s human rights practices, the facts are far from complete. Given the secretive nature of human rights abuses, it is important for HROs to gather information that can be effectively put together to influence the implementation of human rights by governments. In order to inspire corrective efforts by governments, HROs must demonstrate that their factual statements are true and thus constitute a reliable basis for remedial governmental policy. Simply “knowing” human rights violations are occurring is not enough. HROs’ intentions might be laudable, but HROs must also reach concrete factual conclusions. As Roth (2004, 67) suggests, HROs must show that (1) a particular state of affairs amounts to a violation of international human rights standards and (2) a particular violator is principally or significantly responsible for the violation. Otherwise, HROs will lose adherents and influence. HROs make efforts to find information about human rights practices. Their credibility and legitimacy depend in large measure on the reliability of the facts obtained through such efforts. And the reliability of their information depends, in turn, on the methods used to obtain such information.

2.3.1.2 Competition and Survival

Second, all transnational actors are concerned about their organizational survival. They compete for resources and success in order to survive; HROs are not an exception. HROs are concerned with survival (remaining financially solvent), in order to change the behavior of violating governments. They have to make themselves attractive to potential members, donors, and supporters by making themselves look effective in changing or improving governments' human rights practices.

Although existing studies regard transnational actors as neutral and trustworthy (Clark 2001), many scholars have noted explicitly or implicitly that advocacy organizations resemble interest groups and they are not only normative but also strategic (Keck and Sikkink 1998, Princen and Finger 1994, Sweet 2000). HROs, just like interest groups, have to not only consider using their resources effectively and efficiently, but also care about their own organizational survival. In order to continue their campaign on human rights issues, HROs have to exist first and they must be concerned about organizational survival. HROs, as a watchdog, rely primarily on donations from private individuals because of a desire to remain autonomous from states. Consequently, they have to spend a lot of time in fundraising. Andreoni and Payne (2003) have argued that private donors have "latent demands to donate." Although they are generally sympathetic to human rights violations issues, they need to be "waked and asked" to donate. Donors in foreign countries may have some sense or rough awareness, but, generally, they have incomplete information about human rights conditions in repressive states. By providing these donors with

the information (or the opportunities to donate), HROs get monetary and material contribution for their organizational survival.

A number of scholarly works on NGOs behavior have indicated that competition among them is much more common than acknowledged by Keck and Sikkink (1998) and dynamics within the networks is much more complex than described in the Boomerang Model. In other words, the lack of cooperation and solidarity among network members, and the intense competition over resources, information, and success within the network is more common. Even when all the actors in the networks have strongly shared interests in cooperating, distributional concerns can outdo (Salamon, Sokolowski and List 2003, 4). While their motivations may be noble and moral, they behave in much the same way as other actors. All transnational actors are concerned about their organizational interest. They compete for resources and success in order to survive. HROs are not the exception. HROs are concerned with remaining financially solvent, trying to change the behavior of violator governments. They have to make themselves attractive to potential members, donors, and supporters. They have to make themselves look effective in changing or improving governments' human rights practices. When there are multiple HROs they have to compete among one another for donations. For example, analyzing the operational structures of Amnesty International and Human Rights Watch, Rodgers (2006) has explained that the broader HRO community could be a powerful network of activists, but it creates a marketplace of fierce competition, and place increasing pressure on HROs to increase performance to survive. Looking effective is important

for HROs because it is directly related to donations and their organizational survival. And, in principle, HROs are accountable for how they spend the money they receive from donors. It means that, as a non-profit organization, they rely on individual donations. They are not completely independent of what their supporters want. As a result, given the available information, HROs have an incentive to disseminate “high quality” information about abusers. Disseminating inaccurate information can undermine the reputation and legitimacy of the HRO, resulting in less support by donors. At the same time, HROs do not want to hold onto information that they have, lest they be scooped by another HRO or be accused of concealing abuses or not effective organization. Consequently, HROs with an indirect approach are more likely to recognize and report on violations in the states with more information. In other words, when there are a number of other HROs working on the same issue, any HROs are expected to talk about the same issue. They cannot afford to hold onto any information that they have.

2.3.1.3 State Sovereignty

Third, while many governments have guaranteed human rights to their citizens, in a world of sovereign states, human rights is still seen as a matter of domestic jurisdiction (Donnelly 2003, Forsythe 2012*b*). Any outside actors such as HROs or other foreign organizations have no legitimate right to interfere with a states treatment of its own people. Human rights norms have to “fit” with the interests of states in maintaining sovereignty (Florini 1996). Since HROs try to reveal abuses

that governments do not wish to become public knowledge, it is needless to say that they would not welcome critical perspectives from HROs and therefore choose not to cooperate with such organizations.

These challenges imply that HROs essentially choose over two broad approaches to information gathering—an indirect approach in which they leverage the information provided by domestic communities or a direct approach in which they collect information themselves.

2.3.2 Indirect Information Gathering

2.3.2.1 Indirect Approach

One of the central approaches to information gathering is what I will refer to as “indirect,” through which HROs rely on third parties—such as domestic actors—as their source of information. Although sources vary from country to country, HROs normally collect their information by corresponding with domestic informants such as human rights activists, the families and friends of victims, former prisoners, lawyers, journalists, academics, political parties opposing an oppressive regime, other repressed groups, etc. They also review relevant domestic laws and periodicals and interview occasional visitors to know and evaluate human rights conditions in countries. The role of domestic groups and activists is crucial in an indirect approach to information gathering.

As described in the model of “Transnational Advocacy Networks (the Boomerang Model),” domestic human rights activists or groups that are blocked by their repres-

sive governments reach out to international activists (i.e. HROs) to garner global attention (Keck and Sikkink 1998, Ron, Ramos and Rodgers 2005).² Since HROs cannot know about all the human rights issues around the world, the very first step occurs with domestic activists or groups providing HROs with information on human rights conditions, which in turn motivates these organizations to further explore the country's violations (Bob 2005). On the basis of the information provided by domestic activists or groups, HROs assess and publicize information about the human rights conditions in states in order to pressure and persuade governments to comply with international human rights standards. When effective, naming and shaming changes state behavior, reducing or eliminating human rights abuses (Brysk 1993, Franklin 2008, Hawkins 2002, Keck and Sikkink 1998, Risse 2002, Risse and Sikkink 1999, Schneider 2000).

An active domestic civil society, in particular, can contribute to this process as an "enforcement mechanism" by allying with the transnational networks to exert pressure on the government (Clark 2001, Hafner-Burton and Tsutsui 2005, Landman 2005, Murdie 2009, Murdie and Davis 2011, Murdie and Bhasin 2011, Neumayer 2005, Wong 2008). For example, Murdie and Davis (2011) argue that a country's linkage to the international community and HROs can and does influence governments to change their human rights practices for the better. Domestic actors tightly

²For example, Chilean civil groups reached out to Amnesty International and provided information about forced disappearances committed by the Chilean government (Clark 2001, 73). Similarly, domestic human rights activists in Indonesia sought to publicize the issue of political prisoners through the International Commission of Jurists (Jetschke 1999, 140). These domestic groups or activists serve as resources to pass on the latest information about government atrocities to HROs (Haines 2006, Murdie and Bhasin 2011).

linked to the international community tend to have greater awareness of the rights they are entitled to; therefore, they can mobilize political pressures for improvements in human rights. A strong and active domestic civil society also plays a particularly important role as a key information provider. A strong and active civil society means more freedom of expression and political participation; thus individuals can freely provide information to HROs (Figure 2.3).

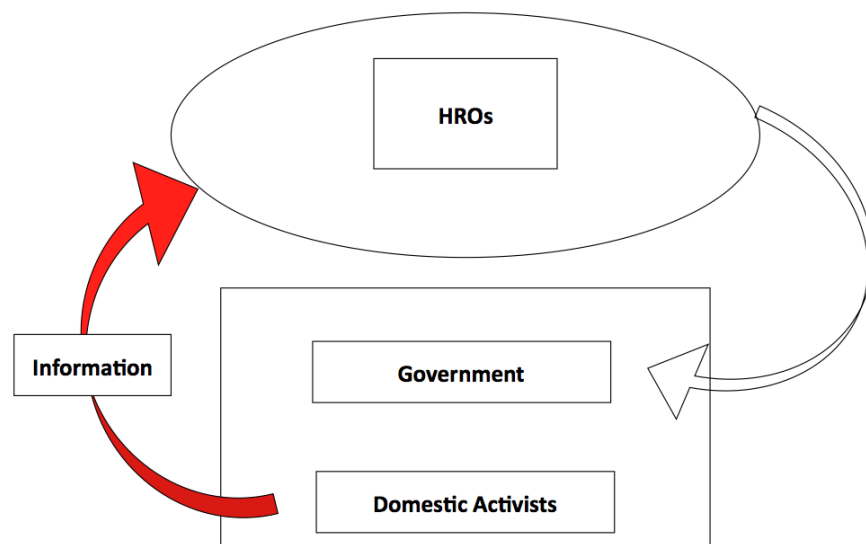


Figure 2.3: Indirect Information Gathering

2.3.2.2 Consequences of Indirect Approach

Depending solely on an indirect approach for information, however, produces problems for HROs. First, HROs are more likely to recognize and report on violations in states that provide more information. It seems to be especially pronounced for states with a good civil society. Although states with an active civil society gen-

erally display a greater respect for fundamental human rights than non-democratic countries with weaker, less active civil societies, HROs report of abuse may be higher for democratic states with strong civil societies and better records, because they encourage the freedom of speech, press, and open debate. In other words, HROs have access to more human rights violations for these countries, though the violations themselves are often less egregious than in more autocratic countries. This skews reporting of human rights violations. In addition to autocratic regimes being more likely than democratic governments to violate basic human rights, they also suppress press freedoms and political opposition that publicize those violations. Therefore, violations perpetrated by democratic states with strong civil societies, while much less frequent, are nevertheless more likely to be spotlighted because of the fewer restrictions placed on the domestic human rights activists, groups and their political activities.

Second, HROs do not want to hold onto the information they have. Although existing studies regard transnational actors as neutral and trustworthy (Clark 2001, Keck and Sikkink 1998, 18), a few scholars have noted explicitly or implicitly that advocacy organizations resemble interest groups that are normative and strategic (Murdie 2009, Princen and Finger 1994, Sweet 2000). In other words, HROs, just like interest groups, have to not only consider using their resources effectively and efficiently, but they also have to care about their own organizational survival. In order to continue their campaign on human rights issues, HROs have to first exist and be concerned about organizational survival. HROs, as watchdogs, rely primarily

on donations from private individuals in order to remain autonomous from states. As a result, they have to spend a lot of time in fundraising. Andreoni and Payne (2003) have argued that private donors have “latent demands to donate.” Although they are generally sympathetic to human rights violations, they need to be “waked and asked” to donate. Donors in foreign countries may have a rough awareness, but, generally, they have incomplete information about human rights conditions in repressive states. By providing these donors with the information (or the opportunities to donate), HROs get monetary and material contribution for their organizational survival.

A number of scholarly works have indicated that competition among NGOs is much more common than acknowledged by Keck and Sikkink (1998), and the dynamics within NGO networks is much more complex than recognized in the simple boomerang model. In other words, the lack of coordination and solidarity among network members, and the intense competition over resources, information, and success within the network is more common. Even when all the actors in the networks have strongly shared interests in cooperating, distribution concerns undermine these incentives (Salamon, Sokolowski and List 2003, 4). While NGOs’ motivations may be noble and moral, they behave in much the same way as other actors (Murdie 2009). They have to make themselves attractive to potential members, donors, and supporters. Furthermore, they have to make themselves look effective in changing or improving governments’ human rights practices. When there are multiple HROs, competition for donations is not an uncommon practice. For example, by analyzing

Amnesty International and Human Rights Watch, Rodgers (2006) explains how the broader HRO community can be a powerful network of activists; however, this creates a marketplace of fierce competition, and places increasing pressure on HROs to increase performance to survive. Looking effective is important for HROs because it is directly related to receiving donations and their organizational survival. In principle, HROs are accountable for how they spend the money they receive from donors. They are not completely independent of what their supporters want. Consequently, given the available information, HROs have incentives to disseminate “high quality” information about abusers. HROs do not want to hold onto the information they have, lest they be scooped by another HRO or be accused of concealing abuses or not having effective organization. Thus, HROs with an indirect approach are more likely to recognize and report on violations in states where they have more information. When multiple HROs are working on the same issue, it is expected that each of them discuss the similar contents of their findings and information, because they cannot hold onto any information that they have.

Third, the skewed perceptions of human rights abuses can be even more exacerbated or exaggerated. Groups and individuals concerned with human rights violations are most often victims or members of political parties and other organizations opposed to certain governments. Thus, getting a complete picture of human rights conditions can be even more difficult because it depends largely on the initiative of interested persons to come forward. While several scholars have called attention to HROs’ reliance on domestic opposition NGOs for information,

less explored are the actual processes of this reliance for the quality of information provided by domestic NGOs regarding government human rights practices. In other words, it would be a mistake to assume that the information from domestic NGOs is always “accurate” and unbiased. Information from domestic opposition NGOs is not necessarily representing the actual situation. In this context, Bob (2005) has argued that, since there is unequal power relations and competition among domestic oppositions, the powerful domestic NGO’s interests are more likely to be reflected in international HROs’ agenda and activities. Although HROs like AI or Human Rights Watch (HRW) try to ensure the integrity, accuracy, and impartiality of their research (Clark 2001, Welch 2001), they cannot be free from this problem as long as they depend on domestic opposition NGOs for information regarding human rights conditions. The following example is illustrative of these issues.

South Korea During the 1980s and 1990s, South Korea was frequently criticized by Amnesty International (AI 2000). Amnesty clearly indicated in their reports that their primary sources of information were domestic civil groups in South Korea such as Min-Ka-Hyup (Family for Prisoners of Conscience), Cham-Yeo-Yeon-Dae (People’s Solidarity for Participatory Democracy), Kyung-Sil-Lyun (The Citizens’ Coalition for Economic Justice), and Min-Ju-No-Chong (Korean Confederation Labor Union). On the basis of the information provided by these domestic NGOs, Amnesty reported the South Korean government’s repressive human rights practices. However, Amnesty’s description of the conditions was not as accurate as it claimed. Consider Table 2.1. Although I do not argue that those domestic civil

Table 2.1: Amnesty Estimates and NHRCK Estimates

Year	Amnesty International Reports	NHRCK Records
1987	“Over 3,400 people were charged with political offence under the National Security Law” (AI 1987:245)	104 (NHRCK 2004:11)
1991	“Over 1,500 people were imprisoned for politically motivated activities during 1990 under the National Security Law” (AI 1991:139)	Convicted and imprisoned: 269 (NHRCK 2004:13)
1993	“Over 600 political prisoners were imprisoned” (AI 1993:181)	The number of people who were serving their sentence: 174 (NHRCK 2004:13)
1995	“Hundreds of opponents were detained and imprisoned” (AI 1995:185)	99 (NHRCK 2004:13)
1996	“Several hundreds political activists were imprisoned under the National Security Law during the year” (AI 1996:200-201)	Total 76 were convicted and imprisoned (NHRCK 2004:13)

Source: National Human Rights Commission of Korea (NHRCK 2004)

groups in South Korea were completely corrupted it is reasonable to think that these domestic actors had incentives to delegitimize the South Korean government and their human rights abuses. Even more ironically, during the same period of time (1980-2000), South Korea was criticized 352 times by Amnesty International, while North Korea was mentioned less than 10 times (AI 2000).

It appears that human rights conditions are better in North Korea than in South Korea, simply because South Korea keeps relatively better and more open records that enable groups like AI to expose, more regularly, the human rights violations that occur there. Clearly, an information infrastructure (i.e. an active, strong civil

society) can produce artificially inflated levels of negative reporting.

Uzbekistan Blaming the violence (or the massacre) at Andijan, Uzbekistan in May 12-13, 2005, HRW (2006) reports that the government Security Forces indiscriminately shot from armored personnel carriers (APC) and helicopters into the crowd ranging 10,000 to 30,000 people. HRW reports that around 300 individuals were killed. Reporting the same event, AI (2006: 226) describes “the security forces killed hundreds of unarmed men, women and children when they fired indiscriminately in the eastern city of Andijan in May.” Relying on these sources, the New York Times also reported that 325 people were killed among the assembled crowd. Some domestic opposition NGOs in Uzbekistan said that it was 800 and other domestic NGOs said that it was 1,500 or more.

Since the accusation of a random massacre by Uzbek security forces, killing hundreds of unarmed civilians, would certainly constitute a grave violation of human rights law, the international community was very interested in establishing the facts of what actually happened in Andijan.

After its own investigation of the event, the Office for Democratic Institutions and Human Rights of the OSCE has estimated between 300 and 500 people lost their lives as a result of the events in Andijan (OHIHR 2005:22). The UN Office of the High Commissioner for Human Rights (OHCHR) has also reported that up to several hundred people may have been killed. The inconsistency in the number of people killed in the event came from the fact that all these HROs relied upon

domestic opposition NGOs for information.³

A member of NATO Research Group was Andijan and witnessed the site a day after the event and provided her direct observation.⁴ She reported that the death toll was closer to the government estimate (i.e. under 200 deaths) than to high estimates of HROs and news media reports. Also, Akiner (2005:36) indicates that domestic NGOs frequently accused each other of not being “genuine,” of making a business out of the information they provide to foreigners. In addition, looking at the square where, according to HRW, 10,000 to 30,000 people gathered, she reports:

“the Square is measured about 30 meters by 50 meters. At a tight squeeze, perhaps 3,000 people might be accommodated in this place. Another way of trying to gauge of the scale of the operation is to seek out the impact on the environment—for example, the numbers and location of bullet marks on surrounding buildings and on the monument in the center of the square. I did not see any such marks in the area. Moreover, a crowd of 10,000 to 30,000 people would not have just stayed still if shot at directly with machine gun. The panic alone would have killed more people.” (Akiner 2005:23)

She also indicates,

³The Uzbek government completely shut down the place and did not allow any foreign group to visit the place without permission from the government. Even the ODIHR from OSCE and the UN OHCHR could not visit the site and had to collect information on the basis of interviews with the refugees who fled to Kyrgyzstan.

⁴According to Akiner, the visit was her spontaneous initiative, while being in Tashkent for an NATO Research Workshop. While this workshop was cancelled on May 18 because of the events in Andijan, Akiner nevertheless travelled to Uzbekistan to finalize “organizational issues” In the course of this, she decided to travel to Andijan “in the spur of the moment” and sought official permission from the Uzbek authorities, which was issued “within the hour” (Akiner 2005: 42-43)

“The daytime temperatures were by this time close 90F. Thus, if there had been a large-scale massacre, by the following day the smell from the decomposing bodies would have been overpowering. Some media reports suggest that the dead (about 1,500) were taken to a school in Andijan. The school year was still in progress, so either alternative arrangements would have had to have been made for quite a large number of children to go to school elsewhere, or a massive decontamination job would have had to be conducted over the weekend. During my visit, there was no mention of children being sent to another school neither did anyone allude to a major decontamination exercise” (Akiner 2005: 21-22)

When HROs had no way to verify the quality of information provided by domestic sources, it often leads to inaccurate description and evaluation of states. Moreover, domestic opposition NGOs have incentives to delegitimize their incumbent government, they are more likely to provide inaccurate (or exaggerated) information regarding their human rights conditions.

While several scholars have called attention to HROs’ reliance on domestic opposition NGOs for information, less explored are the actual process of this reliance for the quality of information provided by domestic NGOs regarding government human rights practices. In other words, it would be a mistake to assume that the information from domestic NGOs is always “accurate” and unbiased. Information from domestic opposition NGOs is not necessarily representing the actual situation. In this context, (Bob 2005) has argued that, since there is unequal power relation and

competition among domestic oppositions, the powerful domestic NGO's interests are more likely to be reflected in international HROs' agenda and activities. Although HROs like AI or HRW try to ensure the integrity, accuracy, and impartiality of their research (Clark 2001, Welch 2001), they cannot be free from this problem as long as they depend on domestic opposition NGOs for information regarding human rights conditions.

This argument suggests the following implications.

H1A: HROs with indirect information gathering are more likely to shame states with a strong civil society.

H1B: HROs with indirect information gathering are more likely to be harsh on states.

2.3.3 Direct Information Gathering

2.3.3.1 Direct Approach

In a direct approach, HROs gather information through on-site investigation and direct communication with governments. Recognizing the problems of indirect information gathering, some HROs believe that it is absolutely essential that they conduct visits and assess the situation of human rights in specific countries because the effectiveness of naming and shaming depends on the reliability of the sources of allegations. Furthermore, in states where domestic activists or groups are few or non-existent, cooperation with governments might be an HROs' only option (Bell and Carens 2004, 322).

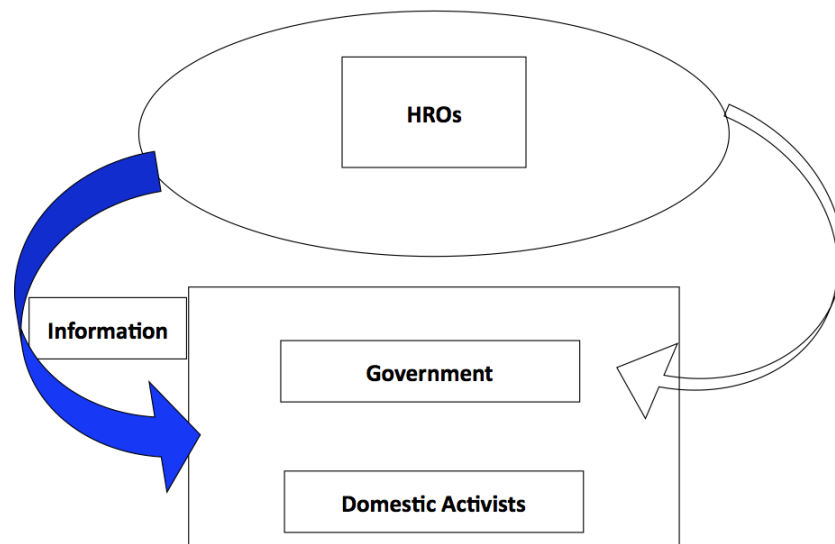


Figure 2.4: Direct Information Gathering

HROs that engage in direct information gathering believe that “producing a clear picture” is the first and most important step for raising the collective self-awareness of the abusive governments. In order to improve existing situations, HROs must first gain this “clear picture,” then use it to persuade government officials to act upon their recommendations. Sometimes governments are sincerely committed to improving the human rights situations in their countries but simply lack the capability to do so. Sometimes human rights violations can occur as a result of institutional inertia rather than active, state-willed perpetration of violations; while at other times, perpetration of these violations might result from a lack of technical skills and know-how. It is crucial in such cases for HROs to formulate clear, precise, and helpful instructions based on direct observation. Techniques for gathering information directly, including on-site investigation and the various fact-finding methods

such as physical visits to places of detention, interviews with detainees, and medical examinations of alleged victims of torture by forensic experts, enable HROs to objectively assess the general conditions and the extent to which human rights are violated in states (Franck and Fairley 1980, Naples-Mitchell 2011, Nowak 2009, Orentlicher 1990, Weissbrodt and McCarthy 1982, Winston 2001).

In general, direct information gathering is conducted by a team of professionals who have trained extensively in documenting evidence of any violations of human rights. For example, a former UN Special Rapporteur on Torture (SRT), Manfred Nowak (2009, 106-107) has explained that visiting detention facilities requires a group of experts such as a lawyer, a psychologist, a medical forensic expert, trained interviewers, photographers, reliable translators with the knowledge of local language, etc. Thus, HROs can get much more specific, vivid, and accurate information about human rights conditions through a direct approach than when relying on word of mouth from third party sources.

2.3.3.2 Consequences of Direct Approach

First, HROs that rely on direct information gathering tend to focus on the worst human rights violators. In order to improve existing conditions, HROs must first gain this “clear picture,” then use it to persuade government officials to act on their recommendations. HROs can receive much more specific, vivid, and accurate information about human rights conditions through a direct approach than when relying on word of mouth from third party sources. The former UN Special Rapporteur on

Torture, Nowak encapsulates the advantage of direct information gathering:

“I had to wait for about 10 minutes outside despite the fact that I could clearly hear that several activities were ongoing inside to ‘clean up’ the place... I entered after long discussions with the police authorities a ‘torture room,’ where more than a hundred detainees, including women and children, were packed together like animals under terrible conditions and subjected in the presence of others to severe physical torture. I requested to speak in private to the detainees to understand exactly what had happened, and requested the forensic expert to carry out a thorough investigation... Police officers said that all allegations of torture by detainees were fabricated as a strategy to avoid conviction and long prison sentences... Often, despite the overwhelming evidence of being tortured, the victims are extremely afraid of reprisals and they invent stories denying that they were subjected to torture. But with certain interview skills and taking into account additional evidence, including forensic expertise, it is possible to find out whether a detainee invents a story or tells the truth” (Nowak 2009, 116).

Thus, direct information gathering is one of the most important tools for an effective investigation of human rights violations such as torture, ill-treatment, and conditions of detention. Through direct observation, HROs can develop a fairly good understanding of the specific root causes of human rights abuses in the respective societies as well as of the measures needed to combat those violations and to improve

the conditions. On the basis of this assessment, HROs publicize their findings and specific recommendations for further actions.

Second, HROs with direct approach tend to cover fewer states than those with indirect approach because it is very costly for HROs to collect information directly. The resources expended to gather information vary, and HROs differ significantly in membership, structure, size of staffs, and financial capability. Most HROs, particularly non-governmental organizations, cannot afford to send staff members on investigative missions around the world (Keck and Sikkink 1998, Murdie and Bhasin 2011, Simmons 2009, Weissbrodt and McCarthy 1982). In general, an on-site investigative mission requires sending in a group of experts who are highly trained in documenting evidence of human rights violations on the basis of agreed terms with the visited governments (including domestic legislation, internal prison rules, security concerns etc.). Thus, employing expert staffs and sending them on these missions is costly; and this costliness often constraints actions that HROs can undertake (Nowak 2009, 107). Missions are expensive, and the availability of funds will affect the quality of a mission's findings, determining how long a mission will last and how many members the mission can include. Former members of the European Commission on Human Rights who conducted on-site investigations during 1990s have pointed out the material costs:

“We worried so much about the costs and they certainly played a role...

It used to cost about 60,000 pound if you carried out a week-long mission.

Air fares alone often ran into thousands. Members of the missions were

rarely paid, but the sponsoring organization paid the cost of the mission's living expenses. They would no doubt be much more expensive now. They are a drain on resources unquestionably" (Leach, Paraskeva and Uzelac 2009, 42)

Even the UN Special Rapporteur on Torture, whose main job is carrying out country visits, agrees with the importance of resource constraints:

"Even if every state consented to a visit and no political obstacles to compliance existed, we could not possibly assess the situation of torture in 192 countries in the world. Although we maintain a number of highly trained experts and spend over 50 percent of our time on the mandate, we could only carry out approximately four country missions a year. Doing a country-by-country assessment of all the countries would require a much bigger office, many more resources" (Naples-Mitchell 2011, 242).

Therefore, a single HRO cannot realistically take on all of the world's human rights violations equally. They have to make strategic and tactical decisions about what issues to prioritize and which countries "deserve" their attention. HROs face a constant tradeoff in their work: they can try to cover all or nearly all instances of human rights abuses in the world, thus increasing the scope of their activities or they can cover fewer cases more deeply.

Third, while the process of direct information gathering might seem to be neutral and technical, it is neither. Instead, it is an inherently political endeavor. Even

with the most exclusive inside information from visited states, HROs cannot always publicly criticize the oppressive states that they observe. Sometimes, HROs must obscure as much as they reveal. The process of direct information gathering may be ideal for HROs, but it may not be ideal for states. states do not have to cooperate with HROs that attempt to gather information. While many governments have guaranteed human rights to their citizens, in a world of sovereign states, human rights is still seen as a matter of domestic jurisdiction (Donnelly 2003, Forsythe 2012*b*). Any outside actors such as HROs or other foreign organizations have no legitimate right to interfere with a states treatment of its own people. Human rights norms have to “fit” with the interests of states in maintaining sovereignty (Florini 1996). Since HROs try to reveal atrocities that governments do not wish to become public knowledge, it should come as no surprise that such governments do not welcome critical perspectives from HROs and therefore can choose not to cooperate with such organizations. States only allow HROs to gather information when it is generally useful to a state’s interests. As long as HROs do not challenge the sovereign state system, they can work in a space where they are “allowed” to operate. In other words, even though states have an incentive to limit HROs’ access to their territory and detention facilities, in general, they do not block them completely. Just like repressive states join the international human rights regimes for the “expressive” benefits in order to signal their commitment to democratic values and reforms (Hafner-Burton and Tsutsui 2005, Hathaway 2002, 2004, Pevehouse 2005, Simmons 2009), states that do allow and cooperate with HROs’ investigative

visits can signal their “intention” or “commitment” to carry through with democratic reform. A few scholars also provide us with insights from a domestic political perspective. Vreeland (2008) explores the domestic institutional dynamics of autocracies and argues that the presence of opposition parties both cause autocrats to torture more heavily and forces these governments to sign human rights treaties. When opposition parties exist, there must be some freedom to engage in speech and activities that contradict the will of the incumbent government. In a similar vein, Hollyer and Rosendorff (2011) illustrate that authoritarian government use the accepting the jurisdiction of the UN Convention Against Torture (CAT)—followed by the willful violation of its provisions—as a costly signal to domestic opposition groups of their willingness to employ repressive tactics to remain in power. Thus, even those authoritarian governments have incentives to allow and cooperate with HROs’ visit. In a similar context, Hyde (2006) argues that the reason why states invite international election monitoring organizations to observe their elections at the “risk” of being caught cheating, is to signal to international audiences that they are trying to be responsible players in the international community. Those states can get international benefits such as foreign aid by “imitating” a committed democracy without actually becoming one. In addition, through actions such as these the governments signal to the power-sharing groups or domestic opposition groups that try to delegitimize the incumbent government. Therefore, while it may seem that cooperating with HROs can only hurt autocratic governments, doing so can in fact provide some benefits to oppressive regimes.

A number of HROs that utilize a direct approach emphasize how harsh public criticism after a country visit may impede further access; as a result, they have to weigh the merits and pitfalls of assigning government responsibility for abuses before moving toward public condemnation. The chair of the UN Committee Against Torture, Claudio Grossman explains this dilemma:

“We need to go softly. We cannot throw everything at them. We cannot demand too much as otherwise they will close the door on us. We must recognize that there are political realities on the ground” (Kelly 2009, 793).

In a similar context, a former UN Special Rapporteur for Myanmar, Sergio Pinherio states:

“On one side the UN Special Rapporteur is supposed to reform, to report about the human rights violations. But on the other hand, always, you have to engage with the government because if you are not able to talk to the government, there is no meaning” (Naples-Mitchell 2011, 239)

Disagreeing with Amnesty International’s critical public approach, the president of the International Committee of Red Cross (ICRC), Jakob Kellenberger emphasizes:

“The reluctance of the ICRC to make public pronouncements is undeniable. The reason for this reserve is: it does not want to risk losing its access to the victims by doing so We are not Amnesty International

After all, if public statement issued in protest and visits are suspended, one has used the last arrow in the quiver; one is out of the game and unable to influence the situation—at least for the time being” (Forsythe 2012*b*, 470).

Thus, I argue that for HROs that use direct information gathering, cooperation or collaboration with governments is necessary given the political context. HROs, then, are “flexible,” and willing to negotiate with governments. I predict that HROs that rely on direct information gathering have tendencies to focus on the worst human rights violators. This is largely due to HROs not needing to rely on domestic opposition groups for information and because they are willing to directly communicate with the governments. However, in order to secure their access to these governments, HROs are limited in the extent to which they can challenge the sovereign states violations. They are willing to work within the space they are given to operate.

This argument suggests the following implications.

H2A: HROs with direct information gathering are more likely to shame states with the worst human rights records.

H2B: HROs with direct information gathering are less likely to be harsh on states.

2.4 The Choice of Information Gathering Approach

Given that there are broadly two approaches to information gathering—an indirect way and a direct way, why do HROs choose one approach over the other? I suggest that there are two primary reasons—material costs and political costs.

2.4.1 Material Costs

As briefly discussed in the previous section, it is very costly for HROs to collect information directly. The resources expended to gather information vary, and HROs differ significantly in membership, structure, size of staffs, and financial capability. Most HROs, particularly nongovernmental organizations, cannot afford to send staff members on investigative missions around the world (Keck and Sikkink 1998, Murdie and Bhasin 2011, Simmons 2009, Weissbrodt and McCarthy 1982). Scholars generally assume that HROs are motivated by altruistic values and the moral impulse to help people with their struggles against abusive regimes (Clark 2001, DeMars 2005, Keck and Sikkink 1998, Kelly 2005). However, although HROs work hard to deal with human rights violations, they are not much different from other organizations in that they have limited time and budgets.

For example, in general, an on-site investigative mission requires sending in a group of experts who are highly trained in documenting evidence of human rights violations on the basis of agreed terms with the visited government (including domestic legislation, internal prison rules, security concerns etc.). Thus, employing expert staffs and sending them on investigative missions is costly; this costliness

often constrains actions that HROs can undertake (Nowak 2009, 107). In other words, investigative missions are expensive, and the availability of funds will affect the quality of a mission's findings, determining how long a mission will last and how many members the mission can include.

Therefore, despite its crucial role, we must acknowledge that direct information gathering is undoubtedly— and sometimes prohibitively—expensive for HROs. Considering their limited resources, HROs cannot realistically take on all of the world's human rights violations equally. They have to make strategic and tactical decisions about what issues to prioritize and which countries “deserve” their attention. In other words, with limited resources, HROs face a constant tradeoff in their work: they can try to cover all or nearly all instances of human rights abuse in the world, thus increasing the scope of their activities, or they can cover fewer cases more deeply.

Moreover, HROs have an incentive to disseminate high quality (accurate and reliable) information about abuses. Disseminating inaccurate information can undermine the reputation and legitimacy of HROs, resulting in less support by donors. As a non-profit organization, they rely on individual donations. They are not completely independent of what their supporters want. At the same time, HROs do not want to hold on to information that they have, lest they be scooped by another HROs or be accused of concealing abuses.

2.4.2 Political Costs

Direct information gathering requires at least the acquiescence, if not the assistance, of the government subject to scrutiny, and such cooperation is not always forthcoming or comes with the costs. As I have discussed in the previous section, HROs have to balance or compromise their strict human rights advocacy principles for information gathering. By communicating with governments, HROs have an opportunity to investigate human rights conditions and may have a chance to directly persuade them directly to improve their conditions. However, direct information gathering is costly because governments have incentives to hide their serious violations. To gain access to the information, HROs may have to negotiate or compromise the information that they will release publicly. This, in turn, gives governments an opportunity to deflect criticisms. In other words, cooperating with HROs may give governments “social camouflage” that can lower the expected costs associated with criticism. Governments use their cooperation with HROs as evidence that their policies on human rights are not so bad and are getting better (Bell and Carens 2004, 323). For example, under the International Committee of Red Cross (ICRC) Doctrine 58, the ICRC reserves the right to make public criticism. When a state has failed to produce ameliorative changes over time under internationally recognized standards, the ICRC decides to engage in public criticism (Forsythe 2012*b*, 466). Therefore, when the ICRC remains silent, it gives the impression that the visited state is not bad and perhaps even making positive progress.

For example, in early 2002, the ICRC started its visits to Guantanamo in Cuba

and Bagram Air Force base in Afghanistan. The president of the ICRC went to Washington several times to meet with leaders of the Bush Administration. Thus, the ICRC knew and clearly reported on much of the mistreatment of detainees in Guantanamo, Afghanistan, and Iraq (Forsythe 2012*b*, 470). However, in a series of press conferences, the ICRC said vaguely that the discreet dialogue with the U.S. was not without results and that Washington had taken into account certain ICRC requests. They maintained this position until the notorious pictures of humiliation, torture, and even wrongful death that appeared in the international press in early 2004 (Forsythe 2012*b*, 472). By privileging discretion in exchange for access to information, HROs could actually, over time, undermine the very human rights standards they supposedly seek to protect by obscuring their findings in the face of continuing abuse of individuals. Criticizing direct engagement with governments, Kenneth Roth, director of Human Rights Watch, argues that any direct contact with governments may grant “moral legitimacy” to those governments, working against HROs’ principles (Roth 2004). Thus, HROs have to ask themselves whether they pay too high a price for this access built on cooperation. In fact, not every HRO is willing to pay this price for information; for some HROs, this kind of information is too costly.

Moreover, vague statement without legal finger pointing, is less likely to bring social mobilization that is necessary for making naming and shaming effective. By engaging in naming and shaming, HROs try to educate the local population about human rights in order to increase domestic mobilization for improved human rights

tation seeks to explain what determines HROs' naming and, once named, how they shame states for their human rights records. While the dominant explanations in the literature recognize that information is important for the naming and shaming, they assume that information is abundant and information gathering is costless. However, the challenge of gathering information is more complex than that.

My argument is built around the idea that information gathering is the key to understanding the politics of naming and shaming. In an indirect approach HROs rely on third parties for information such as the victims, families of the victims, political opposition parties, domestic NGOs, etc. They do not engage with the governments. Yet, one characteristic of this indirect approach is that more information comes from states with a strong civil society. Consequently, HROs tend to name and shame states with a strong civil society despite the fact that they have relatively better human rights records than the ones without it. In a direct approach, HROs get information from direct visits and direct communication with governments. Since HROs can observe the conditions directly, they can have more accurate understanding of the situation. However, one consequence of a direct approach is that HROs have to negotiate with the government for access. If an HRO criticizes the government's performance too harshly or reveal too much information about conditions, the government can cut off access and prevent access to the country. This may lead the HROs hold on to some information and to be less critical of the government in order to maintain access. How HROs gather information influences how they do naming and shaming.

Chapter 3

Measuring Naming and Shaming

3.1 Introduction

Despite the importance of human rights organizations (HROs), it is remarkable that existing studies on HROs have been very limited. One potential reason for the lack of relevant empirical analyses may be due to the coarseness and the lack of data on human rights organizations. Unlike international conflict studies, to date, existing data of HROs have been scant. For example, *the Correlates of War Project (COW)* was founded in 1963 and has assembled the dataset on the incidence, events, and extent of inter-state and extra-systemic war from 1816 to 2010 (Ghosn, Palmer and Bremer 2014). Another example would be *the International Crisis Behavior Project (ICB)*. The ICB project contains 455 international crises, 35 protracted conflicts, and 1000 crisis actor from the end of World War I though 2007 (Brecher and Wilkenfeld 2014). These conflict event data has become a standard for international conflict

scholars. However, despite its growing importance, there is no international “clearing house” for human rights event data and the activities of HROs.

In this chapter, I suggest a new human rights event data on naming and shaming. There have been a few HRO event data, but their scope has been very limited. In order to adequately understand HROs and their activities, we need new data as conflict scholars do in their event data organized in a “who” did “what” to “whom” manner for each event.

3.2 Existing Approaches

3.2.1 Obtaining Sources

3.2.1.1 International Non-Governmental Organization

With the increasing attention devoted to human rights concerns in the international community, there have been a few quantitative works paying attention to the process of “naming and shaming” itself, that is, the dynamics of HRO’s strategies and dynamics (Lebovic and Voeten 2006, Meernik, Aloisi, Sowell and Nichols 2012, Ron, Ramos and Rodgers 2005).

One of the major sources of naming and shaming data is Amnesty International (AI). It is understandable because there is little doubt that Amnesty’s role and status is substantial. AI has the longest history and broadest name recognition in the human rights field, won the Nobel Peace Prize in 1977, and is believed by many to set standards for the movement as a whole (Ron, Ramos and Rodgers 2005). It

was an early pioneer in international human rights advocacy efforts, and has made important contributions to the international normative system (Clark 2001, Ron, Ramos and Rodgers 2005). Over 400 paid and volunteer staff work in its London International Secretariat, and according to annual reports and financial audits, the Secretariat's nominal budget grew from \$22.114 million in 1992, to \$49.296 million in 2013. According to one estimate, Amnesty's global network members, sympathizers, and subscribers include 1.5 million persons living in 150 countries and territories, and 81 national offices (Amnesty International 2013). Former Amnesty employees are spread throughout the broader transnational world, diffusing the groups principles, tactics, and worldviews.

To capture the naming and shaming behavior of Amnesty International, a number of scholars have used background reports and press releases found in *the Amnesty International Cumulative Guide 1962-2000* (Hafner-Burton 2008, Krain 2012, Meernik et al. 2012, Ron, Ramos and Rodgers 2005). Press releases and background reports differ in some ways. The lengthier background reports are sent to human rights professionals, UN officials, academics, and feature journalists, while shorter press releases aim more at the general public and non-specialized media. Some of them also use Amnesty International's Urgent Action (Hendrix and Wong 2013, Meernik et al. 2012). The Urgent Action is one of the tools AI uses to spread information regarding human rights violations and to place pressure on governments in the hope that they will improve their human rights policies. The UA is intended to provide information and emphasize the "urgency" of the situation to prevent further human

rights violations. Individuals and groups who receive the UA notices are encouraged to write letters on behalf of the individual whose human rights have been violated.

3.2.1.2 Inter-Governmental Organizations

A few scholars have used data from inter-governmental organizations (IGOs) like the United Nations (Conrad and Moore 2010, DeMeritt 2012, Forsythe 2009, Hafner-Burton 2008, Krain 2012). In particular, they all rely on Lebovic and Voeten (2006)'s data on the United Nations Commission on Human Rights (UNCHR). Lebovic and Voeten (2006) used available UN documents to understanding the UNCHR's naming and shaming activities during the period of 1977-2001. The UNCHR, for sixty years the traditional center for UN diplomacy on human rights, was dissolved during 2006 and replaced by a new United Nations Human Rights Council. Its major achievement was the drafting of the 1948 Universal Declaration of Human Rights (UDHR). The half-dozen or so major international human rights treaties that followed in the wake of the 1948 UDHR, the norms negotiated through the UNCHR during its existence did establish, for the most part, what was required for a life with human dignity (Forsythe 2009). The UNCHR had the authority to discuss and recommend, eventually to name and shame. The 53-member commission adopted resolutions that specified general human rights standards that governments should maintain and it approved yearly resolutions that singled out individual countries for failing to live up to these international human rights standards (Lebovic and Voeten 2006). It had the "power" to damage governmental reputations within certain circles

of opinion by naming and shaming.

3.2.1.3 Newspapers and News Wires

A number of scholars have used news reports to capture the naming and shaming activities. Ramos, Ron and Thoms (2007) used news reports of specific human rights abuses in *The Economist* and *Newsweek* during the years 1986-2000. Both magazines are major Northern media sources, although they differ in important ways. *The Economist's* readers are comparatively wealthy, educated, influential, and internationally oriented (Ramos, Ron and Thoms 2007). In 2002, *The Economist's* circulation was 880,000, half of which lived in North America; 20% in continental Europe; 15% in the UK; and 10% in Asia (Ramos, Ron and Thoms 2007). So, *The Economist* is an indicator of the type of news read by elite Northerners interested in international affairs. *Newsweek's* 19.5 million North American readers are far more numerous, but their income and educational levels are substantially lower (Ramos, Ron and Thoms 2007). Krain (2012) extended the data to the period of 2001-2008. In this vein, (Franklin 2008) used full text news wire reports from dozens of wire services indexed on *Lexis-Nexis*, and *Keesing's Record of World Events, Facts on Files archives*. This involves the reading of thousands of news articles in order to identify human rights abuses during 1981-1995.

In order to more adequately capture the activities of human rights organizations, Murdie and Davis (2011) examine all *Reuters Global News Reports* during the period of 1992-2007. By using all *Reuters Global News Reports*, Murdie and Davis

(2011) captured considerably more HRO events than any other existing studies. In comparison to other news media sources, Reuters has news offices in over 200 locations around the world (Murdie and Bhasin 2011). It gives the *Reuters*-based measures an advantage in accessing news and more accurate coverage of the country and human rights cases.

3.2.2 Data Coding: Hand-Coded vs. Machine-Coded Data

As I have indicated earlier, one of the primary reasons for the lack of relevant empirical analyses have to do with the lack of data on human rights organizations' naming and shaming. Compared to other topics in the field of international relations such as international conflicts which have variety of event data (e.g, Correlates of War Data (Ghosn, Palmer and Bremer 2014) or International Crisis Behavior Project (Brecher and Wilkenfeld 2014)), why are there so little data on the topic? I argue that it is partially because almost all the existing data are hand-coded.

The most widely used naming and shaming data, Ron, Ramos and Rodgers (2005)'s Amnesty data is a good example of hand-coded data in order to build their human rights event data on naming and shaming. Reading through Amnesty's background reports and news releases, three coders identified shaming targets of each report. Three coders conducted regular coding meetings to ensure consistency. It took a long time for them to complete about 7,000 shaming events. Acknowledging the amount of information to code would be too great to realistically allow a global sample, instead of coding all the countries as the sample of observation, Franklin

(2008) had to use a random sample of 7 Latin American countries (Argentina, Brazil, Chile, Guatemala, Mexico, Nicaragua, and Venezuela).

In the early stage of machine-coded data, the lack of readily available sources and sufficient computing power to support machine coded efforts meant that human coding was the only viable coding option. It is understandable that human coding was initially the only available way to code events.

However, it had three main shortcomings: it is slow, expensive, and subjective. Assuming that the average human coder can code around six to ten events (one event per document) an hour on a sustained basis, and very few people can reliably code more than a few hours a day because the process is so mind-numbingly boring, at that rate, it would take a team of 10 coders at least 1.5 person-years to code 40,000 documents. Paying coders \$10 an hour would cost \$50,000, and the costs to training, re-training, cross-checking and management would at least double that investment. Additionally, due to the inherently subjective nature of human analytical processes, inter-coder reliability between analysts rarely exceeded 70% and often falls in the 30%-40% range (King and Lowe 2003, Mikhaylov, Laver and Benoit 2012) particularly when coding is done across institutions and over long periods of time.

In international conflict studies, the machine-coded data started as early as 1980s. By the late 1980s, computational power had advanced to the point that it was possible to run automated coding software from personal computers. The Kansas Event Data Set (KEDS) in the late 1980s (Schrodt 1994, Schrodt and Donald

1990) was the first attempt to use a computer to parse through electronic text and code relevant events into an event data database, relying on dictionary-driven sparse parsing based on *the World Event/Interaction Survey (WEIS)*. By the late 1990s, machine-coding had become increasingly popular, and almost all time and costs were upfront in the dictionary and software development phase. Because these were open source, they were easily adopted and upgraded. In 2000, the KEDS projects launched *the Textual Analysis By Augmented Replacement Instructions (TABARI)* software, which became the dominant machine coding system in event data. Subsequent automated-coding software, like the proprietary VRA Reader used to build the King and Lowe (2003)'s 10 Million International Dyadic Events datasets, was modeled off of TABARI.

In 2012, Kale Leetaru released a new event dataset called the Global Database of Events, Location, and Tone (GDELT). This dataset combined the strengths of King and Lowe (2003) dataset and the ICEWS dataset. It also uses an advanced natural language processing program to provide latitude and longitude coordinates for each event, which not only combines the strengths of the 10 Million Dyadic Events (i.e. global coverage) with ICEWS, but also provides latitude and longitude coordinates for the events. GDELT uses TABARI and the CAMEO to machine-code the entire content of electronic news stories. Additionally, GDELT obtains news stories from four sources, *LexisNexis*, *Agence France Presse*, *Reuters*, *Associated Press*, and *Xinhua*. GDELT has been considered a new, cutting edge event dataset. The dataset include 200 million plus events from 1979 to 2013. Each observation contains up to

70 columns of additional information regarding the actors and location of the event.

Although these machine-coded data have been flourishing in the field of international conflict studies, it was not until Murdie and Davis (2011) first used it for analyzing HRO's shaming activities. Using the King and Lowe (2003)'s IDEA dataset, Murdie and Davis (2011) created a dataset of all events in *Reuters Global News* reports concerning international NGOs daily from 1992-2007. In order to more adequately capture the activities of human rights INGOs, they created a dictionary of the names of the 432 human rights INGOs listed in *the Yearbook of International Organizations*. It includes a sample of roughly 130 states and around 1,300 country-year observations. This data is organized in a "who" did "what" to "whom" manner for each particular event, over 10 million events in the complete dataset (King and Lowe 2003). In particular, they include 432 HROs which is very impressive because all the other existing studies have only looked at a single HRO, i.e., Amnesty International. This procedure produced 9,543 distinct events where human rights INGOs are the source actor and a government or government official is the target. They are aggregated by yearly-country level.

On the basis of the dataset, Murdie and Peksen (2014) expanded the data to the next level, that is measuring the shaming intensity. They use the Goldstein (1992) scale to measure the level of intensity for each shaming event in order to investigate whether the extent of critical statements and actions by HROs increase the probability of a humanitarian intervention in a potential target country.

3.2.3 Critiques

Although the existing approaches are impressive there are a number of problems. First, as I have discussed in Chapter 2, relying on a single HRO is problematic. While Amnesty International is one of the most influential HROs, there is a great deal of bias that might result from studying only one organization. *The Yearbook of International Organizations* alone includes more than 500 human rights organizations (UIA 2011). It would be very difficult to assume that AI represents the entire HROs in the world.

Second, although UNCHR is one of the most important human rights organs in the UN, it has a quite different mechanism from other human rights organizations in terms of its decision making process. Unlike other HROs, UNCHR is a UN Charter-based body and its decision making is based on 53 state-members consensus.

Third, in many respects, the use of the global news sources make sense and it is quite impressive to include more than 400 HROs, because all the other existing studies have only looked at a single HRO. However, a number of scholars have indicated that the perceptions about newspapers-generated data are not quite accurate (Davenport 2007, Davenport and Ball 2002, Sommer and Scarritt 1999). On the basis of these data and statistical analyses, scholars have shown three problems. First, there are so-called “threshold effects,”—that is, only events above a certain level of significance receive coverage, because they appeal to a large market (i.e. readerships and advertisers) and cost less to cover than smaller events (Davenport 2007, McCarthy, McPhail and Smith 1996). Second, there are so-called “fatigue”

effects—that is, only events of short duration receive coverage, in an effort to hold costs down (Davenport 2007, Gerner and Schrodtt 1996). For example, the coverage of a protracted conflict may be high when hostilities first break out, then decline steadily as reporters, editors, and readers become “bored” with the issue. This is also often called “Media Issue Attention Cycles.” It refers to the sudden ascendance of an issue from previous obscurity to a sustained prominence that dominates the news for a period of time before it fades from media attention. Because issue attention cycles imply the movement of issues on and off the mass media agenda, such cycles may influence the short-term news worthiness of a particular demonstration event. This issue cycles, thus, may bias story selection in part because they can push out the media limelight, but more so because media personnel have a tendency to connect these cycles with ongoing events, selecting for stories those events which will receive subsequent media attention depending on ebbs and flows of media attention cycles. This is even true for online-news or social media. For example, Leskovec, Backstrom and Kleinberg (2009), on the basis of a framework for tracking on-line news, demonstrate dynamics of new cycle between the peaks of media attention using a mathematical model for temporal variation.

Third, there are also “news-hole” effect—that is, coverage is determined by the amount of available space within the pages of a newspaper (Davenport 2007, Hornig, Templin and Walters 1991). Because of space limitations, a full description of international coverage in each country during the time frame would be very difficult. Thus, international news coverage is uneven not only at the level of each individual

country, but also at the level of the world as a whole (Wu, 2001).

Thus, one of the best ways to identify the level of human rights violation is to rely on organizations that are more specifically focused on human rights violations (Ball, Kobrak and Spierer 1999, Davenport 2007, Gibney and Dalton 1996, Poe and Tate 1994). Although news organizations talk about human rights violations or human rights issues, news media are basically profit-making institutions by design, and they are often very successful ones. They depend at least much on advertisers as on consumers for these profits. Therefore, human rights are not necessarily their primary concerns. However, unlike news organizations, these human rights organizations are directly and exclusively interested in state repression and human rights violations. Toward these ends, relevant organizations have well-developed networks of communication both within and across countries. This directly aids them in overcoming many of the difficulties that news-generated data have (Davenport 2007).

3.3 An Alternate Approach

Creating a new data is an effort to overcome and complement the existing data on human rights organizations' naming and shaming. First, I briefly discuss the sources of the data. Second, I will discuss how I measure two levels of naming and shaming, that is—count and intensity.

3.3.1 Data

Instead of relying on newspapers or news wires, I directly collect data from human rights organizations. First of all, I produced a list of all human rights organizations that have a mission statement which focuses on core human rights as listed in *The Yearbook of International Organizations (UIA)*. I use the subject headings in *the Yearbook* to list the organizations (i.e. HROs) devoted to the promotion and protection of human rights. Specifically, I focus on the following subject headings, “Human rights Organizations,” “Peace,” “Peace/Discrimination,” “Peace/Equality,” “Peace/Humanity,” “Peace/Justice,” and “Peace/Rights.” I then examine the aim and activities of each HROs listed in these subject headings to determine whether each organization is relevant HROs for the project.¹ There is an alternative source of the data on HROs such as *the Human Rights Internet (HRI)*’s Master List of human rights organizations. This is an alphabetical list of the addresses of organizations concerned with human rights compiled by the HRI. Some people like Hathaway (2007) and Mosley and Uno (2007) use this data source in order to measure the number of human rights organizations within a particular country. However, there are a couple of serious weaknesses for the purpose of this project. First of all, the HRI reports much smaller number of organizations than *the UIA Yearbook*. Second, the HRI’s Master List is only sparsely available so that it has a serious problem of missing data. For example, the HRI’s Master List is available for only the years

¹*The Yearbook*’s classification has been changed. The subject headings that were relevant were “Human Rights Organizations,” “Innovative Change/Discrimination,” “Innovative Change/Equality,” “Innovative Change/Humanity,” “Innovative Change/Justice,” “Innovative Change/Rights.”

1986, 1991, 1994, and 2000 so that Mosley and Uno (2007, 2) rely on interpolation and extrapolation to impute missing data for their analysis.

In addition to *the Yearbook*, I produced another list from the United Nations Economic, Social, and Cultural Council (ECOSOC)'s list of NGOs with consultative status and all the human rights organizations that provide their background or showdown reports to the UN Human Rights Council's *Universal Periodic Review (UPR)*.

ECOSOC consultative status provides NGOs with access to not only ECOSOC, but also to its many subsidiary bodies, to the various human rights mechanisms of the United Nations (UN ECOSOC resolution 1996/31). To be eligible for consultative status, an NGO must have been in existence (officially recognized by a government) for at least two years, and must have an established headquarters, a democratically-adopted constitution, authority to speak for its members, a representative structure, appropriate mechanisms of accountability and democratic and transparent decision-making processes. The basic resources of the organization must be derived mainly from contributions of the national affiliates or other components or from individual members. Organizations established by governments are not considered NGOs. They are also required to submit a quadrennial report every four years. Consultative relationships may be established with international, regional, sub-regional, national non-governmental organizations, non-profit organizations, public sector or voluntary organizations. NGOs affiliated to an international organization already in status may be admitted provided that they can demonstrate

that their programme of work is of direct relevance to the aims and purposes of the United Nations. In regard to national organizations, consultation with the Member State concerned is required.

The Universal Periodic Review (UPR) is a unique process which involves a periodic review of the human rights records of all 193 UN Member States. The UPR was established when the Human Rights Council was created on 15 March 2006 by the UN General Assembly in resolution 60/251. The UPR is a significant innovation of the Human Rights Council which is based on equal treatment for all countries. It provides an opportunity for all States to declare what actions they have taken to improve the human rights situations in their countries and to overcome challenges to the enjoyment of human rights. Currently, no other mechanism of this kind exists. The review itself takes place in Geneva in a session of the Working Group on the UPR, which is composed of the 47 member States of the Human Rights Council. The review takes the form of an interactive dialogue between the State under review and the member and observer States of the Council. At the end of each review, the Working Group adopts an outcome document, which is subsequently considered and adopted by the Human Rights Council at a later session. The UPR process provides for the participation of all relevant stakeholders, including non-governmental organizations (NGOs) and national human rights institutions (NHRIs). Civil society actors and NHRIs can submit information which can be added to the “other stakeholders” report which is considered during the review. Information they provide can be referred to by any of the States taking part in the interactive discussion dur-

ing the review at the Working Group meeting. NGOs can attend the UPR Working Group sessions and can make statements at the regular session of the Human Rights Council when the outcome of the State reviews are considered.

These lists overlap to a certain degree. That is expected because they are not meant to be mutually exclusive, but they supplement one another. Then, I checked each HROs on the list one by one and collect each HRO's press releases, newsletters, and reports, which including naming and shaming activities. All HROs do not necessarily have the same form of press releases or reports. I excluded any reports that do not deal with human rights issues. For example, some reports are primarily about HROs' administrative news, which are irrelevant to this project. It also helped me to exclude those HROs that have not been active for a long time. By "inactive" I mean that they do not show any activities although they still do exist. It is possible that they might be doing something not in public, but my assumption is that one of the primary purposes of naming and shaming is to mobilize public opinion and advertise their activities to current and potential donors (or members). It means that they are supposed to make their reports (or any sorts of information) available to public. This process narrows down the total number of HROs in my data to 93. I primarily relied on each HRO's web resources and also used hard copy publications, if available. The number of total HRO documents is around 40,000.

3.4 Measuring Naming and Shaming

3.4.1 Shaming Counts

On the basis of this new data, first, I measure the number of HROs shaming events that occurred in a given year. Each document is considered a distinct event in which a HRO discusses a state, that is, the target of naming and shaming. In other words, the main objective is to identify which state a given document is primarily talking about. Once identify target states, I aggregate these data to the HRO-country-year. The final variable, HRO shaming counts, is the yearly count of the number of each HRO's shaming events that occurred toward a state in a given year. Since the target of each document is a state, after converting each document into computer-readable format, I used a well-engineered feature extractor, *Stanford Named Entity Recognizer (NER)* (Finkel, Grenager and Manning 2005) in order to extract target country in each document.

Named-entity recognition (NER) is the problem of segmenting and classifying proper names, such as names of people and organization, in text. Named entity recognition addresses the problem of locating textual mentions of predefined types of entities, where the entity categories can be very diverse, ranging from people to states, etc. For example, when given the sentence “Kenneth Roth from Human Rights Watch blamed China for human rights violations,” a name entity recognizer that is targeted to people and locations should identify the three named entities “Kenneth Roth” of type PERSON, and “Human Rights Watch” of type ORGANI-

ZATION, and “China” of type LOCATION.

More recent work on named entity recognition is usually based on statistical machine learning. Many statistical learning-based named entity recognition algorithms treat the task as a sequence labelling problem. Sequence labelling is a general machine learning problem and has been used to model many natural language processing tasks including part-of-speech tagging, chunking and named entity recognition. It can be formulated as follows. We are given a sequence of observations, denoted as $x = (x_1, x_2, \dots, x_n)$. Usually each observation is represented as a feature vector. We assign a label y_i to each observation x_i . While one may apply standard classification to predict the label y_i based solely on x_i , in sequence labelling, it is assumed that the label y_i depends not only on its corresponding observation x_i but also possibly on other observations and other labels in the sequence. Typically this dependency is limited to observations and labels within a close neighborhood of the current position i . In order to identify the target of each HRO document, I use conditional random fields (CRF) from *Stanford NER* in particular.

3.4.2 Linear-Chain Conditional Random Fields

Conditional random fields (CRF) are a discriminative model for sequence labelling. They were introduced by Lafferty, McCallum and Pereira (2001) to address information extraction problems. In CRFs the label of the current observation can depend not only on previous labels but also on future labels. Also, CRFs are undirected graphical models. Usually linear-chain CRFs are used for sequence labeling problems

in natural language processing, where the current label depends on the previous one and the next one labels as well as the observations. There have been many studies applying CRFs to named entity recognition.

Let Y, X be random vectors, $\theta = \{\theta_k\} \in \mathbb{R}^K$ be a parameter vector, and $\mathcal{F} = \{f_k(y, y', x_t)\}_{k=1}^K$ be set of real-valued feature functions. Then a linear-chain conditional random field is a distribution $p(y|x)$ that takes the form:

$$p(y|x) = \frac{1}{Z(x)} \prod_{t=1}^T \exp \left\{ \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, x_t) \right\} \quad (3.1)$$

where $Z(x)$ is an input-dependent normalization function

$$Z(x) = \sum_y \prod_{t=1}^T \exp \left\{ \sum_{k=1}^K \theta_k f_k(y_t, y_{t-1}, x_t) \right\} \quad (3.2)$$

In a CRF, we can allow the score of the transition (i, j) to depend on the current observation vector, simply by adding a feature $1_{\{y_t=j\}} 1_{\{y_{t-1}=i\}} 1_{\{x_t=o\}}$. A CRF with this kind of transition feature, which is commonly used in text applications. To indicate in the definition of linear-chain CRF that each feature function can depend on observations from any time step, we have written the observation argument to f_k as a vector x_t , which should be understood as containing all the components of the global observations x that are needed for computing features at times t . For example, if the CRF uses the next word x_{t+1} as a feature, then the feature vector x_t is assumed to include the identity of word x_{t+1} . Finally, note that the normalization constant $Z(x)$ sums over all possible state sequences, an exponentially large number of terms.

In a linear-chain CRF, the maximum likelihood parameters can be determined using L-BFGS.

3.4.3 Application: Identifying HROs' Targets

In this section, I discuss measuring HROs' shaming targets using Named-Entity Recognition (NER). NER is the problem of segmenting and classifying proper names, such as names of people and organization, in text. My goal is the identification of mentions of entities in text, country name, that is—Location.

There are generally proposed solutions to NER that fall into three categories: 1) The rule-based (Krupka and Hausman 1998); 2) the machine learning based (Finkel and Manning 2009, Singh, Hillard and Leggetter 2010); and 3) hybrid methods (Jansche and Abney 2002). With the availability of annotated corpora, such as ACE05, Enron (Minkov, Wang and Cohen 2005) and CoNLL03 (Tjong Kim Sang and De Meulder 2003), the data driven methods now become the dominating methods.

Rule-based Approaches The rule-based approaches generally attempt to identify named entities by predefined rules. The rules typically describe common naming structures using either orthographic or lexical clues, or morpho-syntactic features, e.g., word alphanumerical composition, the presence of special symbols, capitalization, and special nouns or special verbs.

The strength of the rule-based approach is that rules can be carefully designed to deal with specific linguistic phenomena. The dictionary-based entity extraction

approach is a special case of rule-based methods. Most rule-based approaches have focused on designing specialized high quality rules for specialized sub-domains (Liu et al. 2012). Because of the carefully crafted rules, accuracy obtained with this approach is generally high. For example, Ananiadou, Albert and Schuhmann (2000) implemented a computational morphological grammar and lexicon, and applied it to medial term recognition. Observing that medical terminology heavily relies on Greek and Latin neoclassical elements for creating terms such as “eruthrocyte” and “angioneurotic,” the author proposed a four-level ordered morphology to describe the term formation patterns. Gaizauskas et al. (2003) also used a manually constructed protein name recognition. They first split a protein name into component terms, based on its apparent syntactic structure such as “calmodulin” and “N-methyltransferase.” The rule-based method is sometimes better because each entity type has its own vocabulary for the extraction task.

However, the main problem is that developing good rules usually requires extensive knowledge of linguistics and domain specific knowledge. The rule construction can also be very time-consuming. In order to extract complete answer entity set, all the possible rules or patterns are required for extractions, but, in fact, it is hard to achieve this level of completeness (Liu et al. 2012). Moreover, the defined rules are usually very specific to a particular domain, which makes it difficult to be applied to other domains. It is hard to adapt to new domains, and time-consuming in rule construction.

Machine Learning-based Approaches Machine Learning (ML)-based approach is language independent and more robust in terms of system performance. A small fixed text set is chosen and manually marked up in order to train the function. The trained function produces similar annotations on unseen texts. Therefore, it turns an extraction problem into a learning problem and extracts the phrases with each argument and the types of entities. ML based approach can be further divided into supervised learning and semi-supervised learning methods. Supervised ML utilizes large annotated corpus while semi-supervised ML only needs small size of annotated corpus along with large un-annotated corpus (Jurafsky and Martin 2009). The supervised ML-methods that have been used for NER include Hidden Markov Models (HMM), Maximum Entropy, and Support Vector Machines (SVM), and Conditional Random Field (CRF).

The main strength of the machine learning based approach is that it has demonstrated better performance (better accuracy rate) than the rule-based approach. Moreover, it can incorporate linguistic information as much as possible into features and it is easier to adapt to new domains, compared to the rule-based approach.

The main problem with this approach is that it requires an annotated training data set to learn the function for the entity extraction tasks, which is hard to obtain. Another main challenge is how to select discriminating features. Besides, as the number of features for machine learning systems increases to cover more linguistic information, data sparseness can be a serious problem for the ML methods used, resulting in degraded generalization capability.

Combined Approaches Since the approaches discussed above all have their own strength and weakness, some research works introduced a hybrid of different approaches. For example, Hanisch et al. (2003) used a machine learning technique for computing optimized parameters of scoring measures in a dictionary-based system, while Zhou et al. (2004) automatically constructed rules to deal with cascaded named entities for their machine learning system.

My goal is the identification of mentions of entities in text, country, that is—Location. This is relatively a straightforward task because of the data structure. First, since these documents were collected by each HRO, identifying actor entity is unnecessary. Second, HRO’s documents generally deal with a single country at a time. Of course, there could be more than one country in a document. The assumption is that the most frequently recognized entity is the main target of a given document.

I utilize *Stanford NER* which is trained on the CoNLL 2003 shared task data (Finkel, Grenager and Manning 2005). The data of the CoNLL 2003 shared task consists of news articles from English and German. The English articles are newswire articles from Reuters taken from between 1996 and 1997. The English data consists of a training set of 946 news articles comprising 203,621 tokens, a development set of 216 articles, and 51,362 tokens, and a test set of 231 articles comprising 46,435 tokens. Each of the articles have been manually annotated to indicate the locations and types of all of the named entities. The average F1 score of the Stanford NER trained on the CoNLL 2003 shared task data achieves 92% on its test set (Ratinov

and Roth 2009). Given that the task is relatively very simple, using already trained *Stanford NER* is sufficient for my project. More to the point, the main targets of my interest is 193 country names. All of them are already included in the CoNLL2003's location list.²

First, I extract state names in each document. Second, on the basis of the most frequently recognized entity, I determine the target country in the document. One could argue that this is not necessarily the best way to identify targets states in a given document. However, my approach is not terribly far away from the purpose of the project. First of all, as Kenneth Roth (2004, 67), the director of Human Rights Watch, indicates, unlike newspapers, HROs document include a particular state of affairs amounts to a violation of international human rights standards AND a particular violator who is responsible for the violation. Second, almost all the HRO documents are issued and classified by country. Third, in order to validate the performance of NER, I calculated *F1 score*, 98.6 (*precision* 96.9, *recall* 98.3) on the basis of manually annotated sample documents.³ This is not about NER entity accuracy, but the document level validation on the basis of the most common state entity at the target country.

²They are the member states of the United Nations: See Appendix B

³

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \cdot 100$$

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \cdot 100$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \cdot 100$$

3.4.4 Shaming Intensity

As I have discussed in Chapter 2, almost all the existing studies on HROs naming and shaming have neglected the level of shaming intensity. There are two notable exceptions. Analyzing the United Nations Commission on Human Rights (UNCHR), Lebovic and Voeten (2006) seeks to explain whether or not a state is targeted by the UNCHR, and once shamed, how severely a targeted country is shamed. The UNCHR was for the past sixty years the traditional center for UN diplomacy on human rights and its main action is naming and shaming. In other words, the UNCHR always dealt with words (Forsythe 2009). It had the authority to discuss and recommend, eventually to name and shame. It had the power to damage governmental reputations within certain circles of opinion. But it had no authority to move beyond words, which themselves were nonbinding, to manipulate threats and use of military force and various other forms of sanctions—whether diplomatic or economic. It could criticize, but it could not impose binding sanctions, and it could not threaten or deploy military forces.

Thus, in order to measure the level of shaming intensity, Lebovic and Voeten (2006) operationalize four different levels of UNCHR actions. First, the UNCHR could choose not to act, either by discontinuing the confidential consideration of an allegation or by considering a matter in public without passing a resolution. Second, the UNCHR could continue consideration under confidential session. Third, the commission could initiate a somewhat mild sanction in the form of the advisory procedure or a critical statement from the chair of the commission. Fourth, the

commission could pass a resolution that publicly condemns a state and expresses the reasons for doing so. These four actions can be regarded as the ranking of shaming intensity. The ranking of these options reflects the following assumptions: (a) having a sanctioning resolution fail (i.e., do nothing) is the most favorable outcome from the standpoint of the targeted offender; (b) a public airing of grievances is more severe (shameful) and politically damaging to the offender than a private discussion (assuming that there is some merit to the charges); and (c) that a public vote of condemnation is the least favorable outcome from the standpoint of an alleged offender.

Murdie and Peksen (2014) assess the impact of HRO naming and shaming on the likelihood of humanitarian interventions by using two different measures, HRO shaming (count) and HRO shaming (intensity). They used the Goldstein Scale to measure HRO shaming intensity from King and Lowe (2003)'s IDEA dataset. The Goldstein Scale is a conflict-cooperation scale created by Goldstein (1992). On the basis of the World Event Interaction Survey (WEIS), Goldstein (1992) created a scale ranging from -10 to +10 to measure how conflictual each event is. In his WEIS data, each article (text) was worded as an action taken by country A toward country B. And, a number of panelists rate each event based on -10 (most conflictual) to +10 (most cooperative).

King and Lowe (2003) used their 10 Million International Dyadic Events dataset and mapped it onto Goldstein conflict-cooperation scores, resulting a score for each event ranging from very conflictual (-10) to very cooperative (+10). Utilizing Murdie

and Davis (2011)'s HRO naming and shaming data, Murdie and Peksen (2014) reversed the scale so that a more intensively negative shaming event would be a given higher weight on the scale.

Although Murdie and Davis (2011)'s approach is reasonable there are critical problems from the dataset. The King and Lowe (2003)'s dataset relies exclusively on Reuters newswires. Like most automated coding systems, King and Lowe (2003)'s approach does not consider the scope or intensity associated with each event. In other words, it is unable to extract information about whether a bombing kills 1 persons or kills 1,000 people. Moreover, the dataset relies on just the title and a lead sentence of *Reuter* newswires. Thus, it would be very difficult to measure shaming intensity by just looking at the title. Additionally, the 10 Million Dyad Dataset was designed to capture inter-state interactions only. It is likely to miss a large number of conflictual events within sub-state groups.

3.4.4.1 An Alternative Approach to Measuring Shaming Intensity

As I have discussed in the previous chapter (Theory Chapter and the section above), there are few scholars who have tried to understand or capture variations in HROs' naming and shaming. In this section, I propose a way for measuring shaming intensity. Although naming and shaming is unwieldy and complex concept to measure, there are a few elements that need to be considered for measuring the concept.

First, HROs should be able to show that a particular state of affairs amounts to a violation of international human rights standards. Explaining how to document

human rights events, Dueck, Guzman and Verstapen (2001) indicate that to call something a human rights violation is to make a legal judgment. This judgment leads to consequences such as an official investigation of abuse (and punishment). Thus, it is important to indicate that the acts are incompatible with international human rights laws.

Along with the first element, HROs should be able to show who is responsible for the violation. The purpose of naming and shaming is to pressure repressive government. It is important to point out who committed human rights violations.

Third, I also consider the length of a document. The length in words of human rights reports itself is not necessarily a sufficient indicator of the depth or specificity of the human rights conditions (Kelly 2009). Nevertheless, given the fact that the space in the report is finite resource, the total length of the report could be informative because lengthy articles carry more detailed and specific information about the event⁴.

Fourth, another important aspect of naming and shaming is the tone. The language in human rights reports is usually expected to be restrained, factual and legal statements rather than polemic or emotional rhetoric. However, they are not lim-

⁴In this context, Hafner-Burton and Tsutsui (2005, 1406-1407) propose three categories: extensive reporting, intermediate reporting, and limited reporting. (1) Extensive reporting, which offers detailed information about individuals tortured, specific individuals or governmental units charged with torture, or specific events of torture such as demonstrations and riots, as well as the enactment and implementation of specific national human rights legislation, commissions, policies, and other programs; (2) Intermediate reporting, which offers some brief detail about individuals tortured or accused of torture, as well as the enactment and implementation of national human rights legislation; and (3) limited reporting, which offers information concerning the general occurrence of torture but very little detail about specific individuals tortured, specific individuals or governmental units charged with torture, or the enactment and implementation of specific national human rights legislation, commissions, policies, and other programs.

ited to such usual expectation (Dudai 2007, Hopgood 2006). For example, describing the Guantanamo Bay detention facility, Amnesty International often refers to it as “gulag.” Amnesty does not hesitate to refer to the U.S. counter-terrorism policy as “foolish” (AI 2008:4). In this context, HROs often use the words “extreme,” “gross,” “hell,” “horrible,” etc. Not all the HROs use such language in their report. Some of them are very minimalist and restrained to factual and human rights legal assertions. In addition, as I have discussed in chapter 2, HROs reports are not just about “blaming.” They also offer “reward” for good behaviors. In other words, it is often neglected that HROs’ reports do include compliments. This is important, because it shows that a given state has been trying to improve its human rights conditions and it could also imply that the lingering violations are not its systematic policy, but “noises.”

3.4.4.1.1 Text Classification In order to estimate the shaming intensity of a given HRO document, I utilize textual analysis using supervised machine learning technique. More specifically, given a set of training documents labelled as either positive or negative, I learn a classifier that can be used on other (unseen) documents to predict the label for those documents. In particular, I use the probability of classification as a measure of the intensity of the documents (for later analysis). If the classifier is confident about the classification on a given document, the probability of the document having the predicted label is high and this can be used as a proxy for how *strongly* the given document reflects its predicted label. On negative documents, the strength is equivalent to the shaming intensity. In the experimental

setup, a random sample of 2,000 documents were drawn from the collected human rights documents. These 2,000 documents were divided into 1,600 documents for training set, 200 documents for development set, and the remaining 200 documents for test set.

After tokenization, each document is represented by a feature count vector. I use a simple unigram, commonly referred to as “the bag of words” model. In this approach, a vocabulary of words is defined for the representation, which are all possible words that might be important to classification. This is usually done by extracting all words occurring above a certain number of times, and defining feature spaces so that each dimension corresponds to one of these words. When representing a given document, the value of each dimension—that is, attribute—is assigned based on whether the word corresponding to that dimension occurs in the given document. I did experiments on using the frequency of the occurrence of the features as the feature values. After feature selection, each document is represented by a vector of word occurrences for each category where each vector component corresponds to a word feature selected for the category in the previous step. Then, it is converted to a numeric vector representing a given document. Because of the vast number of different words that may appear in text, generally the numerical vectors of word occurrences one gets are sparse vectors of very high dimensionality.

Formally, a two-class classification problem is to determine a label, for example, $y \in \{-1, 1\}$ associated with a vector x of input variables. A useful method for solving this problem is using linear discriminant functions, which consist of linear

combinations of the input variables. Various techniques have been proposed for determining the weight values for linear discriminant classifiers from a training set of labelled data. In this project, I use the Logistic Regression Classifier.

3.4.4.2 Logistic Regression Classifier

In text classification, the output y we are trying to predict takes on one from a small set of discrete values. Consider the simplest case of binary classification, where we want to classify whether some observation x is in the class (true) or not in the class (false). In other words, y can only take on the values 1 (true) or 0 (false), and we would like a classifier that can take features of x and return true or false. Furthermore, instead of just returning the 0 or 1 value, we would like a model that can give us the *probability* that a particular observation is in class 0 or 1. This is important because in most real-world tasks we are passing the results of this classifier onto some further classifier to accomplish some task.

Suppose we just tried to train a linear model to predict a probability as follows:

$$P(y = \text{true} | x) = \sum_{i=0}^N w_i \times f_i \quad (3.3)$$

$$= w \cdot f \quad (3.4)$$

The problem with this model is that there is nothing to force the output to lie between 0 and 1. The expression $\sum_{i=0}^N w_i \times f_i$ produces values from $-\infty$ to ∞ . How can we fix this problem? Suppose that we keep our linear predictor $w \cdot f$, but instead of having it predict a probability, we have it predict a ratio of two

probabilities. Specifically, suppose we predict the ratio of the probability of being in the class to the probability of not being in the class. This ratio is called the odds; if an event has probability 0.75 of occurring and probability 0.25 of not occurring, we say the odds of occurring is $0.75/0.25=3$. We could use the linear model to predict the odds of y being true:

This last model is close; a ratio of probability can lie between 0 and ∞ . But we need the left-hand side of the equation to lie between $-\infty$ to ∞ . We can achieve this by taking the natural log of this probability:

$$\ln \left(\frac{p(y = true|x)}{1 - p(y = true|x)} \right) = w \cdot f \quad (3.5)$$

Now both the left and right hand lie between $-\infty$ to ∞ . This function on the left (the log of the odds) is known as the logit function:

$$logit(p(x)) = \left(\ln \frac{p(x)}{1 - p(x)} \right) \quad (3.6)$$

The model of regression in which we use a linear function to estimate, not the probability, but the logit of the probability, is known as logistic regression.

$$P(y = true|x) = \frac{e^{w \cdot f}}{1 + e^{w \cdot f}} = \frac{1}{1 + e^{-w \cdot f}} \quad (3.7)$$

The last equation is now in the form of what is called the logistic function.

$$P(Y = 1|X) = \frac{1}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)} \quad (3.8)$$

Also, we have

$$\begin{aligned} P(Y = 0|X) &= 1 - P(Y = 1|X) \\ &= \frac{\exp(w_0 + \sum_{i=1}^n w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)} \end{aligned} \quad (3.9)$$

3.4.4.3 Estimating Parameters for Logistic Regression

One reasonable approach to training Logistic Regression is to choose parameter values that maximize the conditional data likelihood. The conditional data likelihood is the probability of the observed Y values in the training data, conditioned on their corresponding X values. We choose parameters W that satisfy

$$W = \arg \max_W \prod_l P(Y^l | X^l, W) \quad (3.10)$$

where $W = (w_0, w_1, \dots, w_n)$ is the vector of parameters to be estimated, Y^l denotes the observed value of Y in the l_{th} training example, and X^l denotes the observed value of X in the l_{th} training example. The expression to the right of the $\arg \max$ is the conditional data likelihood. Here we include W in the conditional, to emphasize that

the expression is a function of the W we are attempting to maximize. Equivalently, we can work with the log of the conditional likelihood:

$$W = \arg \max_W \sum_l P(Y^l | X^l, W) \quad (3.11)$$

This conditional data log likelihood, which we will denote $l(W)$ can be written as

$$l(W) = \sum_l Y^l \ln P(Y^l = 1 | X^l, W) + (1 - Y^l) \ln P(Y^l = 0 | X^l, W) \quad (3.12)$$

Note here that we are utilizing the fact that Y can take only values 0 or 1, so only one of the two terms in the expression will be non-zero for any given Y^l .

To keep our derivation consistent with common usage, we will flip the assignment of the boolean variable Y so that we assign

$$P(Y = 0 | X) = \frac{1}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)} \quad (3.13)$$

$$P(Y = 1 | X) = \frac{\exp(w_0 + \sum_{i=1}^n w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^n w_i X_i)} \quad (3.14)$$

In this case, we can reexpress the log of the conditional likelihood as:

$$l(W) = \sum_l Y^l \ln P(Y^l = 1|X^l, W) + (1 - Y^l) \ln P(Y^l = 0|X^l, W) \quad (3.15)$$

$$= \sum_l Y^l \ln \frac{P(Y^l = 1|X^l, W)}{P(Y^l = 0|X^l, W)} + \ln P(Y^l = 0|X^l, W) \quad (3.16)$$

$$= \sum_l Y^l (w_0 + \sum_{i=1}^n w_i X_i^l) - \ln(1 + \exp(w_0 + \sum_{i=1}^n w_i X_i^l)) \quad (3.17)$$

where X_i^l denotes the value of X_i for the l_{th} training example. Note the superscript l is not related to the log likelihood function $l(W)$.

There is no closed form solution to maximizing $l(W)$ with respect to W . Therefore, one common approach is to use gradient ascent, in which we work with the gradient, which is the vector of partial derivatives. The i_{th} component of the vector gradient has the form

$$\frac{\partial l(W)}{\partial w_i} = \sum_l X_i^l (Y^l - \hat{P}(Y^l = 1|X^l, W)) \quad (3.18)$$

where $\hat{P}(Y^l|X^l, W)$ is the Logistic Regression prediction using equation above and the weights W . To accomodate weight w_0 , we assume an illusory $X_0 = 1$ for all l . This expression for the derivative has an intuitive interpretation: the term inside the parentheses is simply the prediction error; that is, the difference between the observed Y^l and its predicted probability. Note if $Y^l = 1$ then we wish for $\hat{P}(Y^l = 1|X^l, W)$ to be 1, whereas if $Y^l = 0$ then we prefer that $\hat{P}(Y^l|X^l, W)$ be 0 (which makes $\hat{P}(Y^l = 0|X^l, W)$ equal to 1). This error term is multiplied by the value of X_i^l ,

which accounts for the magnitude of the $w_i X_i^l$ term in making this prediction.

Given this formula for the derivative of each w_i , we can use standard gradient ascent to optimize the weights W . Beginning with initial weights of zero, we repeatedly update the weights in the direction of the gradient, on each iteration changing every weight w_i according to

$$w_i \leftarrow w_i + \eta \sum_l X_i^l (Y^l - \hat{P}(Y^l = 1 | X^l, W)) \quad (3.19)$$

where η is a small constant (e.g, 0.01) which determines the step size. Because the conditional log likelihood $l(W)$ is a concave function in W , this gradient ascent procedure will converge to a global maximum. In many cases where computational efficiency is important it is common to use a variant gradient ascent called conjugate gradient ascent, which often converge more quickly.

Overfitting the training data is a problem that can arise in Logistic Regression, especially when data is very high dimensional and training data is sparse. One approach to reducing overfitting is regularization, in which we create a modified “penalized log likelihood function,” which penalizes large values of W . One approach is to use the penalized log likelihood function

$$W = \arg \max_W \sum_l P(Y^l | X^l, W) - \frac{\lambda}{2} ||W||^2 \quad (3.20)$$

which adds a penalty proportional to the squared magnitude of W . Here λ is a constant that determines the strength of this penalty term. Modifying our objective

by adding in this penalty term gives us a new objective to maximize. It is easy to show that maximizing it corresponds to calculating the MAP estimate for W under the assumption that the prior distribution $P(W)$ is a normal distribution with mean zero, and a variance related to $\frac{1}{\lambda}$. Notice that in general, the MAP estimate for W involves optimizing the objective

$$W = \sum_l \ln P(Y^l | X^l, W) + \ln P(W) \quad (3.21)$$

and if $P(W)$ is a zero mean Gaussian distribution, then $\ln P(W)$ yields a term proportional to $\|W\|^2$. Given this penalized log likelihood function, it is easy to rederive the gradient descent rule. The derivative of this penalized log likelihood function is similar to our earlier derivative, with one additional penalty term

$$\frac{\partial l(W)}{\partial w_i} = \sum_l X_i^l (Y^l - \hat{P}(Y^l = 1 | X^l, W)) - \lambda w_i \quad (3.22)$$

which gives us the modified gradient rule

$$w_i \leftarrow w_i + \eta \sum_l X_i^l (Y^l - \hat{P}(Y^l = 1 | X^l, W)) - \eta \lambda w_i \quad (3.23)$$

In cases where we have prior knowledge about likely values for specific w_i it is possible to derive a similar penalty term by using a normal prior on W with a non-zero mean.

Thus, learning is formulated as the optimization problem above. After this

learning, we can use the posterior probability with the learned parameters W to predict the class label given a document. And, we use the posterior probability as a measure of confidence of the classifier and thereby the shaming intensity of the documents. The posterior probability, $\sigma(s)$ is given by the dot product as :

$$\sigma(s) = P(Y|s) = P(Y|X, W) \quad (3.24)$$

3.4.4.4 Evaluation

The classification accuracy from using Logistic Regression classifier with unigrams as features is 82.50%. I tried other commonly used classifiers such as Naive Bayes, Support Vector Machine (SVM) with unigram, bigram, and trigram features. However, I find that Logistic Regression with unigram features produces the best accuracy rate (Table 3.1). In the experimental setup, a random sample of 2,000 documents were drawn from the collected human rights documents. These 2,000 documents were divided into 1,600 documents for training set, 200 documents for development set, and the remaining 200 documents for test set.

In order to evaluate the measured intensity of shaming, I manually rate a random sample of 200 documents and compare them with the shaming intensity produced by the classifier (Figure 3.1). Figure 3.2 shows the effects of regularization on shaming intensity. In supervised learning settings, with many input features, overfitting is usually a potential problem unless there is sufficient training data. The goal of regularization is to penalize large weights to avoid the problems of overfitting the

data. Using too small a regularization parameter results in overfitting, and too large a value results in underfitting (Ng 2004). In order to select the best regularization that maximizes the correlation coefficients, I tried a set of λ values. Figure 3.2 illustrates the effect of a set of regularization values on correlation coefficients ($\lambda = 10^{-3}$ maximizes the correlation coefficients between two measures).

Approach	Number of Features	NB	LR	SVM	Baseline
Unigram	38,213	78.50	82.50	81.50	72.50
Unigram + Bigram	306,686	78.00	82.00	81.00	–
Unigram + Bigram + Trigram	609,101	78.50	82.00	80.50	–

Table 3.1: Classifiers Experiments: Approaches and Accuracy Rates

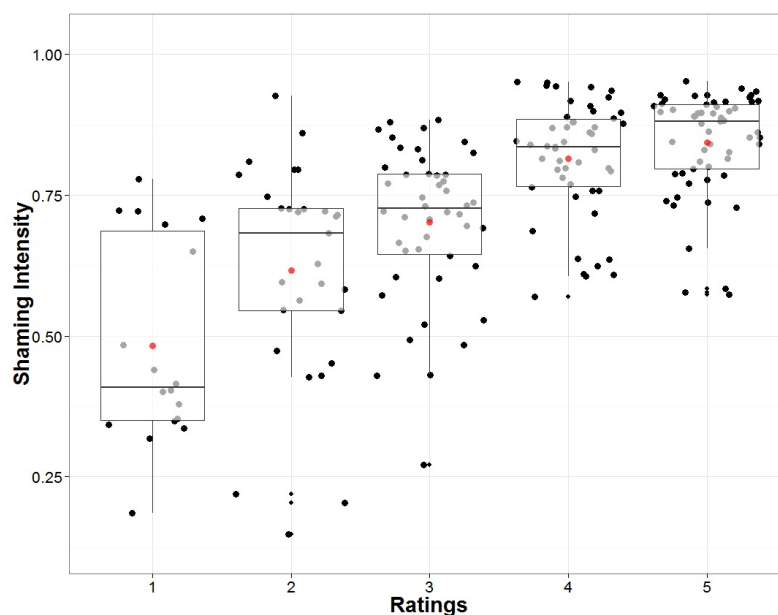


Figure 3.1: Predicted Shaming Intensity and Manual Ratings

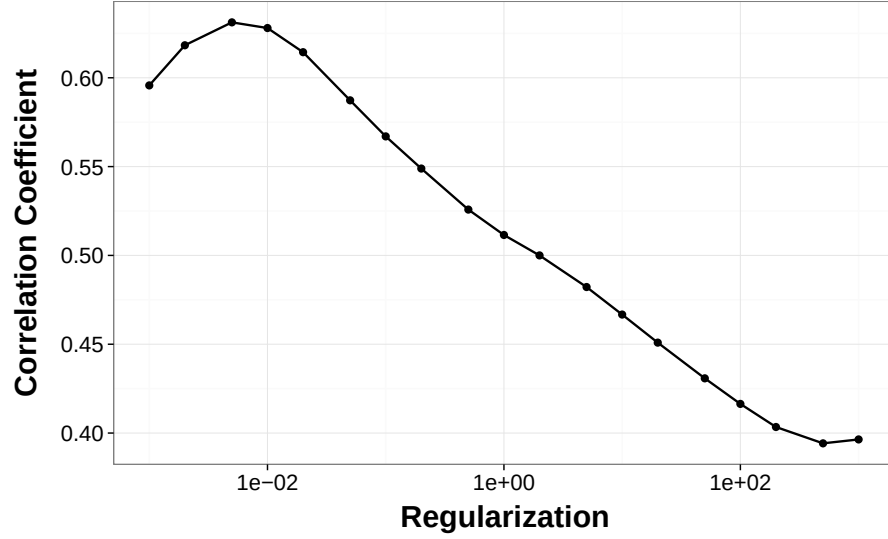


Figure 3.2: Regularization and Correlation Coefficients

The correlation coefficient between manual ratings and posterior probability is around 0.64. In order to use for the analysis, they are all aggregated at the yearly-country level.

3.5 Conclusion

While most of the existing quantitative works focus on examining the efficacy of naming and shaming, they have measured it based on the simple yearly count of the number of HRO shaming events that occurred toward a government in a given year. Thereby, they have implicitly assumed that all the naming and shaming events are essentially the same and equal. In this chapter, I proposed an alternate approach to capture the variations in naming and shaming events across a variety of HROs.

In addition to the theoretical assumption that all the HROs work in a coordi-

nated manner to bring the coherent pressure to repressive governments, a primary reason why scholars have not paid much of the attention to the variation is the lack of empirical data. Thus, I have used the natural language processing approach in order to process thousands of documents much faster and more efficiently. I do not claim that this approach is the best approach. There are a few different approaches to it as discussed in this chapter. Although this approach is not perfect, I believe, it is relatively simpler and easier to execute and process a large number of documents.

Chapter 4

Naming and Shaming Counts Models

4.1 Introduction

Chapter 2 (Theory Chapter) outlines a set of conditions under which states are more likely to get shamed by HROs. This chapter tests these hypotheses through statistical analyses. This chapter proceeds in four parts. Section II discusses the unique data on HROs' shaming. Section III estimates a set of regression models to test hypotheses. Section V concludes.

4.2 HROs and Shaming Targets

The theoretical framework laid out in Chapter 2 suggests that gathering information about human rights practices is difficult. States' human rights practice takes

place domestically; thus, despite the widespread development of the accountability mechanisms, many human rights violations truly are difficult to detect, to observe, and to verify. Given the secretive nature of human rights abuses, it is important for HROs to gather information that can be effectively put together to influence the implementation of human rights by governments. In order to inspire corrective efforts by governments, HROs must demonstrate that their factual statements are true and thus constitute a reliable basis for remedial governmental policy. Simply “knowing” human rights violations are occurring is not enough. Otherwise, HROs will lose adherents and influence. HROs make efforts to find information about human rights practices. Their credibility and legitimacy depend in large measure on the reliability of the facts obtained through such efforts. And the reliability of their information depends, in turn, on the methods used to obtain such information.

In Chapter 2, I suggest that there are two broad approaches to information gathering: an indirect approach and a direct approach.

HROs with indirect approach are more likely to recognize and report on violations in the states with more information. This seems to be especially pronounced for states with a strong civil society. Although states with a strong civil society generally accord greater respect to fundamental human rights than non-democratic countries with a weaker, less active civil society do, HROs report of abuse may be higher for democratic states with a strong civil society and better records, because they encourage freedom of speech, press, and open debate. In other words, HROs have access to more human rights violations for these countries, though the viola-

tions themselves are often less egregious than in more autocratic countries. This skews reporting of human rights violations.

Direct information gathering is one of the most important tools for an effective investigation of human rights violations such as torture, ill-treatment, and conditions of detention. Through direct observation, HROs can develop a fairly good understanding of the specific root causes of human rights abuses in the respective societies as well as of the measures needed to combat those violations and to improve the conditions. On the basis of this assessment, HROs publicize their findings and specific recommendations for further actions. Since they do not need to rely on domestic opposition groups for information, and they are willing to directly communicate with the governments, I predict that HROs that rely on direct information gathering tend to focus on the worst human rights violators.

This argument suggests the following implications.

H1A: HROs with indirect information gathering are more likely to shame states with a strong civil society.

H1B: HROs with indirect information gathering are more likely to be harsh on states.

4.3 Human Rights Shaming: Counts

In order to examine the validity of the hypotheses listed in the theory chapter, I use new data on shaming of countries by HROs during the time period 1991-2011. As I have discussed in the previous chapter, there have been a number of shaming data

(Lebovic and Voeten 2006, Meernik et al. 2012, Murdie and Davis 2011, Ramos, Ron and Thoms 2007, Ron, Ramos and Rodgers 2005), but they are inappropriate for my analyses. I use newly created data for the project (See Chapter 3 for details).

4.3.1 Dependent Variable

On the basis of the new data, I measure the number of HROs shaming events that occurred in a given year. Each document is considered a distinct event in which a HRO discusses a state, that is, the target of naming and shaming. After identifying target states, I aggregate the data to the HRO-country-year. HRO shaming counts are the yearly count of the number of each HRO's shaming events that occurred toward a given in a particular year. Since the target is each state, after converting each document into a computer-readable format, I used a well-engineered feature extractor, *Stanford Named Entity Recognition (NER)* (Finkel, Grenager and Manning 2005) to extract the target country in each document.

My goal in this section is the identification of mentions of entities in text, country name, that is—LOCATION. This is a relatively straightforward task because of the data structure. First, I extract state names in each document. Second, on the basis of the most frequently recognized entity, I determine the target country in the document. One could argue that this is not necessarily the best way to identify targets states in a given document; however, my approach is not terribly far away from the purpose of the project. First of all, as Kenneth Roth (2004, 67), the director of Human Rights Watch, indicates, unlike newspapers, HROs documents

include a particular state of affairs amounts to a violation of international human rights standards AND a particular violator who is responsible for the violation. Second, almost all the HRO documents are issued and classified by country. Figure 4.1. illustrates the data distribution in terms of average count per year for total HRO naming and shaming between 1991 and 2011.



4.3.2 Independent Variables

4.3.2.1 HROs Information Collection: Approach

The key is how each HRO collects the information on human rights violations. In order to measure HROs' approach, first, I focus on organizational level analysis. In other words, it is based on each HRO's mission and value. In general, each HRO explains its operational methods under the section of something like "How We Work." Also, I read through each HRO's document and identify how they work in practice. Let me take an example of direct information gathering. *Dui Hua foundation* explains its working methods as follows:

"Dui Hua engages governments like China in mutually respectful dialogue with the aims of securing the release and better treatment of individuals incarcerated for non-violent expression of political and religious beliefs and promoting transparency and rule of law... Prisoner lists are tools pioneered by Dui Hua for submitting cases of concerns to governments. These lists are compiled through independent and dialogue with governments" (Duihua 2012).

Human Rights Without Frontiers International (HRWF) is an example of HROs with indirect information gathering:

"Our strength is in having access to diverse sources of information being part of wider advocacy networks with flow of information... With trustworthy information of situations of human rights violations, we seek to

bring the human rights record of targeted countries to the light of international scrutiny using moral leverage” (HRWF 2013).

In particular, I take a somewhat conservative approach. In other words, just visiting a country is not sufficient. HROs have to directly contact with governments to get information. For example, do they visit local police stations? Do they actually visit detention facilities to talk with officials in order to handle the issues of human rights violations? This direct contact is the key. Simply collecting secondhand information such as interviewing refugee in a third party state is not sufficient to be considered as a direct information gathering. A direct information gathering is coded as “0” and an indirect gathering is coded as “1.”

4.3.2.2 Human Rights Conditions

In order to measure human rights conditions for targeted states, I utilize the Cingranelli-Richards (CIRI) Physical Integrity Rights Index (Cingranelli and Richards 2014). The Physical Integrity Rights measure is a compilation of scores relating to “torture, extrajudicial killing, political imprisonment, and disappearance” (Cingranelli and Richards 2014). These rights match well with the purpose of this project. The scores of this index range from 0 (No respect for human rights) to 8 (full respect for these rights).

4.3.2.3 State Power

I expect HROs to report more heavily on violations within powerful states because they have greater potential effects on international norms (Goldstein et al. 2000,

Nadelmann 1990, Ramos, Ron and Thoms 2007, Ron, Ramos and Rodgers 2005). If an HRO build support for a new international convention; garner more attention; or boost its fund-raising potential, it makes sense to focus on abuses by high-profile, powerful countries such as the U.S., rather on violations occurring in Zimbabwe.

Wealth is a major determinant of power and prominence, and thus should increase HROs' shaming. Wealth may also have an opposite effect, however, given its association with improved rights conditions and reduced chances of civil war (Poe and Tate 1994, Ron, Ramos and Rodgers 2005). Given these conflicting explanations, I expect only that wealth will have statistically significant impact on HROs' shaming. I measure wealth by a state's GDP per capita, obtained from the World Bank's World Development Indicators (World Bank 2014). My second state power measure is the size of a country's military, which I expect to increase HROs' output. Militarily powerful countries have greater geopolitical prominence, attracting greater HRO attention, and may also be associated with more human rights abuse, given the armed forces' key role in repression and human rights violations. I use the data on the size of military expenditures in a given year (World Bank 2014).

4.3.2.4 Population

I also anticipate that population will increase HROs shaming. Heavily populated countries command greater international attention, and abusive governments in populous states are likely to have greater overall effects on human welfare. In addition, scholars note an association between repression and population size, hypothesizing

that this stems from resource strains (Poe and Tate 1994, Ron, Ramos and Rodgers 2005). In other words, a large population increases the number of occasions on which such coercive acts can occur. As a matter of simple probability, such an increase should lead to the occurrence of more instances of coercion. Also, a large population places stress on national resources and bring the threat of environmental deterioration, further reducing available resources.

4.3.2.5 Civil Society

My measure of domestic civil society is the number of human rights NGOs based in a country in a given year and registered with the Union of International Associations (UIA). Smith and Wiest (2004)'s data have been widely used for a long time, but it is not necessarily accurate. They only counted 4 year terms and the rest of them are imputed. In addition, they included all the NGOs, not necessarily human rights NGOs. On the basis of NER, I identified all the relevant domestic HROs in each country between 1991 and 2011 from *the Yearbook of International Organization*.¹

4.3.2.6 Democracy

My expectations for this indicator is mixed. On the one hand, increased political openness should reduce the volume of HROs reporting, as political participation is itself a basic human rights, and scholars associate it with reduced human rights violations. Political openness might also have the opposite effect, however. Politically

¹Since the Yearbook provides the list of the country names in each organization's membership section, 1) I scrapped the section for each HROs, 2) I extracted country names with NER, 3) I converted them into COW country codes.

open countries have more protests, journalists, and NGOs, all of which stimulate political and moral debate. Also, intermediate levels of democracy and the process of democratization itself are associated with greater risk of violent conflict, moreover, and this could trigger greater levels of abuses (Hegre 2001, Ron, Ramos and Rodgers 2005, Snyder 2000). I thus anticipate only that Polity IV's effect will be significant, but remain agnostic as to its direction.

4.3.2.7 International Human Rights Legal Commitment

Many IR scholars have studied why states ratify international human rights treaties. There may be a number of reasons for governments to ratify human rights treaties. In particular, considering that states (particularly insincere one) have incentives to engage in opportunistic ratification of human rights treaties, scholars have argued that there are two kinds of benefits that states seek for through ratifying the treaties: tangible economic benefits and intangible benefits. For example, Hathaway (2004) argues that states join treaties like the Convention against Torture (CAT) in order to make them look good and attract more foreign aid donations, and international trade, etc. Some governments may ratify human rights treaties as a condition for some intergovernmental organization membership. Governments may ratify them because they expect that it will convince foreign investors of the strength of domestic rule of law or the government's long time horizons (Farber 2000, Simmons 2009). In the same vein, Goodliffe and Hawkins (2006, 361) also argue that by ratifying human rights treaties like CAT, states expect material rewards such as investment,

trade, aid, and positive political relationships. States also expect intangible benefits from joining human rights treaties. Legitimacy-conscious governments care about what other states think of them in the international community. Therefore, with the ratification of human rights treaties, governments express the desire to belong to the international community (Lutz and Sikkink 2003, 659). It provides a good press or an improved image with audiences both at home and abroad. In this context, Hafner-Burton and Tsutsui (2005) argue that participation in human rights regimes give governments a degree of legitimacy. Also, by ratifying treaties, governments can avoid criticism often from external audiences. Thus, governments ratify a human rights treaty even if it does not expect to comply. That is, insincere ratification is common. Given tangible and intangible benefits, governments take the “gamble of ratification” (Simmons 2009). It is the gamble because these benefits can be only short term when insincere ratification is revealed. Therefore, HROs’ shaming is more likely to have impact on governments that ratified international human rights treaties. It gives HROs a kind of justification or legal ground to blame those governments. In order to measure international legal commitment, I look at state’s ratification records of six major human rights conventions: International Covenant on Civil and Political Rights (ICCPR), Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment (CAT), the Convention on the Prevention and Punishment of the Crime of Genocide, Convention on the Elimination of Racial Discrimination (CERD), Convention on the Elimination of Discrimination against Women (CEDAW), and Convention on Enforced

Disappearances (CED). Instead of looking at them respectively, I am going to use state's ratification records with a 7 point scale with "6" indicating ratification of all six conventions and "0" none—that is, it measures the level of international legal commitment.

4.3.2.8 HROs and Target Distance

Several studies in conflict literature have found an association between interstate conflicts and geographical proximity (Diehl 1991, Huth 1996, Kocs 1995, Starr and Most 1976). In other words, boxers, after all, cannot fight until they are physically able to reach one another. Thus, war is more likely between states that share a common border zone. This argument builds on the notion that proximity engenders serious conflicts of interests states, a fraction of which are bound to lead to war. Shared access to a physical area can lead directly to interstate friction, even if the states involved agree as to where the border lies between them. Gochman (1991) reported that about two-thirds of militarized interstate disputes occurring between 1816 and 1976 were between states that shared a common land border or were separated by 150 miles or less of water. Thus, proximity is salient. One of the primary explanations for why contiguous states go to war is that they are fighting over territory. Contiguous states are better able to reach one another militarily and thus are better able to engage one another (Vasquez 1995). They interact more with one another and thus potentially have more issues over which to conflict. Furthermore, the ability of a major power to reach another state declines dramatically with the

increase in geographical distance between these two states.

In this context, some scholars have argued that the origins of human rights organizations influence their activities (Stroup and Murdie 2012). Both formal and informal institutions that concern how INGOs operate in their home country influence their operations with respect to advocacy behaviors overall. Thus, I expect that HROs are more likely to pay attention to states that are close to their home country. In other words, U.S.-based HROs are more likely to pay attention to the issues in the U.S. or their neighboring countries. Egypt based HROs are more probable to cover the problems within their neighboring states.

In order to measure this physical distance between HROs' home country (i.e. headquarter) and target states, I utilize the inter-capital distance between HROs' headquarter and target countries, *the Correlates of War Direct Contiguity Data*. Table 4.1 presents the summary statistics for the data.

4.4 Estimation Method

In this analysis, the dependent variable (shaming event) is a count variable—that is, the number of being shamed by HROs in a given year. It is often suggested that Poisson regression as a possible tool for the analysis of count data. Poisson regression is a special type of non-linear regression that considers the non-negativity and discreteness of the data. The underlying Poisson distribution, however, has several implications that are often neglected. It assumes that the mean and the

Variables	Mean	Std. Dev.	Min	Max
Shaming Counts	0.54	3.04	0	147
Human Rights (CIRI)	4.93	2.28	0	8
Civil Society	3.54	0.98	0	5.97
Approach	—	—	0	1
Democracy	2.78	6.72	-10	10
GDP pc (ln)	7.87	1.63	4.17	12.17
Population (ln)	15.41	2.15	9.11	21.01
Media Reports (ln)	5.75	2.29	0	12.37
Military (ln)	0.63	0.75	-3.06	4.75
HR Treaty	4.28	2.44	0	6
Distance (ln)	8.20	0.79	3.66	9.42

Table 4.1: Descriptive Statistics Summary (Count Model)

variance are equal. This equidispersion assumption may be too restrictive. When there is overdispersion the Poisson estimates are inefficient with standard errors biased downwards, and the maximum likelihood Poisson z-statistics are overinflated.

In order to solve this issue, one could use Poisson Quasi-Maximum Likelihood estimator with corrected standard errors or negative binomial estimator (Hilbe 2011).

In this chapter, for all reported results, I use negative binomial models with standard errors adjusted for clustering on dyads to account for nonindependence by pairs. I do not include year-specific time dummies, but my results are fully robust toward their inclusion. In robustness tests, I also estimate so-called zero-inflated negative binomial regressions.²

²The first stage of zero-inflated models (the inflation equation) uses a binary logit model to estimate whether there is a positive probability of an event within a dyad. The second stage (the count equation) account for variation in the number of events.

4.4.1 Negative Binomial Model

The most widely used model specification for a variable that is a count of events is the Poisson regression,

$$Pr(Y_t = y_i) = f(y_i) = \frac{e^{-\lambda_i h} \lambda_i^{y_i}}{y_i!} \quad (4.1)$$

there is a constant rate (λ) at which events occur—this rate is the expected number of events in period of length h . In this Poisson process, events occur independently with a constant probability equal to λ times the length of the interval i.e. λh . Typically, we assume that all intervals are of the same length equal to one, and so we have

$$Pr(Y_t = y_i) = f(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (4.2)$$

The Poisson model requires that $E[Y_i] = Var(Y_i) = \lambda_i$. Let's imagine that we relax this mean-variance equality restriction and say that $Var(Y) = \lambda_i \sigma$. Then, in general, we can think of the situation like this: Overdispersion: $E[Y] < Var(Y) \leftrightarrow \sigma > 1$ [Overdispersion in Poisson models occurs when the response variance is greater than the mean. It is caused by positive correlation between responses or by an excess variation between response counts. Overdispersion also arises when there are violations in the distributional assumptions of the data. It is problematic because it may cause standard errors of the estimates to be deflated or underestimated, i.e. a variable may appear to be a significant predictor when it is in fact not significant.]

We use a negative binomial model to deal with overdispersion. In the NB model, we still have the mean as $E[Y_i] = \lambda_i = e^{X_i\beta}$. However, we now assume that the variance is

$$\sigma^2 = \lambda_i(1 + \alpha\lambda_i) \quad (4.3)$$

In the Poisson model, we stated that λ_i was fully determined by a linear combination of the X s. The negative binomial essentially adds some unobserved heterogeneity, such as that λ_i is determined by the X s and some unobserved, observation-specific random effect ϵ_i , which is assumed uncorrelated with the X s. So we now have $\tilde{\lambda}_i$ such that

$$\begin{aligned} \tilde{\lambda}_i &= e^{x_i\beta + \epsilon_i} \\ &= e^{x_i\beta} e^{\epsilon_i} \\ &= \lambda_i \delta_i \end{aligned} \quad (4.4)$$

where $\delta_i = e^{\epsilon_i}$. The negative binomial is not identified without an assumption about the mean of the error term and the most convenient assumption is that $E[\delta_i] = 1$ since this gives us $E[\lambda_i] = \lambda_i$. This is good because it means that we have the same expected count as from the Poisson model.

Since the distribution of observations is still Poisson given X and δ , we have

$$Pr(y_i|x_i, \delta_i) = \frac{e^{-\tilde{\lambda}_i} \tilde{\lambda}_i^{y_i}}{y_i!} = \frac{e^{-\lambda_i \delta_i} (\lambda_i \delta_i)^{y_i}}{y_i!} \quad (4.5)$$

Since δ_i is unknown, we cannot compute $Pr(y_i|x_i, \delta_i)$ and instead need to compute the distribution of y given only X . To compute $Pr(y_i|X_i)$ without conditioning on δ , we average $Pr(y_i|X_i)$ by the probability of each value of δ . So, if g is the pdf of δ then

$$Pr(y_i|x_i) = \int_0^\infty [Pr(y_i|x_i, \delta_i) \times g(\delta_i)] d\delta_i \quad (4.6)$$

In order to be able to solve this equation, we need to specify the form of the pdf for δ . Most people assume that δ_i has a gamma distribution with parameter ν_i .

$$g(\delta_i) = \frac{\nu_i^{\nu_i}}{\Gamma(\nu_i)} \delta_i^{\nu_i-1} e^{-\delta_i \nu_i} \quad \text{for } \nu_i > 0 \quad (4.7)$$

where $\Gamma(\nu_i) = \int_0^\infty t^{\nu-1} e^{-t} dt$. The nice thing about this distribution is that $E[\delta_i] = 1$ which was the convenient assumption we wanted to make since it means that the mean function of the negative binomial is the same as the Poisson. $Var(\delta_i) = \frac{1}{\nu_i}$. Using (4.5) and (4.7) to solve (4.6) we obtain the following negative binomial probability distribution:

$$Pr(y_i|x_i, \nu_i) = \frac{\Gamma(y_i + \nu_i)}{y_i! \Gamma(\nu_i)} \left(\frac{\nu_i}{\nu_i + \lambda_i} \right)^{\nu_i} \left(\frac{\lambda_i}{\nu_i + \lambda_i} \right)^{y_i} \quad (4.8)$$

The expected value of this is the same as for the Poisson distribution i.e., $E[y_i|X_i] =$

λ_i . However, the conditional variance now differs:

$$Var(y_i|x_i) = \lambda_i \left(1 + \frac{\lambda_i}{\nu_i}\right) = e^{x_i\beta} \left(1 + \frac{e^{x_i\beta}}{\nu_i}\right) \quad (4.9)$$

Since λ and ν are both positive, the variance will exceed the conditional mean. This increases the relative frequency of low and high counts. So far, so good. But the variance remains unidentified since if ν_i is the same for all observations i.e.,

$$\nu_i = \alpha^{-1} \quad for \quad \alpha > 0 \quad (4.10)$$

This particular formulation just helps simplify some equations. If we substitute (1.10) into (1.9), we have:

$$Var(y_i|x_i) = \lambda_i \left(1 + \frac{\lambda_i}{\alpha^{-1}}\right) = \lambda_i(1 + \alpha\lambda_i) = \lambda_i + \alpha\lambda_i^2 \quad for \quad \alpha > 0 \quad (4.11)$$

The bottom line is that the NB model has:

$$\begin{aligned} E[Y] &= \lambda_i = e^{X_i\beta} \\ Var(Y) &= \lambda_i(1 + \alpha\lambda_i), \alpha > 0 \end{aligned}$$

Thus, the variance is always greater than the mean, but the variance is still (positively) dependent on the mean, i.e., heteroskedastic. Larger values of α correspond

to greater amounts of overdispersion. The density of the NB distribution is:

$$Pr(y_i|x_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} \quad (4.12)$$

Now, we can estimate the NB model by using maximum likelihood:

$$\begin{aligned} \mathcal{L} = \prod_{i=1}^N Pr(y_i|x_i) &= \prod_{i=1}^N \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\nu_i + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} \\ &= \prod_{i=1}^N \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\nu_i + e^{x_i \beta}} \right)^{\alpha^{-1}} \left(\frac{e^{x_i \beta}}{\alpha^{-1} + e^{x_i \beta}} \right)^{y_i} \end{aligned} \quad (4.13)$$

$$\ln \mathcal{L} = \sum_{i=1}^N \ln Pr(y_i|x_i) = \sum_{i=1}^N \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\nu_i + e^{x_i \beta}} \right)^{\alpha^{-1}} \left(\frac{e^{x_i \beta}}{\alpha^{-1} + e^{x_i \beta}} \right)^{y_i} \quad (4.14)$$

After taking the log of the likelihood function, we would maximize it with respect to β and α .

4.4.2 Results

Before presenting my regression findings, I present a couple of bivariate plots between the dependent variable and the key independent variables (the level of civil society and the human rights conditions). Figure 4.2 illustrates the bivariate relationship between the level of civil society and the average total HROs naming and shaming counts (logged).

By simply looking at this bivariate plot, there is a positive relationship between

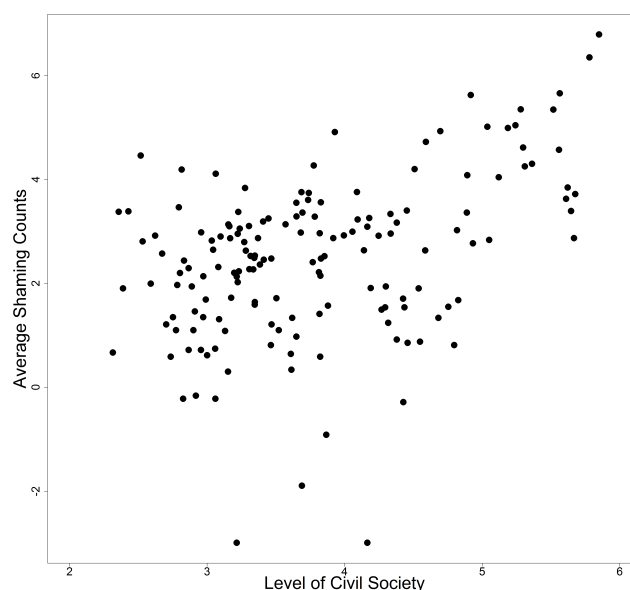


Figure 4.2: HROs Shaming Counts and the Level of Civil Society

the level of civil society and the shaming counts.³

Figure 4.3 illustrates the relationship between the level of human rights conditions (CIRI) and the average total HROs naming and shaming counts (logged). Although it would be too early to say that there is no clearly recognizable pattern, it is not as clear as the pattern above.

Turning to my regression finding, Table 4.2 presents the estimation results. I report the results from a baseline model (model 1) without interaction effects as well as interactive models that assess the effects of information gathering approach conditional on the strength of civil society (model 2), and human rights conditions (model 3). Before I come to my variables of main interest and thus to my hypotheses, I briefly discuss results on the control variables. More populous countries are more likely to get targeted, all other things equal as expected. I found a highly

³It should be noted that the average count is the average count per year for all the HROs in the sample.

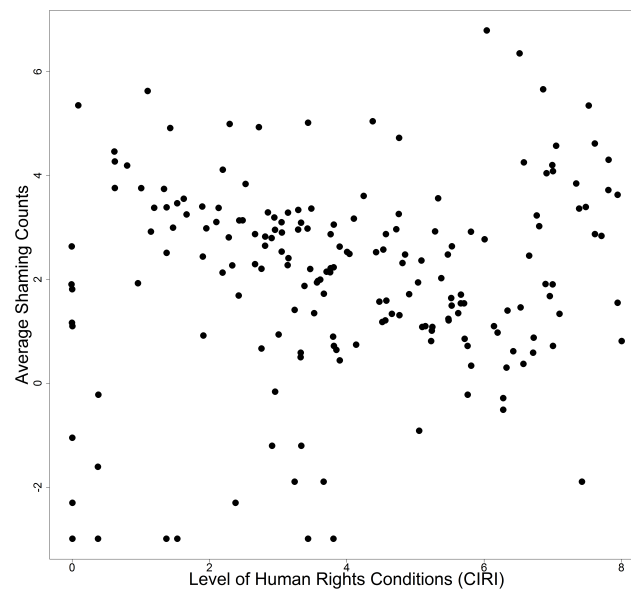


Figure 4.3: HROs Shaming Counts and the Level of Human Rights Conditions

significantly negative effect of distance between HROs' main headquarter and targeted states. The number of overall media coverage also matters. HROs are more likely to target states that were exposed to news coverage. As expected and in line with other studies, the number of HROs' shaming increases with states' increased military capabilities. My findings provide little support for an unconditional effect of democracy and international human rights treaty ratification. This can be seen from the statistically insignificant coefficients for these variables in the models.

Turning to my variables of main interests, at first glance, the coefficient for human rights conditions is not statistically significant, suggesting that human rights condition itself is not a significant determinant in naming and shaming. As expected, the coefficient for civil society is positive and significant, suggesting that, for states, higher civil society are positively associated with higher counts of HRO's shaming. Also, the coefficient for HRO's approach is positive and significant, suggesting that

HROs with indirect approach are more likely to engage in shaming activities than HROs with direct approach. The results from the non-interactive model clearly illustrates the sizable effect of my variables of main interests on the dependent variable. Nonetheless, the more relevant results for my argument are the interactive specifications, which model accurately the theory by testing the conditional relationship between HROs' approach and civil society. The correct interpretation of the interaction results reported in Table [number here] is rendered very difficult by the non-linearity of the negative binomial regression model. In a linear model, the hypothesis that the effect of one independent variable on the dependent variable depends on another explanatory variable can be tested easily by a t-test of the estimated coefficient of the interaction term (Kam 2007, 50). However, it is noteworthy that this easy way of testing the statistical significance of an interaction effect no longer applies when the model is non-linear. As Ai and Norton (2003, 129) have shown, in nonlinear models "the interaction effect cannot be evaluated by looking at the sign, magnitude, or statistical significance of the coefficient of the interaction term." Instead, the interpretation of the interaction term requires computing the cross derivative of the expected value of the dependent variable, which depends on all the covariates in the model and their values.⁴

In order to evaluate the substantive effect of my main variables of interest, Table 4 reports the predicted joint effects of information approach and civil society, and human rights conditions, based on estimation results from Table 3. The reported

⁴See Appendix testing the significance of interaction in non-linear models)

Table 4.2: Determinants of Shaming Counts

Variables	Model (1)	Model (2)	Model (3)
Human Rights	-0.012 (0.012)	-0.013 (0.012)	-0.047** (0.016)
Democracy	-0.005 (0.005)	-0.005 (0.004)	-0.006 (0.006)
Civil Society	0.581** (0.055)	-0.38** (0.056)	0.128** (0.005)
Media	0.206** (0.020)	0.165** (0.017)	0.164** (0.017)
GDP	-0.005 (0.022)	-0.031 (0.022)	-0.028 (0.022)
Population	0.270** (0.024)	0.352** (0.022)	0.356** (0.023)
Military	0.182** (0.040)	0.183** (0.037)	0.187** (0.038)
HR Treaties	0.013 (0.017)	0.014 (0.015)	0.009 (0.015)
Distance	-0.293** (0.030)	-0.332** (0.027)	-0.320** (0.026)
Info.Approach	0.987** (0.058)	-1.76** (0.230)	0.221 (0.120)
Info.Approach*Civil Society		0.690** (0.057)	
Info.Approach*Human Rights			0.141** (0.019)
Constant	-7.453** (0.433)	-4.779** (0.453)	-6.411** (0.425)
ln alpha	1.192** (0.037)	1.055** (0.039)	1.096** (0.385)
N obs.	127,303	127,303	127,303
Wald chi-sq	1927.02**	1696.83**	1482.26**
-ll	-101431.78	-111017.03	-111632.5

Levels of Confidence: * $p < .1$, ** $p < .05$, *** $p < .01$. Standard errors clustered on dyads in parentheses.

values are percentage changes in the predicted count of HRO's shaming relative to a chosen reference category in which HROs take direct approach and the level of civil society is 0 (all control variables at mean values).

Table 4.3: Expected Effect of Approach and Civil Society

Level of Civil Society (percentile)	Direct Approach	Indirect Approach
25 th	-0.05	0.20
50 th	-0.24	0.44
75 th	-0.41	1.29
100 th	-0.70	2.12

Note: Percentage change in predicted count of HRO shaming relative to base category (direct approach, civil society level (0), all other variables at mean values. Estimates based on Model (2)).

Looking at the predicted joint effects, it becomes apparent that the indirect information gathering approach increases the predicted count and the more so the higher the level of civil society. In states where there is stronger civil society (75th), the indirect approach increases the expected count of shaming by 129 percent relative to our base category.

4.4.3 Alternative Statistical Model

Like most studies using statistical analysis, I examined many more models than I am able to conveniently present in the text. In this section, I discuss an alternative statistical model. As discussed in the previous section, my dependent variable of interest is the event of HRO shaming. In a count model, a count refers to the number of specified events that occur in a given interval of time. By definition,

count data consist of only non-negative integers. However, a number of states have not experienced this event and it creates a preponderance of zeros. Such distribution violates fundamental assumptions of OLS regression, most notably normality of residuals. It leads to inaccurate standard errors (Atkins and Gallop 2007).

As a solution, in the previous section, I utilized the negative binomial model. In this section, I re-estimate with Zero-Inflated Negative-Binomial (ZINB). There is no well-defined cutoff that can be used to differentiate Negative Binomial and ZINB (Sano et al. 2005). It is often recommended to perform it as post hoc analyses to investigate the appropriateness of the model.

By specifying the sources of zeros, ZINB allows us to answer to questions, (1) what predicts whether or not the event occurs, and (2) if the event occurs, what predicts frequency of occurrence. In other words, two regression equations are created: one predicting whether the count occurs and a second one predicting differences on the occurrence of the count (Long and Freese 2006). One could consider using Hurdle models (also known as Zero-Altered Models) (Cameron and Trivedi 2013). However, there is the fundamental difference from the hurdle models. The hurdle models do not discriminate the sources of zeros (Gail et al. 2007, 273). In other words, the count process can produce two zero groups: (1) states belong to “the always-zero” group” (i.e. never shamed) and (2) states belongs to “the not-always-zero” group (i.e. could be shamed). I believe that it is more appropriate to model this difference, because data generating process of false zeros (the always-zero) are related to HROs’ errors and capability observe human rights abuses..

Let the dependent variable Y_i denote a non-negative interger count for the i th observation ($i = 1, 2, \dots, N$). We assume that Y_i is a false zero and binomially distributed with probability p , ($0 \leq p \leq 1$.) and the probability that Y_i is not a false zero equal to $1 - p$. On the basis of this assumption, we can write the following equations:

$$\begin{aligned} Pr(Y_i = 0) &= Pr(False\ zeros) + (1 - Pr(False\ zeros)) \\ &\quad \times Pr(Count\ process\ gives\ a\ zeros) \end{aligned} \quad (4.15)$$

$$Pr(Y_i = 0) = p + (1 - p) \times Pr(Count\ process\ gives\ zeros) \quad (4.16)$$

In other words, Y_i follows a Zero-Inflated Negative Binomial (ZINB) distribution,

$$P(Y_i = y_i) : p + (1 - p)\left(1 + \frac{\lambda}{\tau}\right)^{-\tau}, y_i = 0 \quad (4.17)$$

$$P(Y_i = y_i) : (1 - p) \frac{\Gamma(y + \tau)}{y_i! \Gamma(\tau)} \left(1 + \frac{\lambda}{\tau}\right)^{-\tau} \left(1 + \frac{\tau}{\lambda}\right)^{-y_i}, y_i = 1, 2, 3, \dots \quad (4.18)$$

where τ is an overdispersion parameter, and λ is the Poisson mean based on gamma distribution (Γ). Then, the mean and variance of the ZINB distribution are

$$E(Y_i) = (1 - p)\lambda \quad (4.19)$$

$$Var(Y_i) = (1 - p)\lambda(1 + p\lambda + \frac{\lambda}{\tau}) \quad (4.20)$$

Observe that this distribution approaches the zero-inflated Poisson distribution and the negative binomial distribution as $\tau \rightarrow \infty$ and $p \rightarrow 0$, respectively. (Note that if both $\frac{1}{\tau}$ and $p \approx 0$ then the ZINB distribution reduces to the Poisson distribution).

The ZINB distribution model relates p and λ to covariates, that is,

$$\log(\lambda_i) = X_i\beta \quad (4.21)$$

$$\text{logit}(p_i) = Z_i\gamma, \quad (4.22)$$

$$(i = 1, 2, 3...n)$$

where $X_i = (x_1, x_2, \dots, x_n)$ and $Z_i = (z_1, z_2, \dots, z_n)$ are vectors of covariates pertaining to each subject i , and with β and γ the corresponding vectors of regression coefficients, respectively.

Table 4.4 presents the results. There are two sets of coefficients for ZINB relating to the logistic and counts proportions of the model. Overall, ZINB estimations are pretty much consistent with NB estimations and support the hypotheses. The only difference is the inflation stage of the models, because NB does not specifically differentiate the differences in zeros. For the inflation stage of the model, an individual state's chance for being in the false zero ("never shamed"), decrease by 21.3%, 21.0 %, and 20.9% respectively for every one unit increase in the level of democracy

across three models. In other words, better democracy is less likely to be ignored by HROs.

Table 4.4: Determinants of Shaming Counts (Zero-Inflated NB Estimation)

Variables	Model (1)	Model (2)	Model (3)
Count Stage			
Human Rights	-0.011 (0.012)	-0.015 (0.012)	-0.125** (0.016)
Democracy	-0.001 (0.005)	-0.001 (0.006)	-0.001 (0.006)
Civil Society	0.325** (0.045)	-0.195** (0.057)	0.323** (0.045)
Media	0.188** (0.017)	0.188** (0.017)	0.188** (0.017)
GDP	-0.038* (0.021)	-0.031 (0.021)	-0.034 (0.021)
Population	0.270** (0.019)	0.321** (0.021)	0.324** (0.021)
Military	0.130** (0.040)	0.129** (0.039)	0.133** (0.040)
HR Treaties	0.019 (0.015)	0.021 (0.016)	0.018 (0.015)
Distance	-0.260** (0.026)	-0.285** (0.026)	-0.027** (0.026)
Info.Approach	0.997** (0.052)	-1.717** (0.230)	0.189 (0.125)
Info.Approach*Civil Society		0.687** (0.057)	
Info.Approach*Human Rights			0.152** (0.020)
Constant	-7.457** (0.401)	-5.021** (0.435)	-6.646** (0.401)
Inflation Stage			
GDP	-0.030 (0.059)	-0.037 (0.060)	-0.030 (0.060)
Democracy	-1.546** (0.183)	-1.559** (0.184)	-1.564** (0.184)
Constant	-10.527** (1.535)	-10.613** (1.547)	-10.727** (1.550)
N	127303	127303	127303
Log-likelihood	-108851.785	-108248.569	-108248.569

Levels of Confidence: * $p < .1$, ** $p < .05$, *** $p < .01$. Zero-inflated negative binomial model with yearly fixed effects (omitted) and robust standard errors in parentheses.

4.5 Conclusion

This chapter explains how states obtain a position on the international human rights agenda through the expansion of human rights organizations. The main argument is focused on explaining the conditions under which naming and shaming, which target and seek to remedy human rights abuses committed by a state, are generated. On the basis of a larger sample of HROs, I have provided a more comprehensive model that extends and amplifies upon the previous researches. Where previous research on HROs tended to focus on a single organization or news media sources, I have developed a theory of naming and shaming that are shaped by its information collection strategies. Drawing upon this theory, these analyses show how HROs collect information is a critical determinant of HROs' naming and shaming. Moreover, it reveals that when HROs are engaged in an indirect information gathering approach, the strength of civil society in a given state increases the chances of being targeted.

Thus, the findings are mixed. Close examinations of HROs reveal a more complex picture than the conventional wisdom. International activists have made considerable achievements in promoting human rights norms and enforcing them, but the findings suggest that HROs also operate with limited resources against enormous odds, forcing them to pursue more "pragmatic" and politically strategic behaviors. To be effective, HROs engage in "information politics", reporting more heavily on human rights abuses in some countries than others and devoting more attention to some areas than others. Actual human rights conditions alone do not have unconditional effects on the volume of HROs' naming and shaming.

This chapter clarifies the challenge for HROs. They have to ensure that strategic considerations do not play too large a role. Although there is little doubt that information politics is useful for HROs, politically and materially savvy strategies may ultimately contribute to the marginalization of abuses in smaller, poorer, or weaker countries.

Chapter 5

Naming and Shaming Intensity Models

5.1 Introduction

Chapter 2 (Theory Chapter) outlines a set of conditions under which states are more likely to get shamed by HROs and once targeted, how get shamed. The theoretical framework laid out in Chapter 2 suggests that empirically, yet the existing literature has not highlighted large variation in the targeting of particular groups of states and the ways in which those targeted states are shamed. In particular, I argued that for HROs that use direct information gathering, cooperation or collaboration with governments is necessary given the political context. In order to secure their access to these governments, HROs are limited in the extent to which they can challenge the sovereign state's violations. They are willing to work within the space they are

given to operate. Thus, HROs with direct information gathering are less likely to be harsh on states.

This chapter tests this hypothesis through a battery of statistical analyses. This chapter proceeds in four parts. Section II discusses the unique data on HROs' shaming. Section III estimates a set of regression models to test hypotheses. Section V concludes.

5.2 Human Rights Shaming: Intensity

The theoretical framework laid out in Chapter 2 suggests that a gathering information about human rights practices is difficult. States' human rights practice takes place domestically; thus, despite the widespread development of the accountability mechanisms, many human rights violations truly are difficult to detect, to observe, and to verify. Given the secretive nature of human rights abuses, it is important for HROs to gather information that can be effectively put together to influence the implementation of human rights by governments. In order to inspire corrective efforts by governments, HROs must demonstrate that their factual statements are true and thus constitute a reliable basis for remedial governmental policy. Simply "knowing" human rights violations are occurring is not enough. Otherwise, HROs will lose adherents and influence. HROs make efforts to find information about human rights practices. Their credibility and legitimacy depend in large measure on the reliability of the facts obtained through such efforts. And the reliability of their information depends, in turn, on the methods used to obtain such information.

In chapter 2, I suggested that there are two broad approaches to information gathering: an indirect approach and a direct approach.

One of the important aspects of HROs is their strategic motivation. In other words, although existing studies regard transnational actors as neutral and trustworthy (Clark 2001), many scholars have noted explicitly or implicitly that advocacy organizations resemble interest groups and they are normative and strategic. HROs, just like interest groups, have to not only consider using their resources effectively and efficiently, but also care about their own organizational survival. In order to continue their campaign on human rights issues, HROs have to exist first and they must be concerned about organizational survival. While their motivations may be noble and moral, they behave in much the same way as other actors. They have to make themselves attractive to potential members, donors, and supporters. They have to make themselves look effective in changing or improving governments' human rights practices. When there are multiple HROs they have to compete between each other for donations. Thus, given the available information, HROs have an incentive to disseminate "high quality" information about abusers. Disseminating inaccurate information can undermine the reputation and legitimacy of the HRO, resulting in less support by donors. At the same time, HROs do not want to hold onto information that they have, lest they be scooped by another HRO or be accused of concealing abuses or not effective organization. In other words, when there are a number of other HROs working on the same issue, any HROs are expected to talk about the same issue. They cannot afford to hold onto any information that they

have.

As I have discussed in the theory chapter, direct information gathering is one of the most important tools for an effective investigation of human rights violations such as torture, ill-treatment, and conditions of detention. Through direct observation, HROs can develop a fairly good understanding of the specific root causes of human rights abuses in the respective societies as well as of the measures needed to combat those violations and to improve the conditions. On the basis of this assessment, HROs publicize their findings and specific recommendations for further actions. While the process of direct information gathering might seem to be neutral and technical, it is neither. Instead, it is an inherently political endeavor. Even with the most exclusive inside information from visited states, HROs cannot always publicly criticize oppressive states that they observe. Sometimes, HROs must obscure as much as they reveal. A number of HROs that utilize a direct approach emphasize that harsh public criticism after a country visit may impede further access, so they have to weigh the merits and pitfalls of assigning government responsibility for abuses before moving toward public condemnation.

I argue that for HROs with direct information gathering, cooperation or collaboration with governments is necessary given the political context. HROs, then, are “flexible,” and they are willing to negotiate with the governments. Since they do not need to rely on domestic opposition groups for information, and they are willing to directly communicate with the governments, I predict that HROs that rely on direct information gathering tend to focus on the worst human rights violators. However,

in order to secure their access to those governments, HROs are limited in the extent to which they can challenge the sovereign states violations. They are willing to work within the space where they are allowed to operate.

Stated in hypothesis form, my argument thus can be summarized as follows:

H1: *HROs with indirect information gathering are more likely to be harsh on states.*

H2: *HROs with direct information gathering are less likely to be harsh on states.*

5.2.1 Dependent Variable

My dependent variable is the level of shaming intensity. As discussed in Chapter 3, I measure the intensity of shaming on the basis of text classification machine learning. The role of text classification is to assign a document into one of a number of predefined categories. In supervised machine learning approaches, I use the computer program automatically to learn the classification function from a number of annotated example documents. In particular, I deal with binary classification problems where the aim of the classifier is to detect whether a piece of text from HROs belong to one of the two mutually exclusive categories. I apply this setting to a classification problem. The classifier aims to detect whether a preassigned subjective document contains positive or negative opinion. Then, I use this as a proxy for the intensity of shaming on the basis of posterior probability. [See Chapter 3 for measuring shaming intensity.]

5.2.2 Independent Variables

5.2.2.1 HROs Information Collection: Approach

The key is how each HRO collects the information on human rights violations. As I have discussed in the previous chapter, a direct information gathering is coded as “0” and an indirect gathering is coded as “1.”

5.2.2.2 Human Rights Conditions

In order to measure human rights conditions for targeted states, I utilize the Cingranelli-Richards (CIRI) Physical Integrity Rights Index (Cingranelli and Richards 2014). The Physical Integrity Rights measure is a compilation of scores relating to “torture, extrajudicial killing, political imprisonment, and disappearance” (Cingranelli and Richards 2014). The scores of this index range from 0 (No respect for human rights) to 8 (full respect for these rights).

5.2.2.3 State Power

As in the previous chapter, I measure wealth by a state’s GDP per capita, obtained from the World Bank’s World Development Indicators (World Bank 2013). And I use the data on the size of military expenditures in a given year (World Bank 2013) to measure the size of a country’s military.

5.2.2.4 Population

I also anticipate that population will increase HROs shaming. Heavily populated countries command greater international attention, and abusive governments in populous states are likely to have greater overall effects on human welfare. In addition, scholars note an association between repression and population size, hypothesizing that this stems from resource strains (Poe and Tate 1994, Ron, Ramos and Rodgers 2005). In other words, a large population increases the number of occasions on which such coercive acts can occur. As a matter of simple probability, such an increase should lead to the occurrence of more instances of coercion. Also, a large population places stress on national resources and bring the threat of environmental deterioration, further reducing available resources.

5.2.2.5 Civil Society

My measure of domestic civil society is the number of human rights NGOs based in a country in a given year and registered with the Union of International Associations (UIA). Smith and Wiest (2004)'s data have been widely used for a long time, but it is not necessarily accurate. They only counted 4 year terms and the rest of them are imputed. In addition, they included all the NGOs, not necessarily human rights NGOs. On the basis of NER, I identified all the relevant domestic HROs in each country between 1991 and 2011 from *the Yearbook of International Organization*.¹

¹Since the Yearbook provides the list of the country names in each organization's membership section, 1) I scrapped the section for each HROs, 2) I extracted country names with NER, 3) I converted them into COW country codes.

5.2.2.6 Democracy

My expectations for this indicator is mixed. On the one hand, increased political openness should reduce the volume of HROs reporting, as political participation is itself a basic human rights, and scholars associate it with reduced human rights violations. Political openness might also have the opposite effect, however. Politically open countries have more protests, journalists, and NGOs, all of which stimulate political and moral debate. Also, intermediate levels of democracy and the process of democratization itself are associated with greater risk of violent conflict, moreover, and this could trigger greater levels of abuses (Ron, Ramos and Rodgers 2005, Snyder 2000, ?). I thus anticipate only that Polity IV's effect will be significant, but remain agnostic as to its direction.

5.2.2.7 International Human Rights Legal Commitment

As I have discussed in the previous chapter, in order to measure international legal commitment, I look at state's ratification records of six major human rights conventions: International Covenant on Civil and Political Rights (ICCPR), Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment (CAT), the Convention on the Prevention and Punishment of the Crime of Genocide, Convention on the Elimination of Racial Discrimination (CERD), Convention on the Elimination of Discrimination against Women (CEDAW), and Convention on Enforced Disappearances (CED). Instead of looking at them respectively, I am going to use state's ratification records with a 7 point scale with "6" indicating

ratification of all six conventions and “0” none—that is, it measures the level of international legal commitment. I obtained the data from the United Nations Office of High Commissioner for Human Rights (OHCHR).

5.2.2.8 HROs and Target Distance

In order to measure this physical distance between HROs’ home country (i.e. head-quarter) and target states, I utilize the intercapital distance between HROs’ head-quarter and target countries, *the Correlates of War Direct Contiguity Data*.

Variables	Mean	Std. Dev.	Min	Max
Shaming Intensity	0.59	0.38	0	0.99
Human Rights (CIRI)	4.93	2.28	0	8
Civil Society	3.54	0.98	0	5.97
Approach	—	—	0	1
Democracy	2.78	6.72	-10	10
GDP pc (ln)	7.87	1.63	4.17	12.17
Population (ln)	15.41	2.15	9.11	21.01
Media Reports (ln)	5.75	2.29	0	12.37
Military (ln)	0.63	0.75	-3.06	4.75
HR Treaty	4.28	2.44	0	6
Distance (ln)	8.20	0.79	3.66	9.42

Table 5.1: Descriptive Statistics Summary (Intensity Model)

5.3 Estimation Method

With regard to the choice of the appropriate estimation procedure, I consider standard OLS inappropriate, because the dependent variable has a positive probability

mass at zero. As HROs usually shame some states, but not others, the shaming choice of a target is in many cases the corner solution of no shaming. In other words, because shaming intensity data are censored—that is, shaming intensity can be zero but never negative—I use Tobit models of HROs’ shaming with random effects for dyads, which offer advantages over other available methods. Much of the existing studies focuses on estimating the first stage—the decision to shame. One could use the two-part model or the Heckman selection model to allow for separate HROs decision on shaming and the level of shaming intensity. Yet, two-part model requires the assumption that the errors of the targeting model and the level of shaming intensity model are uncorrelated. In fact, it seems very unrealistic that they are independent. The Heckman selection model does not assume uncorrelated errors, but it requires the introduction of at least exclusion restriction, which means an explanatory variable that determines the targeting, but not the level of intensity of shaming. It is not that easy to find it at this point, if not possible.² Thus, I rely on Tobit models to account for this corner solution nature of the dependent variable (Wooldridge 2010). In this case, the Tobit model assumes that HROs simultaneously decide which countries are targeted and how severely it is going to be shamed. The data display heteroskedasticity and serial correlation, which can be corrected using either hierarchical modeling or robust, clustered standard errors (Primo, Jacobsmeier and Milyo 2007). I prefer dyad random effects because they explicitly model the underlying heterogeneity rather than treating it as a nuisance

² This issue has been heavily discussed in economics literature (Alesina and Weder (2002); Berthélemy and Tichit (2004); Neumayer (2006)).

(King and Roberts 2014), but the findings are robust to using either solution.

5.3.1 Tobit Regression Model

The structural equation in the Tobit model is:

$$y_i^* = X_i\beta + \epsilon_i \quad (5.1)$$

where $\epsilon_i \sim N(0, \sigma^2)$. y^* is a latent variable that is observed for values greater than τ and censored otherwise. The observed y is defined by the following measurement equation

$$y_i = \begin{cases} y^* & \text{if } y^* > \tau \\ \tau_y & \text{if } y^* \leq \tau \end{cases} \quad (5.2)$$

where $\tau = 0$, i.e. the data are censored at 0. The likelihood function for the censored normal distribution is:

$$L = \prod_i^N \left[\frac{1}{\sigma} \phi \left(\frac{y - \mu}{\sigma} \right) \right]^{d_i} \left[1 - \phi \left(\frac{\mu - \tau}{\sigma} \right) \right]^{1-d_i} \quad (5.3)$$

where τ is the censoring point. In the traditional tobit model, we set $\tau = 0$ and parameterize μ as $X_i\beta$. This gives us the likelihood function for the tobit model:

$$L = \prod_i^N \left[\frac{1}{\sigma} \phi \left(\frac{y - X_i\beta}{\sigma} \right) \right]^{d_i} \left[1 - \phi \left(\frac{X_i\beta}{\sigma} \right) \right]^{1-d_i} \quad (5.4)$$

The log-likelihood function for the tobit model is:

$$\ln L = \sum_{i=1}^N \left\{ d_i \left(-\ln \sigma + \ln \phi \left(\frac{y_i - X_i \beta}{\sigma} \right) \right) + (1 - d_i) \ln \left(1 + \Phi \left(\frac{X_i \beta}{\sigma} \right) \right) \right\} \quad (5.5)$$

So, the overall log-likelihood is made up of two parts. The first part corresponds to the classical regression for the uncensored observations, while the second part corresponds to the relevant probabilities that an observation is censored.

5.3.2 Results

Before presenting my regression findings, I draw a simple plot to illustrate the link between HROs' information gathering approach and the level of shaming intensity. Figure 5.1 shows a sample of 20 HROs and their average level of shaming intensity over People's Republic of China during 1991-2011. As I have discussed earlier, China has often been shamed by HROs, but HROs do not always have the same opinions about China. All these HROs name and shame China, but some of them are harsher than others. For example, Amnesty International, Human Rights First, and Human Rights Watch along with other HROs (the top half of Figure 4) indicate the higher level of shaming intensity, whereas International Committee of the Red Cross and Duihua Foundation (bottom half of Figure 4) show less severe criticism. In other words, this quick perusal of HROs appearing in Figure 4 provides preliminary support for my hypotheses. HROs with an indirect approach are more likely to be harsher on states than HROs with a direct approach.

Although this information is suggestive, statistical modeling can help highlight

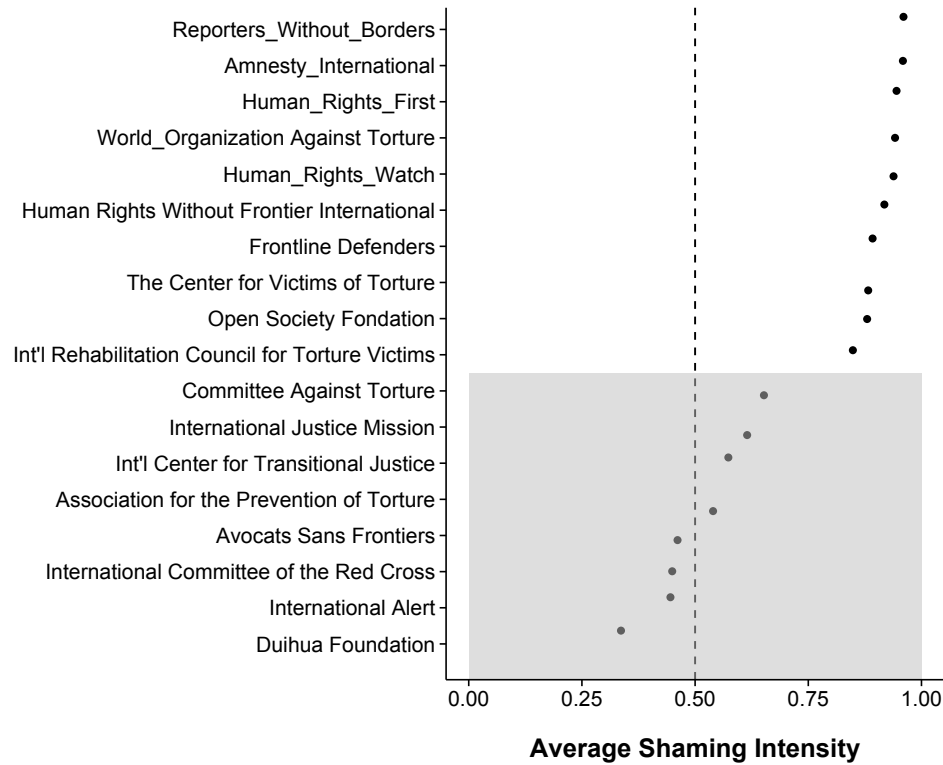


Figure 5.1: HROs Shaming Intensity and China

more specific factors associated with higher level of HROs' shaming. Table 5 presents the estimation results.³ The base model (model 4) assesses the effect of explanatory variables of interest. Before exploring the effects of the main explanatory variables of interest, I briefly discuss the results for the control variables.

As expected, GDP per capita ($\hat{\beta} = -0.033, p < 0.001$), population ($\hat{\beta} = 0.087, p < 0.001$), and military expenditures ($\hat{\beta} = 0.020, p < 0.001$) are statistically significant. International legal commitment ($\hat{\beta} = -0.017, p < 0.001$) and distance ($\hat{\beta} = -0.132, p < 0.001$) negatively affect HROs' shaming intensity. The results provide little support for an unconditional effect of democracy on HROs' shaming intensity. Turning to the effects of the main explanatory variables of interest, I find that states with higher civil society ($\hat{\beta} = 0.037, p < 0.001$) gets shamed more severely. Also, the results show that HROs' information gathering approach ($\hat{\beta} = 0.321, p < 0.001$) is positively related to the level of shaming, as expected—that is, the indirect approach increases the shaming intensity by 0.32 than HROs with a

³The reported estimates indicate the effects of the explanatory variables on the latent propensity to shame states.

direct approach, holding other things equal.

My findings, however, provide little support for an unconditional effect of human rights conditions on HROs' shaming intensity. This can be seen from the statistically insignificant coefficients for this variable, in which HR conditions is not interacted with another variable.⁴

⁴One could think of a number of different statistical models such as the two-stage models. But as I have discussed earlier, I do not place much faith in the two-stage modeling process because the statistical models are known to be biased and seem more aimed at modeling a overly simplistic story. Instead, I re-estimate using OLS with random effects for the entire sample. The results are pretty much similar. See Appendix

Table 5.2: Determinants of Shaming Intensity

Variables	Model (4)
Human Rights	0.001 (0.002)
Democracy	0.001 (0.001)
Civil Society	0.037** (0.008)
Media	0.062** (0.002)
GDP	-0.033** (0.004)
Population	0.087** (0.005)
Military	0.020** (0.006)
HR Treaties	-0.017** (0.002)
Distance	-0.1325** (0.007)
Info.Approach	0.321** (0.012)
Constant	-1.325** (0.098)
Observations	116,337
Left-censored Observations	88,571
Number of dyads	6,678
Log-likelihood	-44404.143

Levels of Confidence: * $p < .1$, ** $p < .05$, *** $p < .01$.
Standard errors clustered on dyads in parentheses.

5.4 Conclusion

The findings in this chapter support that HROs need information about states' actual human rights practices and how they collect information influences how they do naming and shaming. HROs essentially choose over two broad approaches to information gathering: an indirect approach, in which they leverage the information provided by domestic communities and a direct approach, in which they collect information themselves. Choices over these approaches, however, requires HROs to confront a critical dilemma, the resolution of which ensures that either aggregate information about human rights practices will be biased or that the international human rights network cannot be fully coordinated or both.

In this chapter, I have found that a direct approach presents HROs with an opportunity to investigate claims on the ground, to question officials, and to make more personal links with activists. By communicating with a government, HROs may also have a chance to directly persuade them to alter their practices. They may learn better which practices are more or less difficult to reform, in light of political realities, and that information may be critical to real policy change. On the other hand, investigation missions are costly and governments have incentives to hide serious violations from view. HROs have to find ways to observe and understand the contexts they monitor properly. To gain access, they may have to "negotiate" what information they will release publicly.

Thus, the chapter's findings are frustrating and also pose challenges for HROs. Most notably, the real world of abuse alone does not make a meaningful difference.

While HROs' universe of obligation includes all of humanity and human rights, the findings suggest that HROs' savvy dynamics and the lack of coordination may produce confusion in audiences and may provide governments with the "social camouflage" to deflect criticism. Consequently, the findings in the chapter not only clarifies the critical dilemma that HROs confront, but also propose to reevaluate the impact analysis of naming and shaming.

Chapter 6

Choice of Information Gathering Approach

6.1 Introduction

Chapter 2 outlines a set of conditions under which HROs are likely to choose information gathering approach: indirect or direct. There are two primary reasons—material costs and political costs. This chapter tests the hypotheses of HROs' choice of information gathering approach through statistical analyses. Section II discusses the data on HRO's material and political costs. Section III estimates a set of regression models to test hypotheses. Section V concludes.

6.2 HROs and Two Information Gathering Approach

The theoretical framework laid out in Chapter 2 suggests that there are broadly two approaches to information gathering—an indirect and a direct way. It is very costly for HROs to collect information directly. The resources expended to gather information vary, and HROs differ significantly in membership, structure, size of staffs, and financial capability. Most HROs, particularly non-governmental organizations, cannot afford to send staff members on direct information collecting mission (Murdie and Bhasin 2011, Simmons 2009). Investigative missions are expensive, and the availability of funds will affect the quality of a mission’s findings, determining how long a mission will last and how many members the mission can include. Consequently, despite its crucial role, direct information gathering is not always an option for HROs. Considering their limited resources, HROs cannot realistically take on all of the world’s human rights violations equally. They have to make strategic and tactical decisions about what issues to prioritize and which countries “deserve” their attention. This costliness often constraints actions that HROs can undertake.

Information gathering is also politically costly. While communicating with governments, HROs have to balance or compromise their strict human rights advocacy principles for information gathering. In order to gain access to the information, HROs may have to negotiate or compromise the information that they will release publicly. In particular, directly communicating or cooperation with governments can

be used against HROs. Governments may use their cooperation or direct contact with HROs as evidence that their policies on human rights are not so bad and works as “social camouflage” that can lower the expected costs associated with criticism.

Consequently, the politics of information gathering creates a dilemma for HROs. HROs can rely on third parties for information, which allows for wider coverage at lower cost yet risks biased information depending on the nature of domestic civil society. Instead, HROs can seek information directly, but at great political cost and typically with strings attached.

This argument suggests the following implications.

H6A: HROs with material capabilities are more likely to go with direct information gathering.

H6B: HROs with higher political costs are more likely to go with indirect information gathering.

6.2.1 Dependent Variable

As discussed in Chapter 4, in order to measure HRO’s information gathering approach, I pay attention to organizational level on the basis of each organization’s mission and value. In general, a HRO explains its operational methods under the section of something like “How We Work”, “About Us,” etc. In addition, I read through a sample of each HRO’s document and identify how they work in practice. A direct information gathering assumes a value “0” and an indirect approach takes the value “1.”

6.2.2 Independent Variables

6.2.2.1 Material Costs

Material costs is directly related to HROs' financial resources. Well-resources HROs are less likely to be constrained by the cost of information gathering missions. To measure organization's resource availability, I rely on HRO's annual budget as reported in the *Yearbook of International Organizations* and each HRO's financial reports.

I also use the size of staffs as another indicator of HRO's resources. Because having a large number of full-time staffs requires more resources (e.g., maintaining offices, payments, etc.), it is a good indicator. In addition, not all HROs release financial reports. Although one could think that HROs are supposed to be honest and transparent about their spending, it is not always the case. It seems that financially struggling organizations are less likely to report its full financial statements.¹

6.2.2.2 Political Costs

Measuring political costs for HROs is based on HROs' funding sources. As discussed in chapter 2, HROs are not fundamentally different from other actors such as interest groups or business firms. While their motivations may be noble and lofty, they behave in much the same as other actors in that they are all concerned about their organizational survival. HROs are concerned with remaining financially solvent while

¹For example, larger HROs like Amnesty International or Human Rights Watch or Human Rights First, publish their financial reports regularly even with independent auditor's reports. Preparing financial reports for public can be expensive and requires more works.

trying to change the behavior of abusive governments. Consequently, HROs, as a watchdog, rely primarily on donations from private individuals because of a desire to remain autonomous from states. It is especially true for non-governmental organizations rely on donations. They have to make themselves attractive to potential members, donors, and supporters. They have to spend a lot of time in fundraising (Aldashev and Verdier 2010, Andreoni and Payne 2003). Precisely because of this reason, HROs are often held accountable for how they spend the money they receive from donors. In this context, political costs are measured by the sources of funding. If it is primarily funded by private donors—that is, non-government funding, then it would be relatively independent of the government subject to scrutiny. If an HRO is primarily dependent on government funding, it would be difficult to be critical of the government. Therefore, if a HRO is primarily funded by private donors, it takes the value “1,” and if it is primarily funded by governments, it is coded as “0.”

6.2.3 Control Variables

6.2.3.1 ECOSOC Consultative Status

Consultative status to the UN Economic and Social Council (ECOSOC) provides NGOs with access to not only ECOSOC, but also to its many subsidiary bodies, to the various human rights mechanisms of the United Nations, ad-hoc processes on small arms, as well as special events organized by the President of the UN General Assembly. Consequently, it provides HROs with more opportunities to gather information. HROs with consultative status enjoy far more privileges to

participate in intergovernmental meetings of the UN than other association with the UN Department of Public Information. For example, Only NGOs in consultative status with the United Nations Economic and Social Council (ECOSOC) can be accredited to participate in the Human Rights Councils sessions as Observers (UN OHCHR 2015).²

Although HROs with consultative status cannot decide, they can provide for recommendations that are contained in annual reports of the UN ECOSOC. Because HROs with consultative status can attend the relevant international conferences convened by the UN and the meetings of the preparatory bodies of these conferences, etc., they have more opportunities to directly interact with governments officials and other HROs that they are trying to investigate and collect information.³ If a HRO maintains any of three categories of ECOSOC status, then it assumes value “1,” and otherwise “0.” The data can be obtained from the NGO Branch in the Department of Economic and Social Affairs under the UN ECOSOC.⁴

6.2.3.2 Members

In this dissertation, HROs refer to international human rights organizations, but HROs vary in terms of how well they are connected or “networked” to domestic so-

²<http://www.ohchr.org/EN/HRBodies/HRC/Pages/NgoParticipation.aspx>

³There are three categories of status: General consultative status, Special consultative status and Roster status. General consultative status is reserved for large international NGOs whose area of work cover most of the issues on the agenda of ECOSOC and its subsidiary bodies. Special consultative status is granted to NGOs which have a special competence in, and are concerned specifically with, only a few of the fields of activity covered by the ECOSOC. Organizations that apply for consultative status but do not fit in any of the other categories are usually included in the Roster.

⁴<http://csonet.org/content/documents/E-2014-INF-5%20Issued.pdf>

cieties. In their study of the effectiveness of HROs and naming and shaming, Murdie and Davis (2011) captures the presence of domestic HROs within a state using *the Yearbook of International Organization*. *The Yearbook* also lists total number of member states that a HRO has in a given year. It measures how widely a given HRO is connected to the international community and domestic societies around the world. *The Yearbook* lists countries represented or in which members are located. If HROs are better worked, it would give HROs a better information resources and more access to information. For example, Amnesty International has 25 national sections and more than 7 million supporters around the world (Amnesty 2015). The ICRC has 187 national chapters around the world (ICRC 2015). Protection International only has 6 national sections (PI 2015). In order to measure the national members, I count the total number of member states that a HRO has in a given year.

Variables	Mean	Std. Dev.	Min	Max
Approach	–	–	0	1
Political Cost	–	–	0	1
Budget (ln)	15.12	1.62	11.28	19.51
Staffs (ln)	3.39	1.21	0	6.21
ECOSOC	–	–	0	1
Members (ln)	3.35	1.28	0	5.84

Table 6.1: Descriptive Statistics Summary (Approach Model)

6.3 Estimation and Results

I estimate three models. All three models seek to explain if HROs gather information directly or indirectly. The dependent variable takes the value 1 if a HRO takes an indirect approach, and 0 if a direct approach. The first model uses the annual budget as the indicator of material costs. The second model uses the size of staffs as an alternative measure. The third model includes the interaction term between the political cost and the budget to understand the conditional effects. Together, they examine the characteristics of HROs that increase the chances they will take an indirect information gathering approach.

The sample is 93 HROs examined in the previous chapters.⁵ Because my dependent variable is a dichotomous variable, I use a probit regression analysis with robust standard errors. The table 6.1 presents the estimation results.

First and foremost, the model 1 reveals unambiguously that the political costs variable is statistically significant on the probability of choosing indirect information gathering approach. Substantively, the predicted probability that a HRO with high political cost chooses to gather information indirectly is about 0.42, holding continuous and dummy variables constant to their means and modes, respectively). Although the size of budget is negatively associated with the probability of taking indirect approach, the unconditional effect of budget is not statistically significant. ECOSOC consultative status increases the chance that HROs take indirect approach. HROs' members are not significantly related to the choice of information

⁵See Appendix for the full list.

Table 6.2: Determinants of Information Gathering Approach

Variables	Model (1)	Model (2)	Model (3)
Political Cost	0.677** (0.082)	0.707** (0.082)	0.666 (0.487)
Budget (ln)	-0.007 (0.016)		-0.048* (0.023)
Staff (ln)		-0.044 (0.034)	
Political Cost*Budget			-0.095** (0.033)
Members (ln)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
ECOSOC	0.229** (0.088)	0.257** (0.088)	0.197** (0.090)
Constant	-0.228 (0.247)	-0.497** (0.142)	0.388 (0.356)
N	1030	1023	1030
Log-likelihood	-669.704	-660.384	-665.140

Levels of Confidence: * $p < .1$, ** $p < .05$, *** $p < .01$. Robust standard errors adjusted for clustering on HROs in parentheses.

gathering approach. As shown in model 2, using the size of staff does not change the results in a meaningful way.

In order to clarify the effect of political costs on the probability of taking indirect approach conditioned upon financial capability, I look at the interaction effect between political costs and HRO's budget. There has been some recent debate about binary response models and interaction effects. Some scholars have made that it is not necessary to include an interaction term in logit or probit in order to take

account of conditional hypotheses since the non-linearity of these models implicitly force the effect of all the independent variables to depend on each other anyway. (Berry and Berry 1990, Berry, DeMeritt and Esarey 2010). Although the compression effect could be substantively meaningful and can be interpreted as such, I include the interaction term because I would like to specify that there exists the effects of interaction beyond and above compression effects that are always there. Model 3 in table 6.1 shows the probit model with the interaction term between political costs and budgets. Since the coefficient and standard error on the interaction in non-linear model do not tell us the direction, magnitude, or significance of the “interaction effect,” we need to calculate the interaction effect separately (Nagler 1991), Golder 2014).

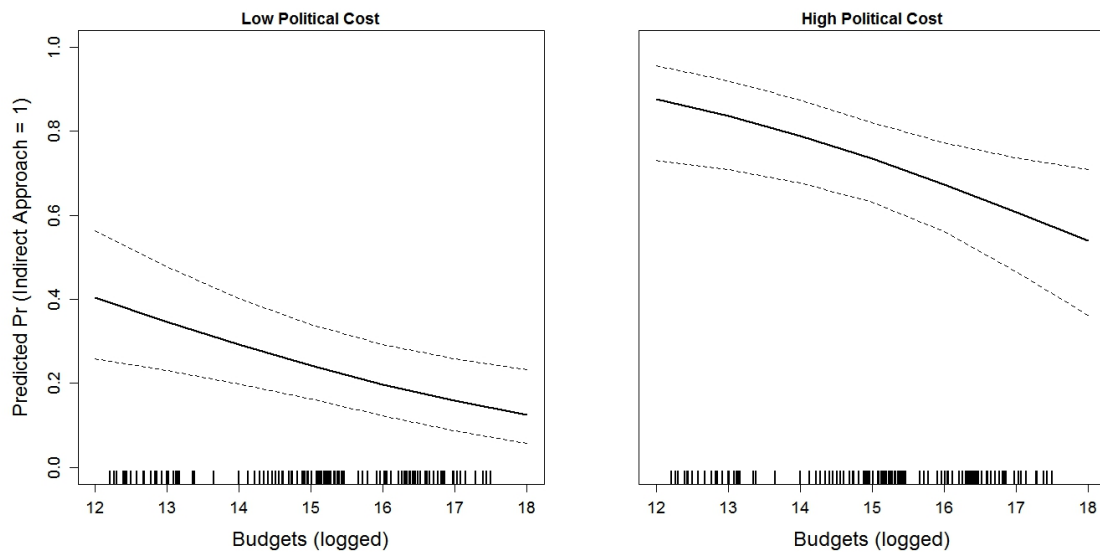


Figure 6.1: The Effect of Political Cost and Budget on Pr (Indirect Approach)

Figure 6.1 illustrates the changes of predicted probability of taking indirect approach for high and low political cost at all levels of budget. I also calculate a

first difference at each level of budget, fixing other variables at its mean and mode respectively. Figure 6.2 presents a graph showing this first difference in $\Pr(\text{Indirect Approach})$ as all levels of HRO budgets. The figure shows that the effect of political costs on the probability of voting increases as the levels of budgets. Specifically, an increase in budget from the minimum to maximum yields an incline in the first difference measuring the effect of political costs on $\Pr(\text{indirect approach})$ of 0.21 ($= 0.57 - 0.36$). So, when all other covariates are fixed at central values, higher political costs results in an increase in the probability of taking indirect approach at any level of material costs. The magnitude of the effect of political costs also increases as the levels of budget increase, although the magnitude is not as substantial as I expected.

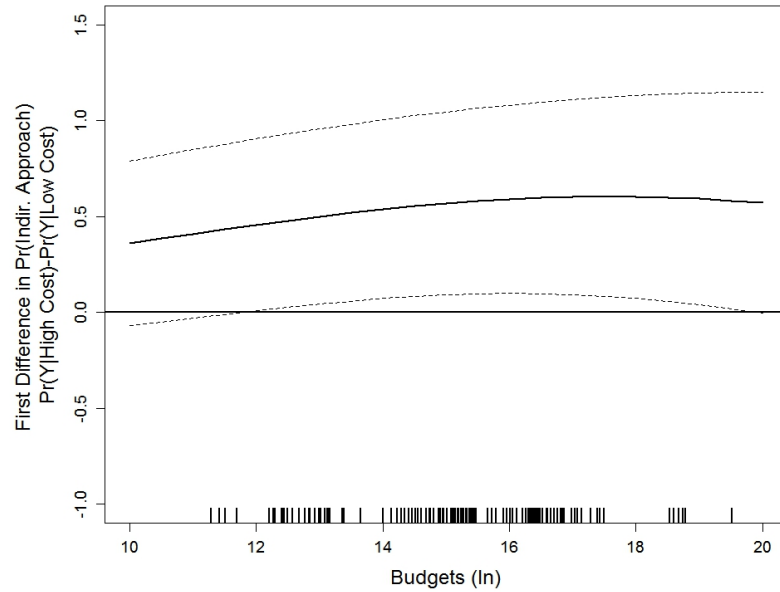


Figure 6.2: First Difference Measuring the Effect of Political Cost on $\Pr(\text{Indirect Approach})$ at All levels of Budgets

6.4 Conclusion

Throughout this dissertation I have emphasized the key role of HROs' information gathering approaches in naming and shaming. How they gather information influences how they do naming and shaming. It is an important factor to explain whether or not a state is targeted, and once targeted, how severely a targeted state is shamed.

In this chapter, I examined why HROs choose particular information collection strategies, focusing on variation in the material costs of investigative missions and the political costs of potentially risking an image of impartiality. The preceding evidence in the findings suggests that HROs' choice is primarily determined by the political costs than the material costs. However, it is worth noting that there exists the conditional effects of the political costs with the material costs on HROs' choices. In other words, HROs with more financial capability tend to have more leeways in following their political principles.

The findings in this chapter clarifies that HROs work hard to deal with the imminence of human rights violations, but at the same time they do not exist in an advocacy vacuum.

Chapter 7

Conclusion

In 2004, the Human Rights Watch (HRW) was very critical of Uzbekistan's human rights conditions, particularly, deaths in police custody. One of the biggest cases against the government was to Andrei Shelkoveko's death in a police station. The statement HRW released on May 21st, 2004 suggested that Shelkovenko detained by officers of the Byzton district police station (Gazalkent, Tashkent region) had died in the detention cell under severe torture (HRW 2004). Solely relying on the information from Shelkvenko's lawyer, witnesses, and relatives, HRW concluded that the detainee had serious cuts on the neck and a fatal wound to his head, legs, and a number of bruises and emphasized that he was beaten to death by police.

After a series of HRW's harsh naming and shaming on the case, the Uzbek government decided to invite an international investigation team that included an independent Canadian forensic expert, Michael Sven Pollanen (his services had been enlisted in various investigation of human rights abuses such as Cambodia, East Timor,

and Kazakhstan) and Freedom House, the U.S.-based rights group. The group was “promised access to each and every step of the way” (Herman 2004). The team and Dr. Pollanen conducted a detailed autopsy and examination of Shelkovenko’s neck revealed evidence of hanging but nothing indicated that the man had been strangled and the results of the autopsy did not show that Shelkovenko had been subjected to tortures and physical abuses that resulted in traumas to the head and scrotum¹ (Pollanen 2004). Consequently, the team concluded that Shelkovenko had committed suicide and did not reveal any signs of torture, but the alleged injuries were misinterpretations of changes that occur in all bodies after death (decomposition and decay). Holding a public press conference, Freedom House thanked the authorities of Uzbekistan for their cooperation and assured that it would continue its activities aimed at protection of human rights and prevention of torture in Uzbekistan (Kudryashov 2004). The Uzbek government immediately backfired on HRW and dismissed all the allegations of torture, saying “HRWs reporting is biased and intended to undermine Uzbekistans international reputation and to sabotage its relationships with other governments” (HRW 2007, 73). HRW had to publicly acknowledge its erroneous conclusions on attributing the cause of Shelkovenko’s death to torture (HRW 2007, 74).

This HRW and Uzbekistan “scandal” sheds some light on a neglected aspect of human rights organizations and their advocacy campaigns. As a non-state actor, HROs try to convert moral authority into “ideational power” to influence national

¹An open wound about 5cm long on the head, bruises and a lacerated wound on the right side of the neck, bruises on the inside of the right forearm, bruises on the shoulders, swelled scrotum, a long lacerated wound on the right thigh (Pollanen 2004, 114)

governments. The power comes from their ability to disseminate information about human rights abuses. In order to do naming and shaming, HROs need information about state's human rights conditions. Accurate information is most critical. Disseminating inaccurate information can seriously undermine the reputation and legitimacy of HROs. Human rights scholars, however, have largely assumed that this information is exogenous to human rights politics or at least freely available. In other words, scholars have neglected the importance of collecting information and how HROs gather information influences how they do naming and shaming. Since HRW solely relied on indirect sources of information, it did not have any direct dialogue or discourses with the government and did not hesitate to criticize them severely. In contrast, Freedom House (FH) did not rush to blame the government. Consequently, FH was invited to the investigation and had the chance to observe the detailed autopsy and released accurate information, and maintained its cooperative relationship with the government.

Many human rights scholars and practitioners have placed greater faith in the abilities of HROs to promote respect for human rights. Unencumbered by partisan interests or politics, HROs seem to have fewer constraints, greater morality, deeper ethical commitments, and more freedom of maneuver. To those scholars, policy makers, and activists, HROs appear to be the ideal global moral agents, and they are able like no other to construct a robust universe of human rights obligation.

Close examination of these HROs, however, reveal a more complex picture. While HROs have made considerable achievements, they also have to operate with

limited resources, forcing them to seek for more pragmatic and politically savvy strategies. This dissertation provide a unique and systematic study of the strategic considerations shaping HROs' agenda.

7.1 Summary of The Dissertation

The broadest objective of this dissertation is to understand the tasks of information collection and HRO naming and shaming. It starts from two empirical patterns of HRO advocacy. Why are some states frequently shamed for human rights violations despite having reasonable human rights records while many states are rarely criticized despite being serious violators? Why are states shamed differently by different HROs for precisely the same human rights violations? I contend that the process by which HROs collect information is the key to understanding to these patterns. HROs that primarily rely on third parties to collect information are more likely to focus on states with relatively good records; and, they tend to criticize these states severely. In contrast, HROs that directly engage states are more likely to pay attention to the worst human rights violators but they tend to shame these states less severely. HROs that indirectly engage states via third parties receive a relatively large amount of information from states in which there is a well-organized civil society prepared and willing to expose violations, and these states generally have reasonable human rights records. HROs that directly engage states focus on serious abuses; however, in order to gain access, HROs tend to negotiate the information that they will release publicly. Moreover, I explain why HROs choose

particular information collection strategies, focusing on variation in the material costs of investigative missions and the political costs of potentially risking an image of impartiality.

In Chapter 3, in an attempt to evaluate the empirical implications of my argument, I explain a comprehensive dataset on the shaming behavior of HROs. Utilizing natural language processing techniques, I examine and process nearly 40,000 human rights reports from 93 HROs. Among other things, I characterize both the targets and the aggressiveness of each HROs shaming activity.

On the basis of the data, in Chapter 4 and Chapter 5, I test the shaming hypotheses derived from the theory in Chapter 2. The results suggest that by choosing an indirect approach, HROs tend to have far more information in places with strong civil societies and they are likely to both over-report on states with generally good records and under-report on states with relatively bad records. HROs that directly engage states; however, tend to focus on states with the worst human rights conditions, but they are less likely to criticize states severely.

In Chapter 6, I test the choice hypotheses. I find that HROs with the high political costs are positively associated with an increase in the probability of taking an indirect information gathering approach. Although the results do not support that the material costs alone are a statistically significant determinant of HROs choice of information collection, I find that the conditional effect of the political costs on the probability of choosing indirect approach increases as the levels of material costs decrease.

7.2 Implications and Future Research Agenda

7.2.1 Impact Analysis

The central goal of naming and shaming is that HROs disseminate information about human rights abuses that governments would prefer to remain hidden in the dark. They collect, assess, and publicize information in order to bring pressure and persuade governments to comply with international human rights standards. In this dissertation, I argue that understanding the determinants of HROs' strategies is a methodological precondition for developing proper assessment of the effectiveness of HRO's activities. Analyses that fail to account for the process by which HROs select their targets are likely to produce biased estimates of the efficacy of the resulting naming and shaming efforts. If the severe human rights abuses and human suffering generated by a government has little bearing on whether and how HROs engage in naming and shaming strategy, or if states that are relatively doing well receive systematically higher levels of criticism than others, no one would take the criticism seriously or feel pressured by international public opinion. In addition, scholars have ignored that some HROs tend to produce weaker and softer criticism than other HROs. It would be difficult to expect that "soft" criticism could bring much pressure to targeted governments. More to the point, as illustrated in this project, it is not unusual to note the competitiveness and the lack of coordination among HROs, resulting in different or conflict opinions about states. When two different HROs say different things about the same country, it would be difficult to mobilize

consistent public opinion about the targeted states. The lack of coordination unduly empowers abusive governments, helping them to evade international pressure by playing a loud-mouthed HRO off against a less-vocal HRO. All the existing works measure naming and shaming on the basis of the yearly count of the number of shaming events that occurred toward a government in a given year. By simply counting the number of events in a given year, they assume that all shaming events are the same. Consequently, the effect of the event count analysis might have been overestimated or underestimated by previous studies. This research suggests that it is important to re-evaluate the effectiveness of naming and shaming. Because HROs' different approaches to information gathering produce different kind of information, it is worth taking this critical difference into estimating the effectiveness of shaming.

7.2.2 HROs and Media Interaction

Another important aspect of naming and shaming is a key role played by the international media. In Chapter 2, I discuss that the international news media is not an HRO and the media's human rights coverage is quite different from HRO reports. However, it would be unfair to dismiss their role in the process of shaming. Widespread media coverage can certainly contribute to pressuring abusive states throughout several avenues (Keck and Sikkink 1998, Risse and Sikkink 1999). First, media coverage make the public more aware of abuses and can help to mobilize people to condemn or protest human rights abuses; media coverage can encourage concerned world citizens to support the efforts of HROs through volunteering

or providing financial support to HROs. Second, media coverage can encourage sympathetic governments to condemn human rights abuses and in some cases take bilateral or multilateral actions to pressure states to end abuses. These government actions may be driven by public pressure or moral outrage, both influenced by media coverage of abuses. Third, media coverage of HRO reports can help focus the attention of international and regional organizations, as well as other HROs, on human rights abuses in certain regimes. Finally, media coverage of HRO reports can increase awareness of citizens within the abusive state of their government's actions, potentially leading to an eroding of the perceived legitimacy of the government (Davis, Murdie and Steinmetz 2012). Clearly, media coverage of HRO reports of human rights abuses can have an important impact on society's awareness of abuses and efforts to prevent abuses.

While there is a great deal of coverage of human rights abuses in the international media, the cross-national empirical academic literature on the causes and effects of HRO reports and international media coverage of HRO reports is still in its infancy. Particularly, it has been largely ignored an important middle step—the interaction between HROs and the international media: how and why certain HRO reports make it into the international media press. HROs are tirelessly working on a number of human rights issues at the same time, often releasing multiple press releases and reports during the same day. Do international media outlets—newspapers, television broadcasters, and cable news networks - pay more attention to some HROs compared with others? Do reports of human rights abuses in some countries garner more

coverage by the international media that abuses in other countries? Why do some of these issues make it to the international stage and others do not? Are there certain characteristics of the abuses HROs report that make certain situations more likely to gain coverage in the international media? Each of these questions merits further research, both for our theoretical understanding of how non-state actors matter in world politics but also because answering these questions could help human rights promotion in repressive states.

Appendix A

Appendix

HROs List

African Centre for Democracy and Human Rights	Commonwealth Human Rights Initiative
Al-HAQ Defending Human Rights	Commonwealth Secretariat
Amman Center for Human Rights	Defenders of Human Rights Center
Amnesty International	Doctors without Borders
Anti-Slavery International Article 19	Duihua Foundation
Asia-Pacific Human Rights Information Center	Euro-Mediterranean Human Rights Network
Asian Centre for Human Rights	European Association for the Protection of Human Rights
Asian Human Rights Commission	European Centre for Law and Justice
Asian Legal Resource Centre	European Committee for the Prevention of Torture
Asslciation for the Prevention of Torture	European Ombudsman Institute
Avocats Sans Frontiers	Families of Victims of Involuntary Disappearance
Canadian Lawyers for International Human Rights	Foundation for Human Rights Initiative
Center for Justice and International Law	Freedom House
Center for Victims of Torture	Front Line Defenders
Child Soldiers International	Global Anti-Aggression Campaign
Civil Society Watch	Global Human Rights Defence
Committee Against Torture	Human Rights Committee
Committee on the Elimination of Racial Discrimination	Human Rights Defence Center

Human Rights First	International Service for Human Rights
Human Rights Foundation	International Society for Human Rights
Human Rights Watch	International Training Centre for Human Rights and Peace Teaching
Human Rights Without Frontier International	International Women's Democracy Network
Institute for Research and Education on Human Rights	Interpeace
Inter-American Press Association	Irish Centre for Human Rights
International Alert	Islamic Human Rights Commission
International Association of Official Human Rights Agencies	Jammu and Kashmir Council for Human Rights
International Bar Association-Human Rights Institute	Journalists for Human Rights
International Center for Transitional Justice	Law Association for Asia and Pacific
International Centre for the legal protection of human rights	Law Association for Asia and Pacific Lawyers' Rights Watch Canada
International Commission on Jurists	Minority Rights Group International
International Committee of Red Cross	No Peace Without Justice International
International Council on Human Rights Policy	Open Society Foundation
International Federation of Action by Christian for the Abolition of Torture	Pax Christi International
International Federation of Journalists	Penal Reform International
International Freedom of Expression Exchange	Physicians for Human Rights
International Human Rights Commission	Protection International
International Humanist and Ethical Union	Public International Law and Policy Group
International Justice Mission	REDRESS
International League for Human Rights	Reporters without Borders
International Movement Against All Form of Discrimination and Racism	Society for Threatened People
International Organization for the Elimination of All Forms of Racial Discrimination	TRIAL
International Peace Bureau	The Advocates for Human Rights
International Rehabilitation Council for Torture Victims	The Arabic Network for Human Rights Information
	The Institute on Religion and Public Policy
	The Latin American Association for Human Rights
	World Council of Churches
	World Organization Against Torture

States List

Australia	Egypt	Libya	Seychelles
Austria	El Salvador	Liechtenstein	Sierra Leone
Azerbaijan	Equatorial Guinea	Lithuania	Singapore
Bahamas	Eritrea	Luxembourg	Slovakia
Bahrain	Estonia	Madagascar	Slovenia
Bangladesh	Ethiopia	Malawi	Somalia
Barbados	Fiji	Malaysia	South Africa
Belarus	Finland	Maldives	South Sudan
Belgium	France	Mexico	Spain
Belize	Gabon	Monaco	Sri Lanka
Benin	Gambia	Mongolia	Sudan
Bhutan	Georgia	Montenegro	Suriname
Bolivia	Germany	Morocco	Swaziland
Bosnia	Ghana	Mozambique	Sweden
Botswana	Greece	Myanmar	Switzerland
Brazil	Grenada	Namibia	Syrian
Bulgaria	Guatemala	Nauru	Tajikistan
Burkina Faso	Guinea	Nepal	Thailand
Burundi	Guinea Bissau	Netherlands	Macedonia
Cabo Verde	Guyana	New Zealand	Togo
Cambodia	Haiti	Nicaragua	Tonga
Cameroon	Honduras	Niger	Tunisia
Canada	Hungary	Nigeria	Turkey
Central African Re- public	Iceland	Norway	Turkmenistan
Chad	India	Oman	Tuvalu
Chile	Indonesia	Pakistan	Uganda
China	Iran	Palau	Ukraine
Colombia	Iraq	Panama	United Arab Emi- rates
Comoros	Ireland	Papua New Guinea	United Kingdom
Congo	Israel	Paraguay	Tanzania
Costa Rica	Italy	Peru	USA
Cte D'Ivoire	Jamaica	Philippines	Uruguay
Croatia	Japan	Poland	Uzbekistan
Cuba	Jordan	Portugal	Vanuatu
Cyprus	Kazakhstan	Qatar	Venezuela
Czech Republic	Kenya	South Korea	Viet Nam
North Korea	Kiribati	Moldova	Yemen
Denmark	Kuwait	Romania	Zambia
Djibouti	Kyrgyzstan	Russia	Zimbabwe
Dominica	Lao	Rwanda	
Dominican Repub- lic	Latvia	Samoa	
Ecuador	Lebanon	Saudi Arabia	
	Lesotho	Senegal	
	Liberia	Serbia	

Document Examples

The Most Negative Case

2005 Brazil Excessive use of police force against protesters, arbitrary detention and abuse of 19 students

The International Secretariat of OMCT has been informed about the excessive use of force used against protesters by police and members of the arbitrary arrest of three students identified as Marcelo Verger Flora and Andr Muller Iron Moro and other 16 students (whose names have not been reported so far), as well as the mistreatment of them when they were in detention, in clashes in Florianopolis, Santa Catarina State. According to the information of Grupo Tortura Nunca Mais-RJ member organization of OMCT network, on May 30, 2005 were arrested by police officers Marcelo students Verger, Flora and Andr Muller Iron Moro when heading home after participating in a peaceful public demonstration. It is alleged that the arrested students, some of them injured during the police action were taken to the prison in the capital of Santa Catarina where they were subjected to serious humiliation. According to reports, these students were subject to torture, among which is denounced the obligation to undress and kneel naked on the floor being cold for over an hour, position, during which they were beaten in the head. Furthermore, according to the complaints were detained in the worst section of the prison, where prisoners were tuberculosis and highly dangerous. According to reports, since that time, 19 students imprisoned in Santa Catarina were released on bail, but are accused of crimes they did not commit and remain intimidated by police repression and at the risk of being arrested again, in which case would no longer have a new right to bail out.

The International Secretariat OMCT expresses its condemnation of these acts of extreme violence, and expresses its concern about the fact that these serious human rights violations do not go unpunished and the perpetrators without penalty therefore urges the authorities of Brazil due continuity of a thorough and impartial investigation into these serious facts, in order to identify those responsible, bring them to trial and apply the penal and/or administrative sanctions provided by law. The Brazilian government should take steps to ensure that the statements made by the civilian population are not suppressed by excessive use of police force; ensure respect for human rights and fundamental freedoms throughout the country in accordance with international human rights standards, in particular the Convention Against Torture, ratified by Brazil.

The Moderate Case

Oct 2010 Training health and legal professionals from Armenia, Azerbaijan and Georgia

In late September 2010, a team of three International Rehabilitation Council for Torture Victims (IRCT) staff conducted an eight-day mission to Georgia as part of the project Use of Forensic Evidence against Torture. The mission had three key objectives: to provide training to a group of approximately 50 health and legal professionals from Armenia, Azerbaijan and Georgia on documentation of torture; to conduct a forensic physical and psychiatric evaluation of an alleged torture victims and to launch an advocacy strategy aimed at improving investigation and prosecution of torture offences in Georgia. “For me, the most impressive aspect of the training was that representatives from Azerbaijan and Armenia two countries ‘in a state of no peace’ not only worked together, but worked together well.” said Dr Bente Rich, Psychiatric at the IRCT. Dr nder zkalipci, IRCT Medical Director, added, “What impressed me was the high caliber of participants in the training - including the Heads of National Medical Associations, Heads of Psychiatric Departments at Universities and Heads of Forensic Medical Departments.”

The case selected and examined by the IRCT in collaboration with local member centre EMPATHY, related to a 16-year-old boy who has allegedly been severely beaten in police custody with the objective of extracting a confession and punishing him. The case, illustrates the way people are tortured and subsequently denied an effective investigation and prosecution in Georgia. An official complaint was filed on 20 April 2010 and now almost 6 months later, the case is still under investigation under the less severe prohibition of abuse of official powers and there is a risk that the investigation will soon be stopped as part of a deal trading a lesser sentence in the boys criminal for withdrawal of his complaint.

In response to this situation, the IRCT and EMPATHY has launched a collective advocacy campaign, which will seek to apply domestic as well as international pressure on Georgia to improve both its legislation and practice in relation to effective investigations of torture cases especially in relation to use of medical evidence. The campaign was launched at a press conference on 28th September in Tbilisi and which was attended by various national media outlets, the Georgian Public Defender (Ombudsman), and representatives of the Ministry of Corrections and Legal Assistance, the EU, ICRC, many national NGOs and several embassies. During the coming 18 months, the IRCT and EMPATHY will implement a number of events and strategic interventions to further the objectives of the campaign.

The Most Positive Case

Sep 2010 Interactive Dialogue regarding children in armed conflict

The Special Representative of the Secretary-General on children and armed conflict, Ms. Radhika Coomaraswamy, presented her report noting significant progress made on children and armed conflict in the past year. During Ms. Coomaraswamy's visit to the Sudan, constructive dialogue with the national authorities resulted in a binding agreement with the Sudan People's Liberation Army (SPLA) to end the recruitment of child soldiers. Similarly, in Uganda the Special Representative is cooperating with the national army to implement procedures for the release and repatriation of children associated with the Lord's Resistance Army. There was broad agreement among States on the necessity to address the issues affecting children as a result of armed conflict, especially the challenges to internally displaced children. The report said children were particularly vulnerable to becoming victims of grave violations, given the majority of child soldiers are recruited from internally displaced camps.

In this vein, Ms. Coomaraswamy recommended steps for national authorities, to enhance the protection of internally displaced children. Ms Coomaraswamy also put forward a working paper on the rights and guarantees of internally displaced children in armed conflict. The paper aims to provide guidance to governments and other authorities on their responsibilities to provide protection to children. It was largely acknowledged that it remains the central responsibility of States to ensure the protection of children, especially during conflicts. Security Council Resolution 1882 of 2009 strengthened the protection framework for children in conflict, by expanding the monitoring and reporting mechanism to include those who kill and maim children and commit sexual violence against them. Ms Coomaraswamy highlighted the UPR process as a useful way of putting pressure on States under review to live up to their responsibilities.

Alternative Intensity Model

Like most studies using statistical analysis, I examined more models than I can conveniently present in the text. This section reports an alternative model. I utilize OLS with random effects. Precisely because of the large amount of zeros—that is, censoring—in the outcome data, OLS estimates are biased and the magnitude of some coefficients are reduced. As presented in Table A.1, the results from OLS with random effects are very similar to the results from Tobit with random effects in the original models.

Table A.1: Determinants of Shaming Intensity: OLS with RE

Variables	Model (1)
Human Rights	-0.001 (0.001)
Democracy	0.001 (0.001)
Civil Society	0.037** (0.008)
Media	0.010** (0.001)
GDP	-0.004** (0.001)
Population	0.010** (0.001)
Military	0.001 (0.001)
HR Treaties	-0.002** (0.001)
Distance	-0.019** (0.002)
Info.Approach	0.067** (0.002)
Constant	-0.056** (0.021)
Obervations	114,807
Number of dyads	6,678

Levels of Confidence: $*p < .1$, $**p < .05$, $***p < .01$.
Standard errors clustered on dyads in parentheses.

Bibliography

- AI. 2000. *Amnesty International Cumulative Guide, 1962-2000*. London: Amnesty International.
- AI. 2008. *China: Legacy of the Beijing Olympics: Issues and Facts: Fair Trials For All-China's Choice*. ASA 17/030/2008 London: Amnesty International.
- Ai, Chunrong and Edward C Norton. 2003. "Interaction terms in logit and probit models." *Economics letters* 80(1):123–129.
- Aldashev, Gani and Thierry Verdier. 2010. "Goodwill bazaar: NGO competition and giving to development." *Journal of Development Economics* 91(1):48–63.
- Alesina, Alberto and Beatrice Weder. 2002. Do corrupt governments receive less foreign aid? Technical report National Bureau of Economic Research.
- Ananiadou, Sophia, Sylvie Albert and Dietrich Schuhmann. 2000. "Evaluation of Automatic Term Recognition of Nuclear Receptors from MEDLINE." *Genome Informatics* 11:450–451.
- Andreoni, James and A Abigail Payne. 2003. "Do government grants to private charities crowd out giving or fund-raising?" *American Economic Review* pp. 792–812.
- Atkins, David C and Robert J Gallop. 2007. "Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models." *Journal of Family Psychology* 21(4):726.
- Ball, Patrick Donnell, Paul Kobrak and Herbert F Spirer. 1999. *State violence in Guatemala, 1960-1996: a quantitative reflection*. Advancement of Science.
- Bell, Daniel and Joseph H. Carens. 2004. "The Ethical Dilemmas of International Human Rights and Humanitarian NGOs: Reflections on a Dialogue between Practitioners and Theorists." *Human Rights Quarterly* 26(2):300–329.
- Bell, Nancy. 2008. "Ethics in child research: rights, reason and responsibilities." *Children's Geographies* 6(1):7–20.

- Berry, Frances Stokes and William D Berry. 1990. "State lottery adoptions as policy innovations: An event history analysis." *American political science review* 84(02):395–415.
- Berry, William D, Jacqueline HR DeMeritt and Justin Esarey. 2010. "Testing for interaction in binary logit and probit models: is a product term essential?" *American Journal of Political Science* 54(1):248–266.
- Berthélemy, Jean-Claude and Ariane Tichit. 2004. "Bilateral donors' aid allocation decisionsa three-dimensional panel analysis." *International Review of Economics & Finance* 13(3):253–274.
- Bloodgood, Elizabeth A. 2011. "The interest group analogy: international non-governmental advocacy organisations in international politics." *Review of International Studies* 37(01):93–120.
- Bob, Clifford. 2005. *The Marketing of Rebellion: Insurgents, Media, and International Activism*. New York: Cambridge University Press.
- Brecher, Michael and Jonathan Wilkenfeld. 2014. *International Crisis Behavior Project*. Maryland: University of Maryland.
- Brysk, Allison. 1993. "From Above and Below: Social Movements, the International System, and Human Rights in Argentina." *Comparative Political Studies* 26(3):259.
- Cameron, A Colin and Pravin K Trivedi. 2013. *Regression analysis of count data*. Vol. 53 Cambridge university press.
- Carpenter, R Charli. 2007. "Setting the advocacy agenda: Theorizing issue emergence and nonemergence in transnational advocacy networks." *International Studies Quarterly* 51(1):99–120.
- Cingranelli, David L and David L. Richards. 2014. *The Cingranelli-Richards (CIRI) Human Rights Dataset*.
- Clark, Ann Marie. 2001. *Diplomacy of Conscience: Amnesty International and Changing Human Rights Norms*. Princeton: Princeton University Press.
- Cleary, Edward L. 1997. *The struggle for human rights in Latin America*. Greenwood Publishing Group.
- Cmiel, Kenneth. 2004. "The recent history of human rights." *The American Historical Review* 109(1):117–135.
- Cohen, Stanley. 1995. *Denial and acknowledgement: the impact of information about human rights violations*. Center for Human Rights, Hebrew University.

- Conrad, Courtenay Ryals. 2010. *Constrained at Home: The Effect of Domestic and International Institutions on Domestic Respect for Human Rights* PhD thesis The Florida State University.
- Conrad, Courtenay Ryals and Jacqueline H.R. DeMeritt. 2011. "Human Rights Advocacy and State Repression Substitutability.".
- Conrad, Courtenay Ryals and Will H. Moore. 2010. "What Stops the Torture?" *American Journal of Political Science* 54(2):459–476.
- Cooley, Alexander and James Ron. 2002. "The NGO scramble: Organizational insecurity and the political economy of transnational action." *International Security* 27(1):5–39.
- Cunningham, David. 2004. *There's Something Happening Here; The New Left, The Klan and FBI Counter Intelligence*. Berkeley: University of California Press.
- Davenport, Christian. 2007. "State repression and political order." *Annu. Rev. Polit. Sci.* 10:1–23.
- Davenport, Christian and Davis Armstrong. 2004. "Democracy and the Violation of Human Rights: A Statistical Analysis from 1976-1996." *American Journal of Political Science* 48(3):538–554.
- Davenport, Christian and Patrick Ball. 2002. "Views to a Kill Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995." *Journal of conflict resolution* 46(3):427–450.
- Davenport, Christian, Will H. Moore and David Armstrong. 2007. "The Puzzle of Abu Chraib: Are Democratic Institutions a Palliative or Panacea?".
- Davis, David R, Amanda Murdie and Coty Garnett Steinmetz. 2012. "“” Makers and Shapers”: Human Rights INGOs and Public Opinion." *Human Rights Quarterly* 34(1):199–224.
- DeMars, Williams E. 2005. *NGOs and Transnational Networks: Wild Cards in World Politics*. London: Pluto Press.
- DeMeritt, Jacqueline HR. 2012. "International Organizations and Government Killing: Does Naming and Shaming Save Lives?" *International Interactions* 38(5):597–621.
- Diehl, Paul F. 1991. "Geography and war: A review and assessment of the empirical literature." *International Interactions* 17(1):11–27.
- Donnelly, Jack. 2003. *Universal Human Rights in Theory and Practice*. Ithaca: Cornell University Press.

- Donnelly, Jack. 2007. "The relative universality of human rights." *Human Rights Quarterly* 29(2):281–306.
- Dudai, Ron. 2007. "A to Z of Abuses: State of the Art in global human rights monitoring." *Development and Change* 38(6):1255–1265.
- Dueck, Judith, Manuel Guzman and Bert Verstapen. 2001. *Micro-Thesauri: A Tool for Documenting Human Rights Violations*. Versoix: HURIDOCs.
- Duihua. 2012.
URL: <http://duihua.org/wp/>
- Farber, Daniel A. 2000. "Rights as signals." *Available at SSRN 254349*.
- Finkel, Jenny Rose and Christopher D Manning. 2009. Nested named entity recognition. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics pp. 141–150.
- Finkel, Jenny Rose, Trond Grenager and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics pp. 363–370.
- Finnemore, Martha and Kathryn Sikkink. 1998. "International norm dynamics and political change." *International organization* 52(04):887–917.
- Florini, Ann. 1996. "The Evolution of International Norms." *International Studies Quarterly* 40:363–389.
- Forsythe, David P. 2009. *Encyclopedia of Human Rights: Vol. 1*. Oxford University Press.
- Forsythe, David P. 2012a. *Human rights in international relations*. Cambridge University Press.
- Forsythe, David P. 2012b. *The Humanitarians: The International Committee of the Red Cross*. New York: Cambridge University Press.
- Franck, Thomas M. and H. Scott Fairley. 1980. "Procedural Due Process in Human rights Factfinding by International Agencies." *American Journal of International Law* 74:304–345.
- Frank, Donner. 1990. *Protectors of Privilege: Red Squads and Police Repression in Urban America*. Los Angeles: University of California Press.
- Franklin, James E. 2008. "Shame on You: The Impact of Human Rights Criticism on Political Repression in Latin America." *International Studies Quarterly* 52(1):187–211.

- Gail, Mitchell, Klaus Krickeberg, J Samet, Anastasios Tsiatis and Wing Wong. 2007. "Statistics for biology and health."
- Gaizauskas, Robert, George Demetriou, Peter J. Artymiuk and Peter Willett. 2003. "Protein structures and information extraction from biological texts: the PASTA system." *Bioinformatics* 19(1):135–143.
- Gerner, Deborah J and Philip A Schrodt. 1996. The Kansas event data system: a beginner's guide with an application to the study of media fatigue in the Palestinian intifada. In *Poster Session presented at the American Political Science Association meetings, San Francisco, August*.
- Ghosn, Faten, Glenn Palmer and Stuart Bremer. 2014. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21:133–154.
- Gibney, Mark and Matthew Dalton. 1996. "The political terror scale." *Policy Studies and Developing Nations* 4(1):73–84.
- Gochman, Charles S. 1991. "Interstate metrics: conceptualizing, operationalizing, and measuring the geographic proximity of states since the Congress of Vienna." *International Interactions* 17(1):93–112.
- Goldstein, Joshua S. 1992. "A conflict-cooperation scale for WEIS events data." *Journal of Conflict Resolution* 36(2):369–385.
- Goldstein, Judith, Miles Kahler, Robert O Keohane and Anne-Marie Slaughter. 2000. "Introduction: Legalization and world politics." *International organization* 54(03):385–399.
- Goodliffe, Jay and Darren G Hawkins. 2006. "Explaining commitment: States and the convention against torture." *Journal of Politics* 68(2):358–371.
- Hafner-Burton, Emilie. 2008. "Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem." *International Organization* 62(4):689–716.
- Hafner-Burton, Emilie M. and Kiyouteru Tsutsui. 2005. "Human rights in a Globalizing World: The Paradox of Empty Promises." *American Journal of Sociology* 110(5):1373–1411.
- Haines, Herbert H. 2006. "Dangerous Issues and Public Identities: The Negotiation of Controversy in Two Movement Organizations." *Sociological Inquiry* 76(2):231–263.
- Hanisch, D., J. Fluck, H.T. Mevissen and R. Zimmer. 2003. "Playing biology's name game: identifying protein names in scientific text." *Proceedings of the 8th Pacific Symposium on Biocomputing*. pp. 403–414.

- Hathaway, Oona. 2002. "Do Human Rights Treaties Make a Difference?" *Yale Law Journal* 111:1935–2042.
- Hathaway, Oona. 2004. The Promise and Limits of the International Law of Torture. In *In Torture*, ed. Sanford Levinson. New York: Oxford University Press pp. 199–212.
- Hathaway, Oona. 2007. "Why Do Countries Commit to Human Rights Treaties?" *Journal of Conflict Resolution* 51(4):588–621.
- Hawkins, Darren G. 2002. *International Human Rights and Authoritarian Rule in Chile*. Lincoln: University of Nebraska Press.
- Hegre, Håvard. 2001. Toward a democratic civil peace? Democracy, political change, and civil war, 1816–1992. In *American Political Science Association*. Vol. 95 Cambridge Univ Press pp. 33–48.
- Hendrix, Cullen S and Wendy H Wong. 2013. "When is the pen truly mighty? regime type and the efficacy of naming and shaming in curbing human rights abuses." *British Journal of Political Science* 43(03):651–672.
- Herman, Burt. 2004. "Death of Uzbek Prisoner Prompts Unusual Investigation: Man's family says he was tortured." .
- Hertel, Shareen. 2006. "New Moves in Transnational Advocacy: Getting Labor and Economic Rights on the Agenda in Unexpected Ways." *Global Governance* 12:263–281.
- Hilbe, Joseph M. 2011. *Negative binomial regression*. Cambridge University Press.
- Hollyer, James R and B Peter Rosendorff. 2011. "Why do authoritarian regimes sign the convention against torture? Signaling, domestic politics and non-compliance." *Signaling, Domestic Politics and Non-Compliance (June 1, 2011)* .
- Hopgood, Stephen. 2006. "Keepers of the Flame." *Understanding Amnesty International. Cornell U. P. Ithaca–London* .
- Hornig, Susanna, Julie Templin and Lynne Walters. 1991. *Voices in the news: Newspaper coverage of Hurricane Hugo and the Loma Prieta earthquake*. National Emergency Training Center.
- HRW. 2004. "Crating Enemies of the State: Religious Persecution in Uzbekistan." .
- HRW. 2007. "Nowhere to Turn: Torture and Ill-treatment in Uzbekistan." .
- Huth, Paul K. 1996. "Enduring rivalries and territorial disputes, 1950-1990." *Conflict Management and Peace Science* 15(1):7–41.

- Hyde, Susan Dayton. 2006. *Observing Norms: Explaining the Causes and Consequences of Internationally Monitored Elections* PhD thesis University of California, San Diego CA: .
- ICRC. 2009. *Annual Report 2009*. Geneva, Switzerland: International Committee of The Red Cross.
- Jansche, Martin and Steven P Abney. 2002. Information extraction from voicemail transcripts. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing- Volume 10*. Association for Computational Linguistics pp. 320–327.
- Jetschke, Anja. 1999. Linking the Unlinkable? International Norms and Nationalism in Indonesia and Philippines. In *The Power of Human Rights: International Norms and Domestic Change*, ed. Steve C. Ropp and Kathryn Sikkink. Cambridge, UK: Cambridge University Press.
- Jurafsky, D. and J.H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall series in artificial intelligence Pearson Prentice Hall.
- Kam, Cindy D. 2007. “andRobert J. Franzese, Jr. 2007. Modeling andinterpreting interactive hypotheses in regression analysis.” *Ann Arbor: University of Michigan Press* .
- Keck, Margaret E. and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca: Cornell University Press.
- Kelly, Robert Edwin. 2005. *A Lot More than the NGOs Seem to Talk: The Impact of Non-Governmental Organizations on the Bretton Woods Institutions* PhD thesis The Ohio State University.
- Kelly, Tobias. 2009. “The UN Committee Against Torture: Human Rights Monitoring and the Legal Recognition of Cruelty.” *Human Rights Quarterly* 31(3):777–800.
- King, Gary and Margaret E Roberts. 2014. “How robust standard errors expose methodological problems they do not fix, and what to do about it.” *Political Analysis* p. mpu015.
- King, Gary and Will Lowe. 2003. “An Analysis Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design.” *International Organization* 44(2):341–355.
- KINU. 2000. *White Paper on Human Rights in North Korea*. Korea Institute for National Unification White Paper.
- Kocs, Stephen A. 1995. “Territorial disputes and interstate war, 1945–1987.” *The Journal of Politics* 57(01):159–175.

- Korey, William. 1998. *NGOs and the Universal Declaration of Human Rights*. New York: St. Martin's Press.
- Krain, Matthew. 2012. "Jaccuse! Does Naming and Shaming Perpetrators Reduce the Severity of Genocides or Politicides? 1." *International Studies Quarterly* 56(3):574–589.
- Krupka, George R and Kevin Hausman. 1998. Isoquest inc.: Description of the netowl (tm) extractor system as used for muc-7. In *Proceedings of MUC*. Vol. 7.
- Kudryashov, Andrei. 2004. "Independent Experts Do Not Confirm Detainee Andrei Shelkovenko's Death under Torture."
- Lafferty, John, Andrew McCallum and Fernando CN Pereira. 2001. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data."
- Landman, Todd. 2005. *Protecting Human rights: A Comparative Study*. Washington D.C.: Georgetown University Press.
- Leach, Philip, Costas Paraskeva and Gordana Uzelac. 2009. *International Human Rights and Fact-Finding: An Analysis of the fact-finding missions conducted by the European Commission and Court of Human Rights*. London: UK: Human rights and Social Justice Research Institute, London Metropolitan University.
- Lebovic, James H. and Erik Voeten. 2006. "The Politics of Shame: The Condemnation of Country Human Rights Practices in the UNCHR." *International Studies Quarterly* 50:861–888.
- Leskovec, Jure, Lars Backstrom and Jon Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM pp. 497–506.
- Liu, Xiaohua, Ming Zhou, Furu Wei, Zhongyang Fu and Xiangyang Zhou. 2012. Joint inference of named entity recognition and normalization for tweets. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics pp. 526–535.
- Long, J Scott and Jeremy Freese. 2006. *Regression models for categorical dependent variables using Stata*. Stata press.
- Lupu, Yonatan. 2013. "Best evidence: the role of information in domestic judicial enforcement of international human rights agreements." *International Organization* 67(03):469–503.
- Lutz, Ellen L. and Kathryn Sikkink. 2003. "The International Human rights Law and Practice in Latin America." *International Organization* 54(3):633–659.

- McCarthy, John D, Clark McPhail and Jackie Smith. 1996. "Images of protest: Dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991." *American sociological review* pp. 478–499.
- McPhail, Clark, David Schweingruber and John McCarthy. 1998. *In Policing Protest: The Control of Mass Demonstrations in Western Democracies*. Minneapolis: University of Minnesota Press.
- Meernik, James, Rosa Aloisi, Marsha Sowell and Angela Nichols. 2012. "The impact of human rights organizations on naming and shaming campaigns." *Journal of Conflict Resolution* 56(2):233–256.
- Mihr, Anja and Hans Peter Schmitz. 2007. "Human rights education (HRE) and transnational activism." *Human Rights Quarterly* 29(4):973–993.
- Mikhaylov, Slava, Michael Laver and Kenneth R Benoit. 2012. "Coder reliability and misclassification in the human coding of party manifestos." *Political Analysis* 20(1):78–91.
- Minkov, Einat, Richard C Wang and William W Cohen. 2005. Extracting personal names from email: Applying named entity recognition to informal text. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics pp. 443–450.
- Mosley, Layna and Saika Uno. 2007. "Racing to the Bottom or Climbing to the Top? Economic Globalization and Collective Labor Rights." *Comparative Political Studies* 40(8):923–948.
- Murdie, Amanda. 2009. Signaling Beyond Borders: The Conditional Impact of INGOs PhD thesis Emory University.
- Murdie, Amanda and David R. Davis. 2012. "Shaming and Blaming: Using Events Data to Assess the Impact of Human Rights INGOs." *International Studies Quarterly* 56(1):1–16.
- Murdie, Amanda and Dursun Peksen. 2014. "The Impact of Human Rights INGO Shaming on Humanitarian Interventions." *The Journal of Politics* 76(01):215–228.
- Murdie, Amanda M. and David R. Davis. 2011. "Shaming and Blaming: Using Events Data to Assess the Impact of Human Rights INGOs." *International Studies Quarterly* 56(1):1–16.
- Murdie, Amanda and Tavishi Bhasin. 2011. "Aiding and Abetting: Human Rights INGOs and Domestic Protest." *Journal of Conflict Resolution* 55(2):163–191.
- Nadelmann, Ethan A. 1990. "Global prohibition regimes: The evolution of norms in international society." *International Organization* 44(04):479–526.

- Nagler, Jonathan. 1991. "The effect of registration laws and education on US voter turnout." *The American Political Science Review* pp. 1393–1405.
- Naples-Mitchell, Joanna. 2011. "Perspectives of UN Special Rapporteurs on Their Role: Inherent Tensions and Unique Contributions to Human Rights." *The International Journal of Human Rights* 15(2):232–248.
- Neumayer, Eric. 2005. "Do International Human Rights Treaties Improve Respect for Human Rights?" *Journal of Conflict Resolution* 49(6):925–953.
- Neumayer, Eric. 2006. "Unequal access to foreign spaces: how states use visa restrictions to regulate mobility in a globalized world." *Transactions of the Institute of British Geographers* 31(1):72–84.
- Ng, Andrew Y. 2004. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*. ACM p. 78.
- NHRCK. 2004. *2004 Report: National Human Rights Commission*. National Human Rights Commission of Korea.
- Nowak, Manfred. 2009. "Fact-Finding on Torture and Ill-Treatment and Conditions of Detention." *Journal of Human Rights Practice* 1(1):101–119.
- Orentlicher, Diane F. 1990. "Bearing Witness: The Art and Science of Human Rights Fact-Finding." *Harvard Human Rights Journal* 3:83–135.
- Pevehouse, Jon C. 2005. *Democracy from above: Regional organizations and democratization*. Cambridge University Press.
- Pieper, JE. 1998. "Moral opposition to authoritarian rule in Chile, 1973-90."
- Poe, Steven C and C Neal Tate. 1994. "Repression of human rights to personal integrity in the 1980s: a global analysis." *American Political Science Review* 88(04):853–872.
- Pollanen, Michael Sven. 2004. "Observing Forensic Pathologist's Report on the Second Autopsy of Andrey Shelkovenko."
- Prakash, Aseem and Mary Kay Gugerty. 2010. *Advocacy organizations and collective action*. Cambridge University Press.
- Price, Richard. 2003. "Transnational civil society and advocacy in world politics." *World Politics* 55(04):579–606.
- Primo, David M, Matthew L Jacobsmeier and Jeffrey Milyo. 2007. "Estimating the impact of state policies and institutions with mixed-level data." *State Politics & Policy Quarterly* 7(4):446–459.

- Princen, Thomas and Matthias Finger. 1994. *Environmental NGOs in world politics*. Routledge.
- Ramos, Howard, James Ron and Oskar N.T. Thoms. 2007. "Shaping the Northern Media's Human Rights Coverage, 1986-2000." *Journal of Peace Research* 44(4):385-406.
- Ratinov, Lev and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics pp. 147-155.
- Ratner, Steven R. and Jason S. Abrams. 2001. *Accountability for Human Rights Atrocities in International Law: Beyond the Nuremberg Legacy*. New York: Oxford University Press.
- Richards, David L, Ronald D Gelleny and David H Sacko. 2001. "Money with a mean streak? Foreign economic penetration and government respect for human rights in developing countries." *International Studies Quarterly* pp. 219-239.
- Risse, Thomas. 2002. *Transnational Actors and World Politics*. Thousand Oaks, CA: Sage Publications.
- Risse, Thomas and Kathryn Sikkink. 1999. The Socialization of International Human Rights Norms into Domestic Practices: Introduction. In *The Power of Human Rights*, ed. Stephen Ropp Thomas Risse and Kathryn Sikkink. Cambridge, UK: Cambridge University Press.
- Ron, James, Howard Ramos and Katheleen Rodgers. 2005. "Transnational Information Politics: Human Rights NGO Reporting, 1986-2000." *International Studies Quarterly* 49(3):557-587.
- Ropp, Stephen C and Kathryn Sikkink. 1999. *The power of human rights: international norms and domestic change*. Vol. 66 Cambridge University Press.
- Roth, Kenneth. 2004. "Defending Economic, Social and Cultural Rights: Practical Issues Faced by an International Human Rights Organization." *Human Rights Quarterly* 26(1):63-73.
- Salamon, Lester M, S Wojciech Sokolowski and Regina List. 2003. *Global civil society*. Johns Hopkins Center for Civil Society Studies.
- Sano, Y, Y Jeong, AC Acock and AM Zvonkovic. 2005. Working with count data: Practical demonstration of poisson, negative binomial, and zero-inflated regression models. In *35th Annual Theory Construction and Research Methodology Workshop-Proceedings, National Council of Family Relations*.
- Schmitz, Hans Peter. 2002. "From Lobbying to Shaming: The Evolution of Human Rights Activism Since the 1940s."

- Schneider, Volker. 2000. The Global Social Capital of Human Rights Movements: A Case Study on Amnesty International. In *Private Organisations in Global Politics*, ed. K. Ronit and Volker Schneider. London: Routledge pp. 146–164.
- Schrodt, Philip A. 1994. “Event data in foreign policy analysis.” *Foreign Policy Analysis: Continuity and Change*. Prentice-Hall pp. 145–166.
- Schrodt, Philip A and Christopher Donald. 1990. Machine coding of events data. In *International Studies Association meetings, Washington, DC*.
- Sikkink, Kathryn. 2007. *Mixed signals: US human rights policy and Latin America*. Cornell University Press.
- Simmons, Beth A. 2009. *Mobilizing for human rights: international law in domestic politics*. Cambridge University Press.
- Singh, Sameer, Dustin Hillard and Chris Leggetter. 2010. Minimally-supervised extraction of entities from text advertisements. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics pp. 73–81.
- Snow, Clyde Collins, Fredy Armando Peccerlli, Jose Samuel Susanavar, Alan G. Robinson and Jose Maria Najera Ochoa. 2008. Hidden in Plain Sight: X.X. Burials and the Desaparecidos in the Department of Guatemala, 1977-1986. In *Statistical Methods for Human Rights*, ed. Jana Asher, DAVID L. Banks and Fritz Scheuren. Springer.
- Snyder, Jack L. 2000. *From voting to violence: Democratization and nationalist conflict*. Norton New York.
- Sommer, Henrik and James R Scarritt. 1999. “The utility of reuters for events analysis in area studies: The case of Zambia-Zimbabwe interactions, 1982–1993.” *International Interactions* 25(1):29–59.
- Starr, Harvey and Benjamin A Most. 1976. “The substance and study of borders in international relations research.” *International Studies Quarterly* pp. 581–620.
- Stroup, Sarah S and Amanda Murdie. 2012. “Theres no place like home: Explaining international NGO advocacy.” *The Review of International Organizations* 7(4):425–448.
- Sweet, Alec Stone. 2000. *Governing with Judges: Constitutional Politics in Europe: Constitutional Politics in Europe*. Oxford University Press.
- Tjong Kim Sang, Erik F and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics pp. 142–147.

- UIA. 2011. *Yearbook of International Organizations*. Vol. 1-4 Munich, Germany: UIA and De Gruyter Saur.
- Vasquez, John A. 1995. "Why do neighbors fight? Proximity, interaction, or territoriality." *Journal of Peace Research* 32(3):277–293.
- Vreeland, James Raymond. 2008. "Political Institutions and Human Rights: Why Dictatorships Enter into the United Nations Convention Against Torture." *International Organization* 62(1):65.
- Weissbrodt, David and James McCarthy. 1982. Fact-Finding by Nongovernmental Organizations. In *International Law and Fact-Finding in the Field of Human rights*, ed. B.G. Ramcharan. Martinus Nijhoff pp. 186–230.
- Welch, Claude E. 2001. *NGOs and human rights: Promise and performance*. University of Pennsylvania Press.
- Winston, Morton. 2001. Assessing the Effectiveness of International Human Rights NGOs: Amnesty International. In *NGOs and Human Rights*, ed. Claude E. Welch Jr. University of Pennsylvania Press.
- Wiseberg, Laurie S. 1992. "Human rights nongovernmental organizations." *Human Rights in the World Community: Issues and Action* pp. 372–83.
- Wong, Wendy H. 2008. Centralizing Principles: How Amnesty International Shaped Human Rights Politics through its Transnational Network PhD thesis University of California, San Diego CA: .
- Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.
- Zhou, Guodong, Jie Zhang, Jian Su, Dan Shen and Chewlim Tan. 2004. "Recognizing names in biomedical texts: a machine learning approach." *Bioinformatics* 20(7):1178–1190.