Check for updates

Regular Article

Journal of Peace Research

Get the word out: Monitoring human rights reduces abuse

Journal of Peace Research
1–17
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00223433231164443
journals.sagepub.com/home/jpr
Sage

José Kaire

School of Politics and Global Studies, Arizona State University

Abstract

Does human rights advocacy make a difference? Many are skeptical. Studies often find that advocates have an impact only under limited circumstances. I argue that these underwhelming results are a by-product of an identification problem. Research so far has effectively focused on whether shaming campaigns reduce ongoing abuse. But such cases are only part of the story. Another big aspect of advocacy is preventing abuse from ever starting. We must then pay attention to the deterred, those who chose not to violate human rights because of the threat of shaming. These cases do not repress and are never shamed. They are easy to miss because they look the same as those who never considered abuse in the first place. However, identifying deterred cases is crucial for judging the effectiveness of advocacy. I argue that we can resolve this issue by focusing on the degree of human rights monitoring in a country. Doing so allows researchers to recognize those under the scrutiny of advocates, and hence those that could have been deterred even if they were never shamed. Once we do this analytical refocus it is easy to see the positive effect of human rights advocacy. Human rights monitoring reduces abuse, and it does so in most countries.

Keywords

human rights, monitoring, repression, shaming

No one gave the *Zapatistas* a shot. They were a small indigenous guerilla group trying to reform the powerful Mexican dictatorship. Critics appeared to be right at first. Government forces had the rebels surrounded just days after their insurrection in 1994. Many expected more violent reprisals to follow. Instead, the government chose restraint. Thanks to the burgeoning Internet, the *Zapatistas* managed to become a global topic of interest. The political costs of repression soared as the world turned to Mexico, forcing the government to compromise. Foreign monitoring – while it lasted – tempered the repressive nature of the regime (Martínez-Torres, 2001; Garrido & Halavais, 2013).

Cases like the one of the *Zapatistas* highlight a limitation in the human rights advocacy literature. Most of this research has asked whether shaming can force an actor committing repression to stop. Yet the Mexican case shows that the impact of human rights advocacy is not only to halt ongoing abuse through shaming. It can also prevent abuse from starting by monitoring governments. The Mexican regime did not stop repression after being shamed. It chose restraint because it could anticipate the political costs of abuse. This distinction is important because it suggests a bias in the literature. Studies that examine how governments respond to shaming miss cases where shaming never occurred because its threat was enough to prevent repression. This deterrent effect is easy to overlook, but it is a fundamental part of the effectiveness of human rights advocacy.

An exclusive focus on shaming undersells the impact of human rights promotion. The complication arises from an identification problem similar to the one in deterrence research (Schultz, 2001; Fearon, 2002; Drezner, 2003; Drury & Li, 2006). The chance of being shamed is public knowledge. Potential perpetrators can deduce it from the amount of monitoring they face. Those under the spotlight of the international community, who cannot control the flow of information, and who face strong human rights activism, know that the

Corresponding author:
kaire@asu.edu

chance of shaming is high. Some potential perpetrators may be deterred by this looming threat of shaming, but it is difficult to distinguish them from those who never considered repression in the first place. The identification problem is that, for researchers focused on shaming, the cases deterred by the anticipation of public opprobrium are observationally equivalent to those that are sincerely committed to human rights. They both refrain from abuse, and they both avoid shaming. This makes our inferences depend on the shameless, those who repress knowing that they will be shamed. Research has then underestimated the effect of human rights advocacy by giving undue weight to the most resolute actors (Danilovic, 2001; Signorino, 2002).

Indeed, recent research on shaming has not been optimistic about its effectiveness – usually restricting it to countries with specific political institutions (Hendrix & Wong, 2013; Escribà-Folch & Wright, 2015), a strong civil society (Meernik et al., 2012; Murdie & Davis, 2012), or foreign aid dependence (Franklin, 2008; Dietrich & Murdie, 2017). Others conclude that shaming is entirely ineffective (Hafner-Burton, 2008; Hill & Jones, 2014). However, these studies adopt the most demanding test. They focus on cases that have already accepted the costs of shame, while excluding the possibility that cases without shaming may be examples of the deterrent effect of advocacy. To get a more accurate perspective, we must correctly distinguish those who would have repressed if not for the threat of shaming from those sincerely committed to human rights.

I argue that switching analytical focus from shaming to monitoring allows researchers to sidestep this identification problem. I refer to monitoring as the amount of human rights information in a country and the infrastructure that produces it. Monitoring then serves as a form of general deterrence (Quackenbush, 2010). It seeks to prevent unwanted actions by putting potential perpetrators on notice (Hu & Conrad, 2020). Monitoring communicates the cost of abuse *ex ante* and thus influences the decision to repress before shaming takes place. By quantifying monitoring, we can recognize actors under pressure to respect human rights even if they were never shamed. Actors with low shaming but high monitoring can then still serve as evidence of the effectiveness of human rights advocacy, making it unnecessary to focus on the deterrence of the most committed perpetrators.

This article addresses two challenges in the study of monitoring. The first obstacle is a lack of data. Measures of monitoring are limited. I present a new measure based on text analysis that covers almost every country for the last 40 years. The second challenge is that monitoring is a cause and a consequence of human rights. Monitoring prevents abuse, but human rights conditions also shape the decision of who gets monitored. I use vector autoregression models to study this endogenous relationship. The results paint a complicated picture. Monitoring is a powerful guarantor of human rights. However, as human rights conditions improve, monitors shift their limited resources to other countries to try to maximize their impact. As monitoring decreases in the original set of countries abuse becomes likely again. Overall, the results show that the efficacy of the human rights regime depends more on its ability to threaten punishment than in carrying it out.

What is monitoring, and why is it different from shaming?

Human rights have become a normative standard. Public support for human rights has been increasing in rich and developing countries for some time (Ron et al., 2017). Today, almost every country adheres to an international human rights treaty, and many governments have domestic institutions enforcing those commitments (Koo & Ramirez, 2009; Hafner-Burton, 2013). A global infrastructure now documents human rights across the world. Even smaller or nondemocratic countries may have dozens of globally connected organizations reporting on perpetrators, and many previously unconcerned international actors have joined those efforts (Tsutsui & Smith, 2019). In the age of information and globalization, many political actors have made it their priority to document human rights across the globe.

Human rights advocates have adopted a multipronged strategy. Human Rights Watch, for example, explains that 'the core of our methodology is our ability to investigate, expose, and shame' (Roth, 2004: 67). These are not synonyms for the same approach, but distinct strategies. Previous work interested in the overall efficacy of human rights organizations (HROs) has often analyzed them in tandem (Keck & Sikkink, 1998). However, it is useful to recognize each of them as following distinct processes and goals. I use the word 'monitoring' to refer to the investigation of abuses. It refers to human rights information and the infrastructure that produces it, such as journalists and civil society organizations. It emphasizes fact-finding and is proactive in nature. It occurs preventively, even in the absence of abuse. It is distinct from naming (expose), which is a deliberate effort to increase awareness about specific human rights abuses that have occurred. While monitoring is a mechanism of surveillance, naming divulges its findings.

Monitoring is also distinct from shaming. The latter is the most salient but also the last tool advocates wield against perpetrators. Like previous work (Lebovic & Voeten, 2009; Meernik et al., 2012), I understand shaming as advocacy carried out in response to abuse and that aims to generate material and reputational costs for the perpetrator. Unlike monitoring, it is reactive in nature and has the explicit goal of changing the behavior of abusers. While monitoring documents human rights conditions, shaming campaigns build upon it to challenge and steer abusers. To do so, shaming leverages networks of powerful actors, sanctions and delegitimization campaigns. Shaming then implies an active and costly commitment by advocates. It is, therefore, not as pervasive as monitoring. A recent measure finds that only 7% of country-year observations are the target of shaming (Murdie & Davis, 2012). It is nevertheless, along with monitoring and naming, a key part of the repertoire of human rights advocates.

Importantly, monitoring, naming and shaming are not necessarily sequential steps. Human rights advocates may investigate abuses but avoid shaming campaigns that they believe will not be impactful. The costliness of shaming means that advocates are strategic when implementing it, only choosing to shame when doing so can have the most effect. A senior manager of Amnesty International (AI) considered that 'you can work all you like on Mauritania, but the press couldn't give a rat's ass' (Ron, Ramos & Rodgers, 2006: 27). A lack of political opportunities keeps AI from directing resources to actively shaming the Mauritanian government, but they still produce yearly monitoring reports on its abuses. Doing so allows AI to quickly pivot to a stronger shaming campaign if the right political conditions were to arise. Advocates then have an incentive to maintain their fact-finding efforts despite not having an immediate intention to shame perpetrators.

Monitoring and shaming also have distinct motivations. While the *expectation* of repression drives monitoring, shaming occurs because of *observed* abuse. This too can create a gap between levels of monitoring and shaming. Ron, Ramos & Rodgers (2005) collected data on background reports and press releases by AI. While the former are instruments of monitoring, the latter focus on shaming. They find a low correlation (.35) between the two. For example, they find monitoring and shaming were both present during Pinochet's most repressive years in Chile. As repression decreased, so did shaming. However, human rights advocates continued to expect abuses and redoubled their monitoring efforts. In other cases, monitoring occurs as a part of democratic accountability and does not lead to shaming in the absence of abuse. For example, in 1997, France had the same level of monitoring as Rwanda and Kenya but, unlike them, it was not the target of shaming (Ron, Ramos & Rodgers, 2005). Monitoring and shaming share the goal of protecting human rights, but their motivation and application varies.

International organizations similarly separate monitoring from shaming. The International Labor Organization, for example, maintains separate bodies to investigate and to shame perpetrators. A representative group of experts focuses on monitoring, while a partisan body of political appointees decides on shaming (Koliev & Lebovic, 2018). In the United Nations, shaming is partly the result of strategic considerations and is not always directed at the worst transgressors (Freedman, 2013; Hug, 2016; but see Lebovic & Voeten, 2006). Donor agencies likewise monitor human rights conditions of their beneficiaries, but they do not want to bring attention to the fact that they might have aided human rights transgressors. Hence, organizational incentives keep donors from shaming fund recipients (Swedlund, 2017). Across the board, the decisions to monitor and shame follow different logics and remain empirically distinct.

This is not to say that monitoring and shaming are not intrinsically linked. They are. Monitoring signals to potential perpetrators the probability of shaming. Their abuses are more likely to be picked up when there are networks of journalists and advocates looking for violations. This signal is imperfect, as other factors can keep monitoring from translating into shaming. Nevertheless, shaming campaigns need the evidence generated by monitoring agents to make their case. Monitoring makes shaming a possibility, and intense monitoring makes the threat of shaming more credible. Likewise, shaming gives monitoring its teeth. The deterrent effect of monitoring weakens without strong shaming.

The identification problem in the shaming literature

Monitoring makes the potential for shaming public knowledge, and gives possible perpetrators the ability to adapt to it. This creates an identification challenge for researchers. Two hypothetical examples illustrate the problem. Consider a first scenario where HROs signal through monitoring their intention to shame perpetrators. Assume that these perpetrators are undeterrable. These may be regimes that are too reliant on repression or that simply lack the means to stop it (Escribà-Folch, 2013; Kaire, 2019). After observing repression, HROs follow through and start shaming. However, because the

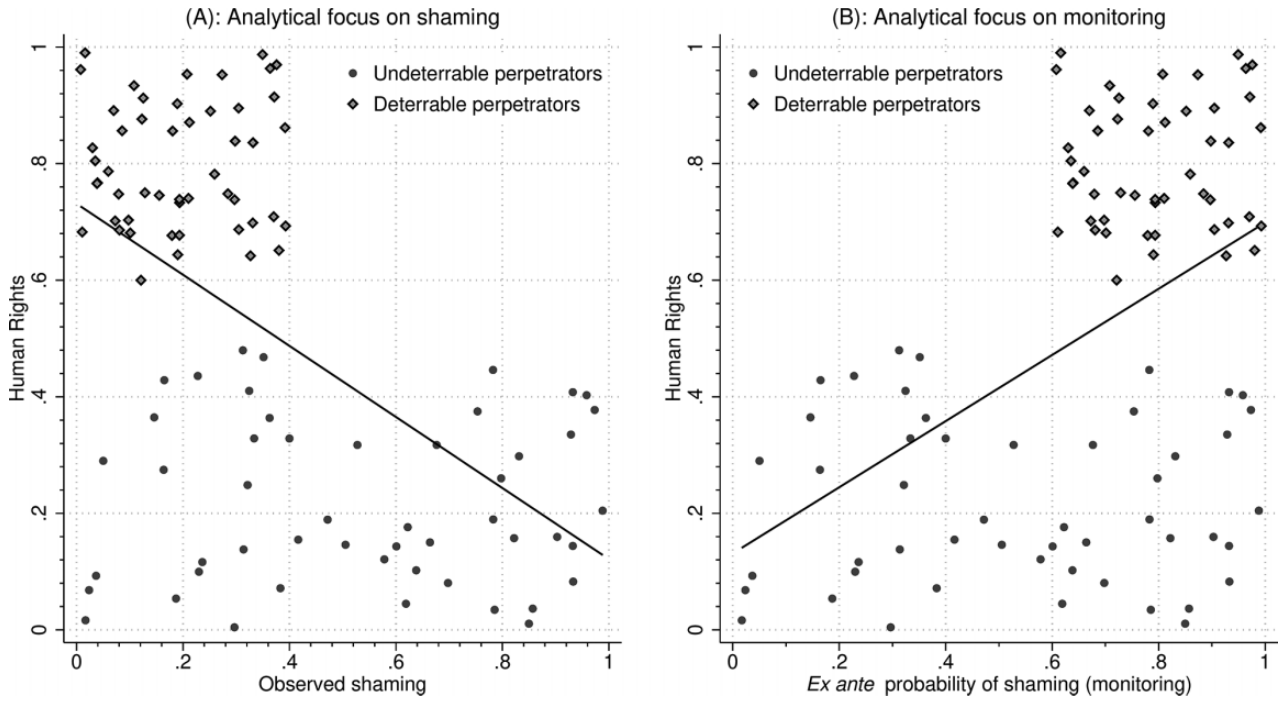

Figure 1. Expected effect of human rights advocacy by analytical focus

Figure shows hypothetical data linking human rights to (A) observed shaming and (B) monitoring by type of perpetrator. Panel A shows that a focus on observed shaming leads to a perceived negative relationship between shaming and human rights because deterred incumbents are counted as low shaming and high human rights. Importantly, these same cases can be characterized as facing high monitoring because monitoring and the occurrence of shaming are negatively correlated among deterrable types. Panel B shows that through the monitoring lenses, the same set of cases show a positive relationship between human rights and human rights advocacy.

perpetrators are undeterrable, repression continues. They had already accepted the cost of shaming, so its realization does not reveal any new information that would dissuade them. As these undeterrable types continue to repress, scholars would conclude that shaming is ineffective. Perpetrators are not always undeterrable, but in this case, the conclusion would be warranted.

Now consider a scenario where perpetrators are deterrable. This is where the identification problem arises. In these cases, the cost of shaming outweighs the perpetrators' utility of abuse. Like before, HROs use monitoring to signal their willingness to shame. Deterrable perpetrators observe this signal and conclude that repression is not worth the costs of shaming, deciding instead to respect human rights. Consequently, HROs do not shame. Scholars focused on realized forms of shaming would observe no shaming and no repression. This may lead them to conclude that shaming was unnecessary to protect human rights. However, in reality, it is the threat of shaming that is causing the absence of repression. If scholars focus only on whether shaming occurred, they will miss this deterring effect. Compare this with an alternative focus on monitoring. Under these lenses, researchers would capture the high *ex ante* probability of shaming when observing intense monitoring. This would make it possible to recognize these cases as responsive to HROs.

Figure 1 uses hypothetical data to depict how the identification problem can lead us to the wrong conclusion. It shows how the inferred relationship between advocacy and repression changes depending on whether we focus on observed shaming or the threat of shaming (i.e. monitoring). Undeterrable types always repress. This makes the interpretation of deterrable types crucial for inference. Deterrable cases do not repress widely if monitoring is high. The key distinction is whether we describe them as having low shaming or high monitoring. Perpetrators can anticipate the costs of shaming and be deterred by them. This strategic response induces a negative relationship between monitoring and observed shaming among deterrable types: with more monitoring there is less need for shaming. Therefore, the same set of cases can be described as either experiencing low shaming or high monitoring. Crucially, this distinction is not an empirical one, but one imposed by the observer. Panel A shows that if we opt for a focus on shaming,

then advocacy appears to damage human rights. Indeed, some studies have reached this conclusion (Snyder, 2020; Terechshenko et al., 2019). In contrast, Panel B shows that a focus on monitoring reveals the positive effect of advocacy on human rights. This difference occurs despite the underlying empirical process staying the same, the only change is how we categorize the cases.

In short, the identification problem is that by focusing on shaming, scholars cannot distinguish its deterrent effect from sincere adherence to human rights norms. We therefore miss many of the cases in which human rights advocacy was effective, making it appear less impactful. Cases that were never the target of shaming cannot be shown to be responding to advocates if we only pay attention to realized forms of shaming. We instead need to look at the off-equilibrium path to see the cases that could have repressed but chose not to because of the threat of public opprobrium. A focus on monitoring gives us a way to do this. Through it, we can distinguish between those under pressure to respect human rights, and those who are not. Hence, we can study how the former respond to the demand for human rights even if they were never shamed.

The monitoring game

This section formalizes the preceding argument. It shows that – in the presence of monitoring – further shaming campaigns are unlikely to induce better human rights. Importantly, it shows that monitoring is an effective deterrent because of the threat of shaming. Monitoring and shaming are both essential tools for advocates, but research must pay close attention to monitoring to better understand the effects of the latter. The monitoring game is different from previous models of deterrence (Fearon, 2002; Drezner, 2003) because the monitoring agent does have an incentive to carry out the punishment after detecting abuse. The model then shows a distinct structure under which the identification bias documented in previous work can emerge.

The monitoring game consists of an incumbent considering repression and a monitoring agent, call it an HRO, trying to dissuade the incumbent. The incumbent has a choice, $h \in \{0, 1\}$. They can comply with human rights norms $h = 1$ or violate them $h = 0$. When incumbents choose $h = 1$, they incur the loss associated with compliance $\rho \in (0, \infty)$. When an incumbent under monitoring choses repression, they suffer the costs of shaming $\sigma \in (0, \infty)$. The human rights group simultaneously decides whether to monitor the incumbent $m = 1$, or not $m = 0$. The HRO must pay $c > 0$ to implement monitoring but gains a benefit $\beta > 0$ if it targets a repressive incumbent. The utilities are:

$$u_{Incumbent}(h,\ m) = 1 - h(\rho - \sigma m)$$

$$u_{HRO}(h,\ m) = m(\beta - c - h\beta)$$

The payoffs for the players can be expressed as:

Incumbent \ HRO	$m = 0$	$m = 1$
$h = 0$	$1, 0$	$1, \beta - c$
$h = 1$	$1 - \rho, 0$	$1 - \rho + \sigma, -c$

Equilibrium analysis

The equilibrium for the monitoring game varies depending on how much the incumbent benefits from repression (ρ), the costs of shaming (c) and the relative value of monitoring to the HRO (β). When $\beta < c$, monitoring is not valuable to the HRO, and incumbents can choose to repress without worrying about shaming. When the benefit of monitoring outweighs its costs, $\beta > c$, incumbents will still choose repression if $\sigma < \rho$. These could be states facing heavy opposition (high ρ) or those so powerful that the relative costs of shame cannot dissuade them (low σ). While these undeterrable types cannot be shamed into respecting human rights, there will also be deterrable types for whom $\sigma > \rho$. When $m = 0$, incumbents of this type have an incentive to join undeterrables in repression. However, I show next that HRO can use monitoring to dissuade deterrables from imitating their undeterrable counterparts.

There is a unique Nash equilibrium in mixed strategies when dealing with deterrable types, $\sigma > \rho$, and when monitoring is valuable $\beta > c$. In such cases, the incumbents respect human rights with probability π, such that $\pi(\beta - c) - c(1 - \pi) = 0$. This gives a dominant strategy of choosing to violate human rights ($h = 0$) with probability $\pi^* = c/\beta$. Note that this equality is independent of σ, granted $\sigma > \rho$. This shows that – in the presence of monitoring – the sanctions that result from shaming do not induce additional compliance with human rights. When dealing with deterrable types ($\sigma > \rho$), the choice to repress depends on the extent of monitoring, not on the incumbents' cost of shaming. This counterintuitive result explains why the literature has struggled to identify a positive effect of shaming. It also gives the main hypothesis of this study: *human rights will increase when monitoring is prevalent* (i.e. $\frac{c}{\beta} \to 0$).

Shaming still plays a vital role in human rights advocacy. Shaming makes the amount of monitoring required to dissuade would-be perpetrators manageable. In the mixed strategy equilibrium, the HRO will try to maximize its impact while minimizing its costs. It chooses to monitor with just enough probability to dissuade the incumbent: $\phi(1-\rho)+(1-\phi)(1-\rho+\sigma)=1$, where ϕ is the probability of not monitoring. This means that the HRO can save on monitoring in proportion to the effectiveness of shaming $\phi^*=(\sigma-\rho)/\sigma$. As the international community and other factors drive the impact of shaming up, the HRO can reasonably expect that the incumbent will be less willing to repress, allowing it to scale back monitoring $\left(\frac{\mathrm{d}\phi^*}{\mathrm{d}\sigma}=\frac{\rho}{\sigma^2}\right)$. In addition to imbuing monitoring with a credible threat of sanctions, shaming makes monitoring a sustainable strategy for HROs with limited resources. I present a test of this secondary hypothesis below.

Monitoring as a tool to prevent abuse

The previous section showed that monitoring should protect human rights by making the costs of shaming public knowledge. However, the political consequences of monitoring remain underexplored in current research. Studies suggest that monitoring can positively affect state behavior in other areas, such as electoral competition and foreign aid transparency (Hyde, 2007; Donno, 2013; Honig & Weaver, 2019). Kelley & Simmons (2015) argue that monitoring is the first step in developing indicators that help the legal adoption of human rights principles. Subsequent research has focused on the soft power of monitoring agencies, particularly on how rankings can shape public discourse and elite norms (Merry, Davis & Kingsbury, 2015). The effect of monitoring in the absence of such rankings is still unclear.

Some qualitative studies make the connection between monitoring and human rights more explicit (Robinson, 2010; Wiseberg, 2017; Hu & Conrad, 2020). Research has also documented how perpetrators are mindful of monitoring. Oppressive governments often target the monitoring infrastructure before violating human rights in hopes of escaping punishment (Gohdes & Carey, 2017). Recently, the Belarusian government barred dozens of foreign journalists before deciding to repress ongoing protests (Shotter, 2020). India restricted Internet access before using deadly force against the protests that swept the country in 2019 (Findlay, 2019). Perpetrators try to shield themselves from the costs of abuse by undermining monitoring. This suggests that potential transgressors will hesitate to repress while monitoring remains in place. Indeed, some research suggests that abuse is less likely when monitoring agents, such as HROs, are active (Murdie & Davis, 2012).

The effect of monitoring is not a forgone conclusion. Sometimes repression occurs despite monitoring (Kent, 2001). Scholars cite the cases of Rwanda and the Former Yugoslavia as examples of monitoring failing to prevent atrocities (Steele & Amoureux, 2005; Straus, 2006). Incumbents can also take over the monitoring structure and subvert its role, using it for censorship or to justify abuses (Weidmann & Rød, 2019; Gohdes, 2020). Monitoring may also push states to enforce the human rights in areas that receive the most attention but disregarding them in others (Rejali, 2007; Bisbee et al., 2019). It is then important to systematically examine the effect of monitoring on abuse.

Measuring human rights monitoring

While scholars have provided several measures of naming and shaming, there are no cross-national measures of monitoring. Existing measures of press freedom or NGO density touch on the potential for monitoring, but do not quantify actual monitoring. A higher capacity for monitoring may remain unrealized if there is no expectation of abuse or under repressive spells. The potential for monitoring is then not a suitable replacement for actual monitoring. Considering these limitations, I develop an original measure of human rights monitoring. I present it here before proceeding to the analyses.

I use the US State Department reports on human rights to measure monitoring. While these reports are normally used to quantify repression, they can also be used to assess the extent of human rights monitoring. The reports offer comprehensive accounts of the presence of monitoring agents, the information they generate and their relationship with government. For example, the 2017 report on Venezuela references how 'a variety of independent domestic and international human rights groups generally operated with some government restrictions. Major domestic human rights NGOs conducted investigations and published their findings on human rights cases.' Importantly, mentions of monitoring occur independently of mentions of abuses. For example, the 2017 Costa Rica report notes that 'the Office of the UN High Commissioner for Refugees (UNHCR) and the government ombudsman monitored detention conditions, with UNHCR visiting monthly and the ombudsman preparing annual reports'. This is a key advantage since it allows us to distinguish between monitoring and violations of human rights. I expand on this point later in the measure validation section.

There are several other advantages to using the State Department reports over other possible sources of

information. The reports cover almost every country annually since 1977, excluding the United States, ensuring variation across time and space. The reports also aggregate the information produced by a diverse set of sources, including multilateral organizations, NGOs and journalists. This is an important strength because monitoring agents usually have political agendas that can affect the amount of coverage they give to a specific country. For example, in 2003 AI focused most of its resources to covering the human rights situation in Iraq. As a consequence, other countries saw a decrease in coverage by this organization of four times the normal variation (as measured by Ron, Ramos & Rodgers, 2005). By using a variety of sources, the reports are less sensitive to the strategic decisions of individual monitoring agents.

While there is a concern that there is bias in the State Department reports, researchers have not found systematic bias (Poe, Carey & Vazquez, 2001; Bagozzi & Berliner, 2018). The incentives to misrepresent the levels of monitoring are likely not as strong as the incentives to misrepresent human rights conditions. Nevertheless, I explore this possibility later through validity checks. I also replicate the measure using AI yearly reports. The correlation between the two measures varies depending on the year, with an average of 0.63 (SD=.08). This represents a significant correlation considering AI, by its own admission, tailors its coverage to complement its advocacy agenda.

Methodology

I use text analysis to quantify the monitoring described in the State Department reports. They cover the period from 1977 to 2017 for most countries. I use computational methods to process the resulting 7,318 documents. There are a variety of quantitative text analysis methods to choose from. I opt for a dictionary-based approach because it best serves the complexity of the task[1] and the interpretability of the results. This method consists of creating a list (i.e. dictionary) of keywords that signal the prevalence of monitoring. For example, 'journalists' and 'reports' are words that usually indicate a higher level of monitoring. My coding simply counts the number of times these keywords appear in a document. This count constitutes the core of the measure. The appendix shows the complete list of keywords.

A key step in this process is building the dictionary. I opt for a dictionary of general terms. I include two families of terms, one for the presence of monitoring agents (e.g. journalist, human rights group), and another one to capture the amount of actual information being produced (e.g. report, document). The general nature of these terms allows them to pick up on many instances of monitoring mentioned in the reports. However, this high sensitivity also means that they have a tendency to produce false positives. The reports may refer to monitoring that is no longer in place, or two keywords may be used for the same event. I account for some of this by excluding certain combinations of words (i.e. 'no reports') but also through an error-in-variables measurement model.

I use Bayesian measurement models to correct the dictionary counts (McElreath, 2015). I treat false positives as an error in variables problem. The idea is to make a distinction between the true count of keywords in a document and the count recovered by the dictionary. The latter is assumed to have some error, but if the error can be modeled in some systematic way, then it can be corrected.

The first step in this process is to get an idea of how much error the dictionary is producing. I hand-coded 200 randomly chosen reports to identify the rate of false positives. The rate converged at an average of three false positives per document (SD=3.1). Two other coders completed the same process simultaneously. There was agreement 83% of the time. I revised cases with disagreement and adjusted the count of false positives accordingly. I then related the number of false positives to the characteristics of the document. The best predictor of number of false positives is document length ($r=0.83$, $R^2=0.69$). Documents had about one false positive for every 4,500 words. The error in variables model corrects for false positives based on this finding.[2] The resulting measure ranges from 0 to 256, with a mean of 26 and a standard deviation of 23.

[1] A variety of supervised and unsupervised latent space models would be a natural alternative, but these models tended to recover respect for human rights and not monitoring because of their unidimensional assumption.

[2] The measurement model is:

$$Dictionary\ count_i \sim N(Regularized\ count_i,\ \epsilon_i)$$

$$\epsilon_i = 1 + \beta * Word\ count_i$$

where the regularized count is the final, corrected measure, and the dictionary count is the one produced by the dictionary before correcting for false positives. The regularized count is a combination of the dictionary count and some error ϵ_i determined by the rate of false positives β per word count (McElreath, 2015).

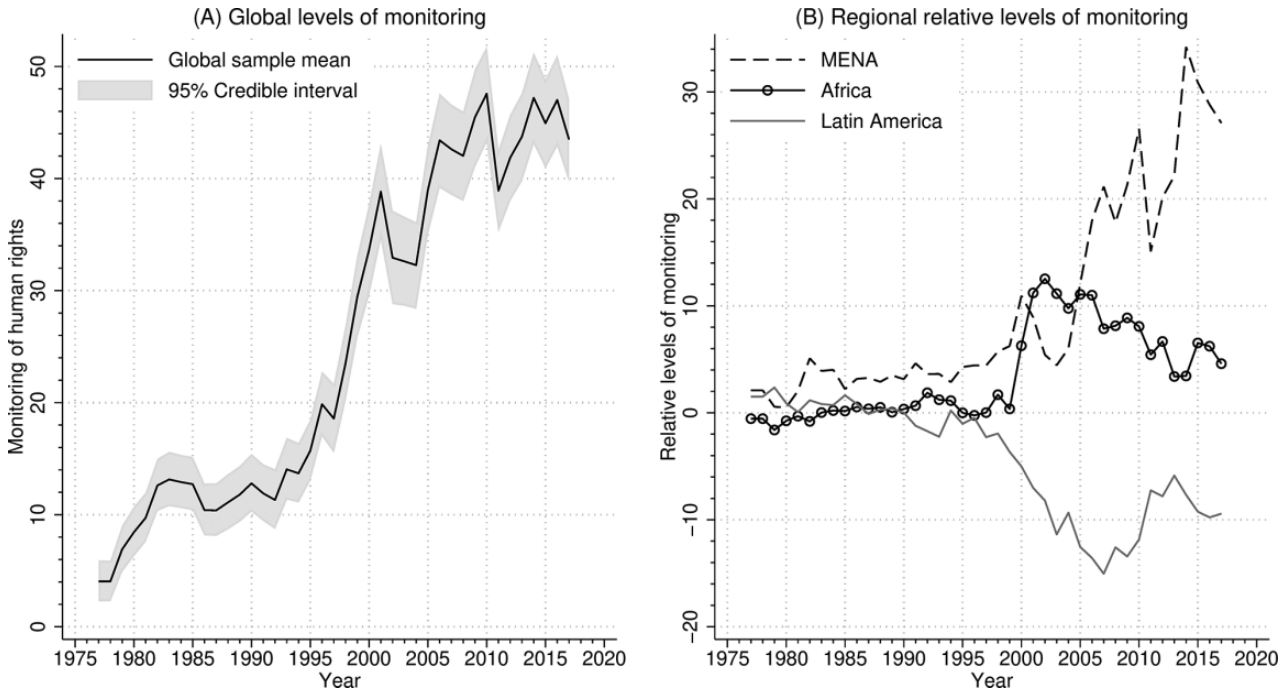

Figure 2. The evolution and regional distribution of human rights monitoring

Validity checks

I now assess the face validity of the resulting monitoring measure. Does it behave as one might expect? Qualitative accounts suggest that monitoring should increase with time (Simmons, 2009). In the 1980s international treaties such as the CAT, CEDAW and CRC went into effect, each buttressing the monitoring capacity of the international community. International donors have also increased their support for domestic human rights groups in the last 30 years. Panel A in Figure 2 plots the estimated global level of monitoring since 1977. It confirms a steady increase during the 1980s. It also shows rapid growth after 1993, the year the UN created the High Commissioner for Human Rights, and prominent donors such as the Open Society Foundations came into play. States also accelerated the adoption of international legal commitments during this year, strengthening the legitimacy and salience of human rights (Hafner-Burton & Tsutsui, 2005).

The second panel in Figure 2 compares monitoring in four regions. As one might expect, the MENA region has attracted the most attention in recent years. Africa saw a sharp increase in monitoring once international actors became involved in the conflicts of the DRC, Sierra Leone and Burundi in the early 2000s but not before. Recently, HROs in Mexico have noted how it is harder to find funding as donors have started to look elsewhere (Ron et al., 2018). The monitoring measure confirms that Latin America's share of monitoring has decreased as international donors have shifted priorities. Overall, the proposed measure mirrors what we know about the global evolution of human rights monitoring.

The monitoring measure also shows a good degree of discriminant validity. It remains distinct from the related, but separate, concepts of human rights and shaming. I rely on factor analyses to assess the distinctiveness of the variance in the monitoring measure. Table I shows the results. I examine the covariance of monitoring with other four variables. These include respect for human rights as measured by Fariss, Kenwick & Reuning (2020) and the civil liberties and physical integrity index by V-DEM. These two variables load into a resulting human rights factor that has an eigenvalue of 1.3. While the eigenvalue indicates a common latent quantity, monitoring does not load into it. Monitoring has a factor loading close to zero, and 90% of its variance is independent from the factor. The monitoring measure is not simply reflecting patterns of human rights.

The other two variables included in the factor analysis are the shaming measures by Murdie & Davis (2012) and Meernik et al. (2012). These two measures of shaming constitute the second factor. However, all its factor loadings fall short of the usual recommendation of 0.60. Combined with its eigenvalue of 0.5, the results indicate that the variables in the second factor are in fact tapping into different dimensions. Indeed, 79% of the variance

Table I. Factor analysis of monitoring, human rights and shaming

	Human rights factor		*Shaming factor*	
Variable	*Factor loading*	*Variance uniqueness*	*Factor loading*	*Variance uniqueness*
Respect for human rights	.75	.32		
Civil liberties	.82	.36		
Monitoring	-.05	.90	.42	.80
Shaming			.31	.90
AI urgent actions			.42	.79
Factor eigenvalue	1.3		0.5	

Variable sources: Respect for human (Fariss, Kenwick & Reuning, 2020), Civil liberties (Lindberg et al., 2014), Shaming (Murdie & Davis, 2012), AI urgent actions (Meernik et al., 2012).

in monitoring is independent from the shaming measures. This not only shows the discriminant validity of the measure but also confirms that it is possible to study the relationship between these three variables while avoiding circularity.

Statistical analyses

Estimation strategies and data

Identifying the effect of monitoring on human rights presents three problems. First, the average treatment effect may be unrepresentative. Research suggests that human rights reporting may be effective only under specific circumstances (Meernik et al., 2012; Escribà-Folch & Wright, 2015; Dietrich & Murdie, 2017). To account for this, I use hierarchical models with varying slopes to estimate the effect of monitoring for each country. This makes it easy to assess heterogeneity and to test if the effect of monitoring is limited to only a few cases.

Endogenity presents a second obstacle. Monitoring can be a cause and a consequence of human rights. Monitoring may improve human rights, but repression can also undermine or invite monitoring. I address this potential endogenity in two ways. First, I use change in human rights as the dependent variable and used lagged predictors. This allows for an imperfect but straightforward way to sidestep the possibility that *levels* of human rights correlate with monitoring. I also use a hierarchical vector autoregression (VAR) model for a more substantive exploration of this endogenous relationship (Freeman, Williams & Lin, 1989; Savage, 2016). This approach allows me to capture the long-term effects of monitoring on human rights, while also describing how current human rights shape future monitoring.

Finally, governments can hide their abuses from monitoring agents. Instead of reducing abuse, monitoring may push perpetrators to adopt coercive methods that are harder to detect (Rejali, 2007). I account for this strategy of coercive substitution to avoid overstating the effectiveness of monitoring. The final set of models follow the approach of Conrad, Hill & Moore (2018) to see if monitoring can curb stealth abuse. This is a two-step process. First, they use ill-treatment and torture data, which explicitly codes for torture designed to go undetected, such as sleep deprivation and waterboarding. Second, they use a negative binomial model that corrects for torture undercounting based on the probability of detecting stealth abuse. This approach provides a final estimate for monitoring adjusted by the probability of detecting abuse.

I use the preceding measure to estimate the impact of monitoring on human rights. I present results for the monitoring variable based on the State Department and AI reports. To ensure the robustness of the results, I use a variety of measures of human rights as dependent variables. I use the Latent Human Rights Protection Scores (Fariss, Kenwick & Reuning, 2020) as the first dependent variable. Fariss and colleagues aggregate a variety of human rights measures through an IRT model to estimate the overall level of physical integrity rights. It is worth noting that while this measure also uses the State Department reports as one of its 16 inputs, the different theoretical approach and coding procedures results in two distinct quantities, as shown in Table I. Nevertheless, as an alternative, the results also show the effect of monitoring on torture as operationalized by the ill-treatment and torture data (Conrad, Haglund & Moore, 2014). The appendix shows results using the civil liberties and physical integrity index by V-DEM, and the event-based measure by Bagozzi, Berliner & Welch (2021).

The following models also include shaming as measured by Murdie & Davis (2012). Its inclusion allows me to test whether the effect of monitoring is distinct from the one of shaming. I also control for potential

Table II. The effect of monitoring on human rights and types of torture

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
Dependent variable	*Change in human rights*	*Change in human rights*	*Change in human rights*	*Scarring torture*	*Stealth torture*
Monitoring (logged)		.48 (.14)†		-.98 (.33)*	-.98 (.40)*
AI monitoring (logged)			.39 (.07) †		
Shaming	-.14 (.14)	-.16 (.14)	-.16 (.14)	1.36 (.33)*	1.18 (.41)*
Level of democracy	.06 (.02) †	.05 (.02)	.08 (.02) †	- .13 (.14)	-.85 (.19)*
Number of protests	-.21 (.05) †	-.16 (.05) †	-.28 (.06) †	-.45 (.33)	-.42 (.45)
GDP	-.09 (.05)	-.04 (.06)	-.08 (.06)	.63 (.30)*	.85 (.45)
Population	-.15 (.11)	-.22 (.11)†	-.38 (.13) †	1.65 (.54)*	2.71 (.74)
Foreign investment	.07 (.09)	.08 (.08)	.08 (.09)	-.56 (.50)	-2.0 (.65)*
Number of observations	2,267	2,267	1,863	1,183	1,183
Number of groups	153	153	150	139	139

Models 1 to 3 report results of Bayesian multilevel models. Models 4 and 5 report results of random effects negative binomial models. Parentheses indicate standard errors. *p<.05. † less than 5% of posterior samples overlap with zero.

confounders. Levels of democracy, as measured by Polity, should affect the availability of monitoring and respect for human rights. Similarly, the number of protests should be associated with monitoring and levels of abuse, which could lead to confounding if not controlled for. For the same reasons I control for GDP and population size. These controls represent the typical set of covariates included in statistical analyses of human rights (Hill & Jones, 2014). Additionally, I include a time polynomial to guard against time trends and mean levels of monitoring within countries to prevent bias due to the exchangeability assumption of multilevel models (Bafumi & Gelman, 2006). Hierarchical models use Bayesian estimation.

Results

Table II presents the results of the analyses. Models 1 to 3 use hierarchical regression to explain changes in human rights. Models 4 and 5 use scarring and stealth torture as their dependent variable instead. For ease of interpretation, I only present the VAR result graphically. The results show a consistent picture: more monitoring means better human rights for most countries.

Figure 3 shows the results of Models 1 and 2. It mirrors the results foreshadowed earlier in Figure 1. Panel A finds a slight negative relationship between shaming and human rights, similar to the one other studies have reported. This does not necessarily mean that human rights advocacy is ineffective, however. It may also be that, as argued above, realized shaming is the wrong place to look for successful deterrence. In contrast, Panel B shows monitoring is associated with better human rights. A 10% increase in monitoring is associated with a 4.8% increase in human rights. That is equivalent to more than a two standard deviation change. This effect remains consistent when using AI reports (Model 3). Moreover, the positive impact of monitoring benefits most countries. In Figure 3, the gray lines represent the effect of the independent variable for each country. In the case of monitoring, this is positive for 84% of countries in the sample. By and large, monitoring is a strong safeguard against abuse.

Coercive substitution under monitoring

Can perpetrators turn to stealth abuse to escape monitoring? A comparison of scarring and stealth torture suggests that this is not the case. Monitoring reduces both types of abuse. The coefficient for monitoring in Models 4 and 5 is about the same, about an incidence rate ratio of 0.37. This means that a country that goes from no monitoring to the highest levels of monitoring in the sample would diminish torture by 63% (1-.37). Figure 4 shows the results in terms of predicted instances of torture. In both cases, more monitoring implies less torture. While scarring torture diminishes more in absolute terms, stealth torture diminishes at the same rate (63%). This suggests that monitoring prevents coercive substitution. Perpetrators cannot turn to stealth abuse to escape monitoring. The great effort HROs put on documenting abuse that is designed to go undetected seems to be paying off.

Dealing with endogenity

I now turn to the results of the vector autoregression model. The results so far have not substantially incorporated the endogenous relationship between monitoring

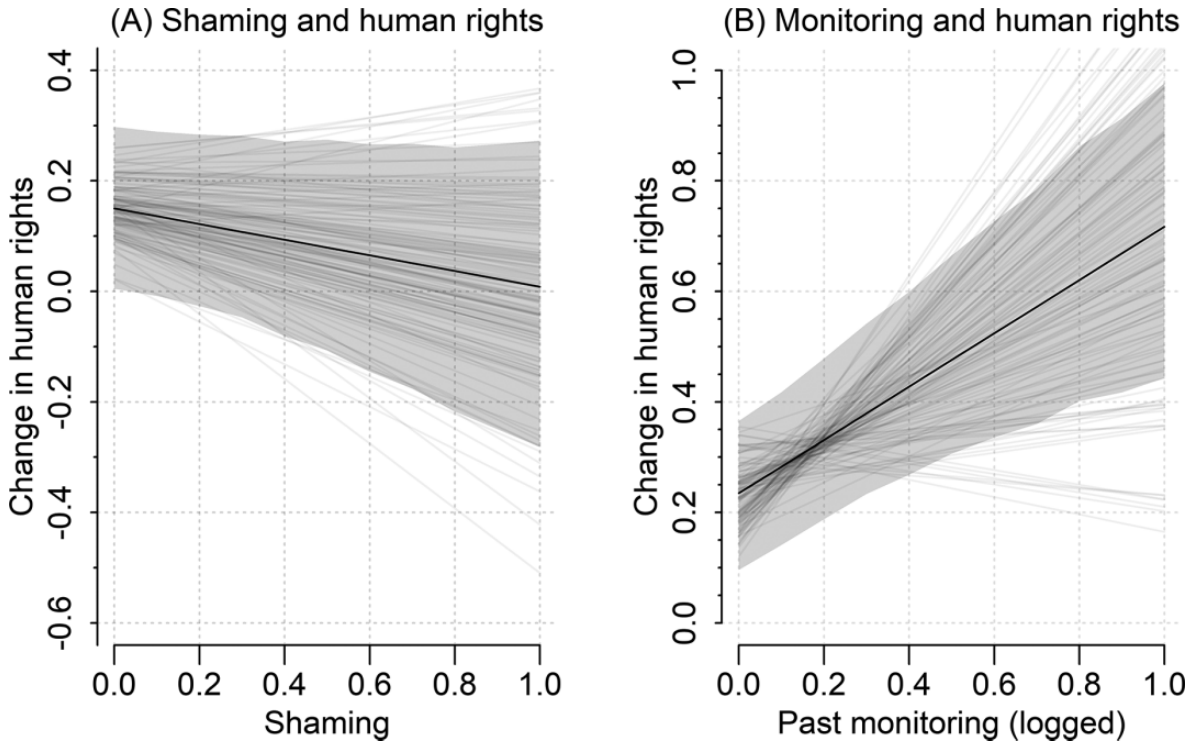

Figure 3. The impact of shaming and monitoring on human rights

Panel A and Panel B show results for Model 1 and Model 2, respectively. All variables are standardized to range from 0 to 1. Shaded areas depict 95% credible intervals, and the grey lines are country-specific effects calculated through multilevel estimation.

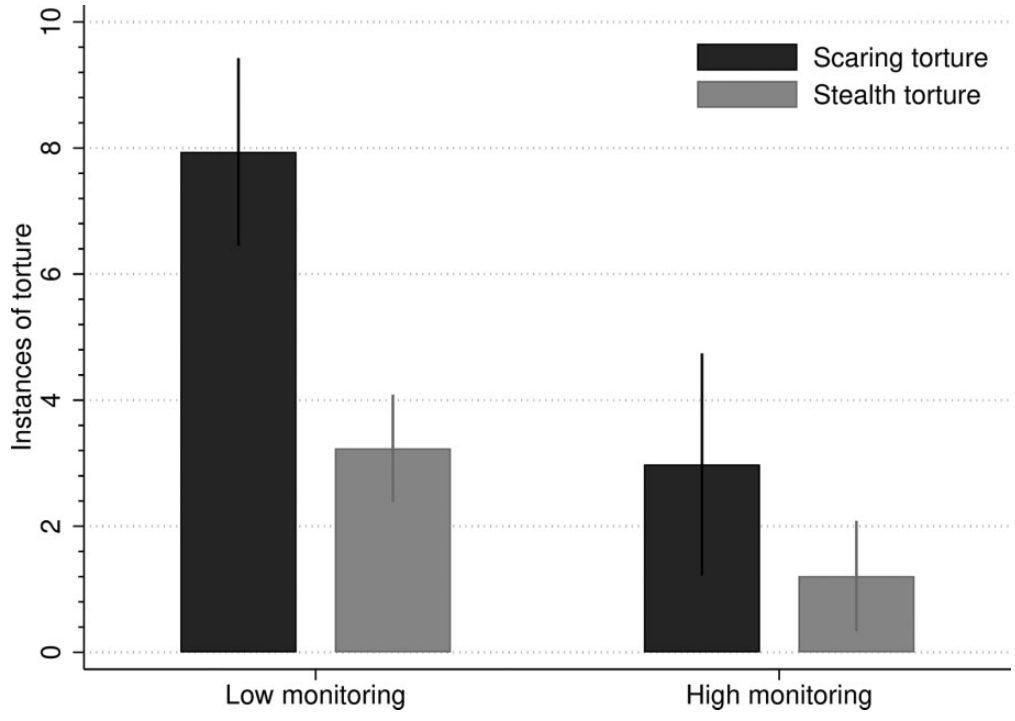

Figure 4. Types of torture under monitoring

Low and high monitoring refer to the 5th and 95th percentile of monitoring in the sample.

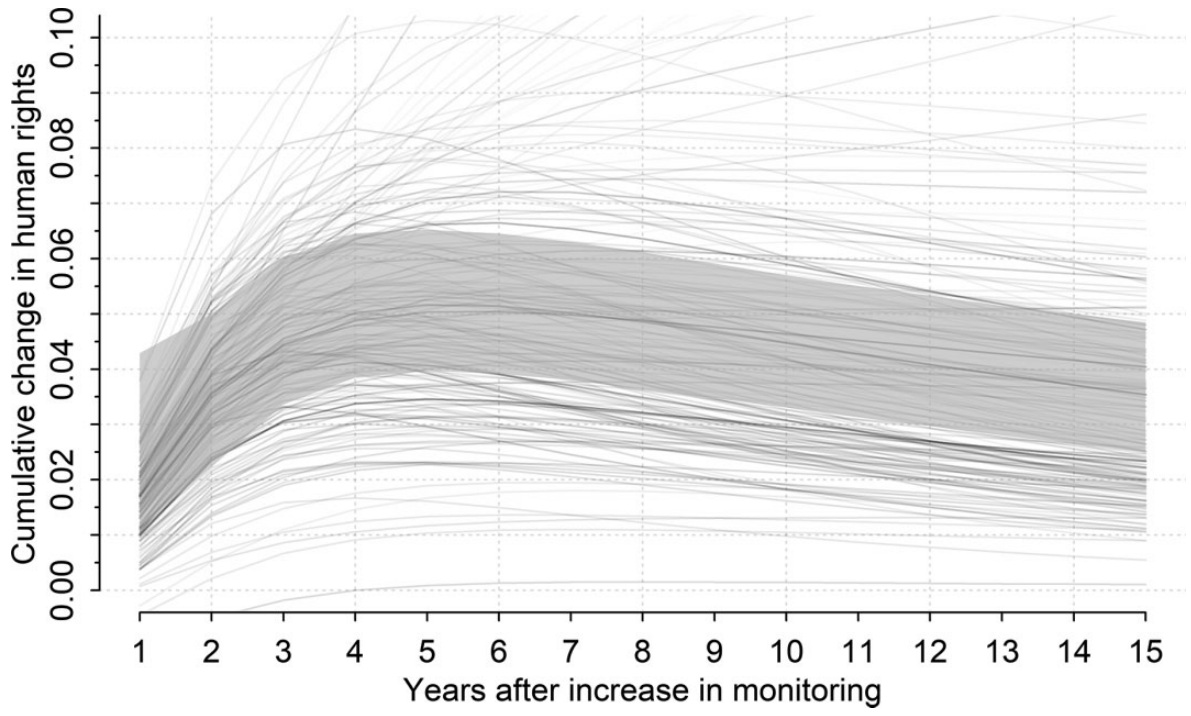

Figure 5. Response of the monitoring–human rights system to a shock to monitoring

Impulse response function of changes in human rights and an exogenous shock to monitoring. The *y*-axis shows the fractionally integrated changes in human rights scores, which have a 95th percentile of 0.35. Series are fractionally integrated through Robinson's Log-Periodogram regressions to allow for accurate estimation in the presence of unit roots. Shaded area indicates credible interval of the average effect. Gray lines show country-level effects.

and human rights. VAR models address this deficiency while providing useful theoretical insights. A limitation of the model is that since both variables are fractionally integrated to ensure statistical validity, it is difficult to interpret the coefficients. Instead, Figure 5 shows the long-term changes in human rights as a country receives an exogenous unit shock to monitoring. The results show that such an increase produces an immediate improvement in human rights. This is in line with the results of the previous models. I also find that the monitoring shock continues to have a lagged positive effect for the next few years.

The direct response to the monitoring shock is only part of the story. The system continues to react as it internalizes the change. Monitoring keeps increasing even after the original shock subsides. This is why Figure 5 shows that the cumulative effect of an increase in monitoring continues to increase for several years. Research has documented how it becomes easier for monitoring agents to continue their work once they have developed their infrastructure and networks in a country, encouraging them to continue their coverage even after human rights improve (Hafner-Burton and Ron 2013). The model indeed finds that a shock in monitoring leads to more monitoring down the line, which further reinforces future human rights.

Despite this reinforcing cycle, the effect of the shock tends to dissipate. The reason is that the model finds that human rights improvements reduce the monitoring a country receives. This could be consequence of international advocates and donors focusing their efforts on the worst offenders. While better human rights open the door for more monitoring in the short term, this is not enough to offset the loss of international attention as conditions improve over the long term. While the initial decision to implement monitoring can reach far into the future, it is necessary to maintain it to prevent subsequent abuses.

The need for shaming

The results so far should not be interpreted as evidence that shaming is a counterproductive strategy. While the models reported in Table II showed that shaming had a negative impact on human rights, this is not because it is incentivizing abuse. Instead, this negative impact arises from the fact that the threat of shaming, communicated

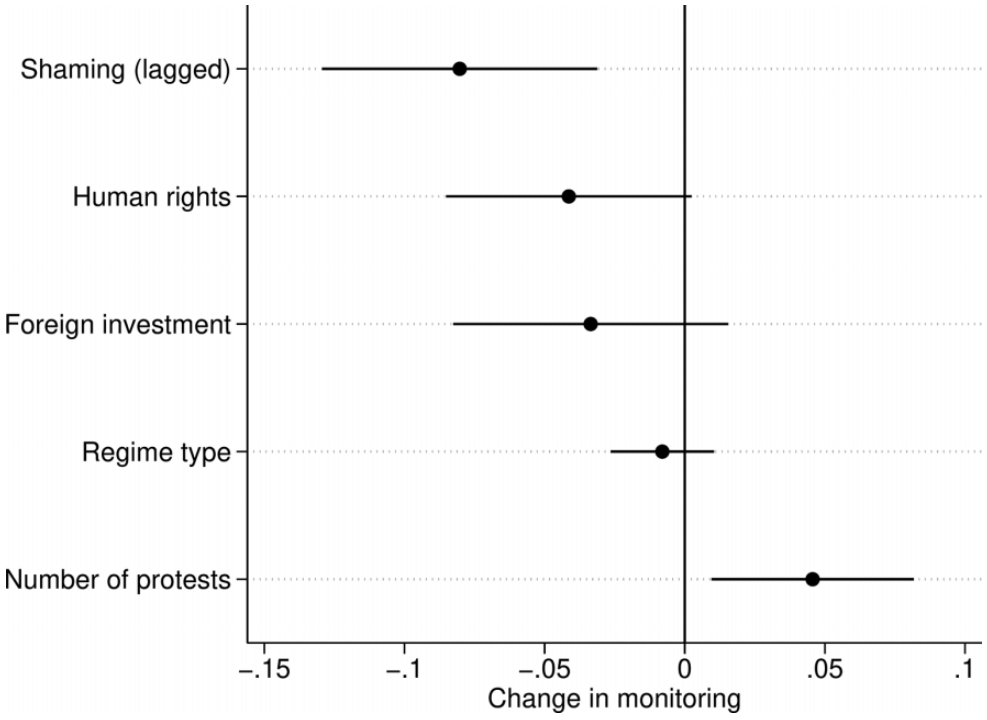

Figure 6. Determinants of monitoring

Random effects regression with 95% credible intervals. The dependent variable is yearly change in monitoring. All variables range from 0 to 1 and come from the sources discussed previously.

by monitoring, makes deterrable potential perpetrators respect human rights *before* shaming occurs. Moreover, the preceding monitoring game shows that shaming is integral to human rights advocacy. The amount of effort advocates need to invest on monitoring diminishes when the costs of shaming are high ($\phi^* = (\sigma - \rho)/\sigma$). Strong shaming allows advocates to monitor less without facilitating abuse.

Figure 6 shows the results of a final analysis testing this second prediction of the monitoring game. It shows the estimated effect of shaming on changes in monitoring. It shows that countries that have experienced shaming in the past receive less monitoring in the future. This evidence supports the strategic adaptation suggested by the monitoring game. Advocates can 'save' on monitoring because they know that perpetrators will be dissuaded by the higher costs of being caught. Shaming imbues monitoring with the threat of punishment, and thus makes it more efficient. The conclusion then should not be that monitoring is a preferable strategy to shaming, both are necessary for effective human rights promotion.

Conclusion

Monitoring human rights works, and it does so across a wide variety of contexts. That is the good news. Unfortunately, monitoring is also expensive and unevenly distributed. During the civil war in the Republic of Congo, international attention was instead set on the events in Cambodia. Similarly, the Salvadorean civil war received half of the attention of the neighboring Nicaraguan revolution despite being a more extensive conflict. These differences matter. Decisions about which stories to cover have a big impact on human rights. Countries with monitoring experience better human rights for long periods of time. In contrast, people in countries that receive little attention are at greater risk of facing abuse. In a world where human rights advocates operate under tight constraints, they must make difficult decisions about the issues and contexts they want to monitor.

Fortunately, the international community can help. It can do a lot by supplementing monitoring in countries where domestic conditions cannot sustain it. By subsidizing monitoring costs, international actors dissuade would-be perpetrators and help protect human rights. Another avenue of action is keeping the reputational costs of abuse high. Previous work has

focused on whether shaming improves human rights directly. This is an important question, but shaming can also be beneficial by making monitoring efficient. Less monitoring with a high threat of shaming may be as dissuading as intense monitoring with lower potential reputational costs. Indeed, the results indicate that past shaming allows advocates to invest less on monitoring. This also suggests that international shaming campaigns can help keep the repressive calculus in favor of human rights as governments continue to clamp down on monitoring agents.

Replication data

Replication materials are available at https://www.prio.org/jpr/datasets.

Acknowledgements

I would like to thank Aaron Hoffman, Amanda Murdie, John Patty, Victor Peskin, Tim Peterson, James Ron, David Samuels, Carie Steele and participants of the Northern Arizona University's research colloquium (25 March 2022). This article also greatly benefited from the feedback of two anonymous reviewers and the editor.

ORCID iD

José Kaire https://orcid.org/0000-0002-2435-9953

References

Bafumi, Joseph & Andrew Gelman (2006) Fitting multilevel models when predictors and group effects correlate. In: *Midwest Political Science Association Conference*. Chicago, IL.

Bagozzi, Benjamin E & Daniel Berliner (2018) The politics of scrutiny in human rights monitoring: Evidence from structural topic models of US State department human rights reports. *Political Science Research and Methods* 6(4): 661–677.

Bagozzi, Benjamin E; Daniel Berliner & Ryan M Welch (2021) The diversity of repression: Measuring state repressive repertoires with events data. *Journal of Peace Research* 58(5): 1126–11336.

Bisbee, James H; James R Hollyer, B Peter Rosendorff & James Raymond Vreeland (2019) The millennium development goals and education: Accountability and substitution in global assessment. *International Organization* 73(3): 547–578.

Conrad, Courtenay R; Daniel W Hill & Will Moore (2018) Torture and the limits of democratic institutions. *Journal of Peace Research* 55(1): 3–17.

Conrad, Courtenay R; Jillienne Haglund & Will Moore (2014) Torture allegations as events data: Introducing the Ill-Treatment and Torture (ITT) specific allegation data. *Journal of Peace Research* 51(3): 429–438.

Danilovic, Vesna (2001) Conceptual and selection bias issues in deterrence. *Journal of Conflict Resolution* 45(1): 97–125.

Dietrich, Simone & Amanda Murdie (2017) Human rights shaming through INGOs and foreign aid delivery. *Review of International Organizations* 12(1): 95–120.

Donno, Daniela (2013) *Defending Democratic Norms: International Actors and the Politics of Miscoonduct*. Oxford: Oxford University Press.

Drezner, Daniel W (2003) The hidden hand of economic coercion. *International Organization* 57(3): 643–659.

Drury, Cooper A & Yitan Li (2006) US economic sanction threats against China: Failing to leverage better human rights. *Foreign Policy Analysis* 2(4): 307–324.

Escribà-Folch, Abel (2013) Repression, political threats, and survival under autocracy. *International Political Science Review* 34(5): 543–560.

Escribà-Folch, Abel & Joseph Wright (2015) *Foreign Pressure and the Politics of Autocratic Survival*. New York: Oxford University Press.

Fariss, Christopher J; Michael Kenwick & Kevin Reuning (2020) Latent human rights protection scores Version 4. *Harvard Dataverse* V2: UNF:6: QPg88sybNJyuljPY ph2OXQ== [fileUNF].

Fearon, James D (2002) Selection effects and deterrence. *International Interactions* 28(1): 5–29.

Findlay, Stephanie (2019) Internet cut off in parts of India as protests continue. *Financial Times* 27 December.

Franklin, James C (2008) Shame on you: The impact of human rights criticism on political repression in Latin America. *International Studies Quarterly* 52(1): 187–211.

Freedman, Rosa (2013) *The United Nations Human Rights Council: A Critique and Early Assessment*. Abingdon: Routledge.

Freeman, John R; John T Williams & Tse-min Lin (1989) Vector autoregression and the study of politics. *American Journal of Political Science* 33(4): 842–877.

Garrido, Maria & Alexander Halavais (2013) Mapping networks of support for the Zapatista movement: Applying social-networks analysis to study contemporary social movements. In: Martha McCaughey & Michael D Ayers (eds.) *Cyberactivism: Online Activism in Theory and Practice*. New York: Routledge pp. 165–184.

Gohdes, Anita R (2020) Repression technology: Internet accessibility and state violence. *American Journal of Political Science* 64(3): 488–503.

Gohdes, Anita R & Sabine C Carey (2017) Canaries in a coal-mine? What the killings of journalists tell us about future repression. *Journal of Peace Research* 54(2): 157–174.

Hafner-Burton, Emilie M (2008) Sticks and stones: Naming and shaming the human rights enforcement problem. *International Organization* 62(04): 689–716.

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

Hafner-Burton, Emilie M & Kiyoteru Tsutsui (2005) Human rights in a globalizing world: The paradox of empty promises. *American Journal of Sociology* 110(5): 1373–1411.

Hendrix, Cullen S & Wendy H Wong (2013) When is the pen truly mighty? Regime type and the efficacy of naming and shaming in curbing human rights abuses. *British Journal of Political Science* 43(3): 651–672.

Hill, Daniel W & Zachary M Jones (2014) An empirical evaluation of explanations for state repression. *American Political Science Review* 108(3): 661–687.

Honig, Dan & Catherine Weaver (2019) A race to the top? The aid transparency index and the social power of global performance indicators. *International Organization* 73(3): 579–610.

Hu, Shengkuo & Courtneay Conrad (2020) Monitoring via the courts: Judicial oversight and police violence in India. *International Studies Quarterly* 64(3): 699–709.

Hug, Simon (2016) Dealing with human rights in international organizations. *Journal of Human Rights* 15(1): 21–39.

Hyde, Susan (2007) The observer effect in international politics: Evidence from a natural experiment. *World Politics* 61(1): 37–63.

Kaire, José (2019) Compensating elites: How international demands for economic liberalization can lead to more repressive autocracies. *International Studies Quarterly* 63(2): 394–405.

Keck, Margaret & Kathryn Sikkink (1998) *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.

Kelley, Judith G & Beth A Simmons (2015) Politics by number: Indicators as social pressure in international relations. *American Journal of Political Science* 59(1): 55–70.

Kent, Ann (2001) States monitoring states: The United States, Australia, and China's human rights, 1990-2001. *Human Rights Quarterly* 23(3): 583–624.

Koliev, Faradj & James Lebovic (2018) Selecting for shame: The monitoring of workers' rights by the international labour organization, 1989 to 2011. *International Studies Quarterly* 62(2): 437–452.

Koo, JW & FO Ramirez (2009) National incorporation of global human rights: Worldwide expansion of national human rights institutions, 1966-2004. *Social Forces* 87(3): 1321–1353.

Lebovic, James H & Erik Voeten (2006) The politics of shame: The condemnation of country human rights practices in the UNCHR. *International Studies Quarterly* 50(4): 861–888.

Lebovic, James H & Erik Voeten (2009) The cost of shame: International organizations and foreign aid in the punishing of human rights violators. *Journal of Peace Research* 46(1): 79–97.

Lindberg, Staffan I; Michael Coppedge, John Gerring & Jan Teorell (2014) V-Dem: A new way to measure democracy. *Journal of Democracy* 25(3): 159–169.

Martínez-Torres, María Elena (2001) Civil Society, the Internet, and the Zapatistas. *Peace Review* 13(3): 347–355.

McElreath, Richard (2015) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: Chapman and Hall/CRC.

Meernik, James; Rosa Aloisi, Marsha Sowell & Angela Nichols (2012) The impact of human rights organizations on naming and shaming campaigns. *Journal of Conflict Resolution* 56(2): 233–256.

Merry, Sally Engle; Kevin E Davis & Benedict Kingsbury (2015) *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge: Cambridge University Press.

Murdie, Amanda M & 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.

Poe, Steven C; Sabine Carey & Tanya Vazquez (2001) How are these pictures different? A quantitative comparison of the US State Department and Amnesty International human rights reports, 1976–1995. *Human Rights Quarterly* 23(3): 650–677.

Quackenbush, Stephen L (2010) General deterrence and international conflict: Testing perfect deterrence theory. *International Interactions* 36(1): 60–85.

Rejali, Darius (2007) *Torture and Democracy*. Princeton, NJ: Princeton University Press.

Robinson, Geoffrey (2010) *'If You Leave Us Here, We Will Die': How Genocide Was Stopped in East Timor*. Princeton, NJ: Princeton University Press.

Ron, James; Howard Ramos & Kathleen Rodgers (2005) Transnational information politics: NGO human rights reporting, 1986-2000. *International Studies Quarterly*. 49(3): 557–587.

Ron, James; Howard Ramos & Kathleen Rodgers (2006) What shapes the West's human rights focus? *Contexts* 5(3): 22–28.

Ron, James; José Kaire, Archana Pandya & Andrea Martinez (2018) What will it take to reduce NGO dependence on foreign aid? In: César Rodríguez-Garavito & Krizna Gomez (eds.) *Rising to the Populist Challenge: A New Playbook for Human Rights Actors*. Bogotá: Antropos, 153–160.

Ron, James; Shanon Golden, David Crow & Archana Pandya (2017) *Taking Root: Human Rights and Public Opinion in the Global South*. New York: Oxford University Press.

Roth, Kenneth (2004) Defending economic, social and cultural rights: Practical issues faced by an international human rights organization. *Human Rights Quarterly* 24(26): 63–73.

Savage, James (2016) Hierarchical vector autoregression. *Modern Statistical Workflow* (https://rpubs.com/jimsavage/hierarchical_var).

Schultz, Kenneth A (2001) Looking for audience costs. *Journal of Conflict Resolution* 45(1): 32–60.

Shotter, James (2020) Belarus cracks down on foreign journalists ahead of protests. *Financial Times* 29 August.

Signorino, Curtis S (2002) Strategy and selection in international relations. *International Interactions* 28(1): 93–115.

Simmons, Beth A (2009) *Mobilizing for Human Rights: International Law in Domestic Politics*. 1 edition. Cambridge: Cambridge University Press.

Snyder, Jack (2020) Backlash against naming and shaming: The politics of status and emotion. *British Journal of Politics and International Relations* 22(4): 644–653.

Steele, Brent J & Jacque L Amoureux (2005) NGOs and monitoring genocide: The benefits and limits to human rights Panopticism. *Millennium: Journal of International Studies* 34(2): 403–432.

Straus, Scott (2006) Rwanda and Darfur: A comparative analysis. *Genocide Studies and Prevention* 1(1): 41–56.

Swedlund, Haley J (2017) Can foreign aid donors credibly threaten to suspend aid? Evidence from a cross-national survey of donor officials. *Review of International Political Economy* 24(3): 1–43.

Terechshenko, Zhanna; Charles Crabtree, Kristine Eck & Christopher J Fariss (2019) Evaluating the influence of international norms and shaming on state respect for rights: An audit experiment with foreign embassies. *International Interactions* 45(4): 720–735.

Tsutsui, Kiyoteru & Jackie Smith (2019) Human rights and social movements: From the boomerang attern to a sandwich effect. In: David Snow, Sarah Soule & Hans Kriesi (eds.) *The Wiley Blackwell Companion to Social Movements* Second edition. Oxford: Blackwell pp. 586–601.

Weidmann, Nils B & Espen Geelmuyden Rød (2019) *The Internet and Political Protest in Autocracies*. New York: Oxford University Press.

Wiseberg, Laurie (2017) The Role of Non-Governmental Organizations (NGOs) in the protection and enforcement of human rights. In: Janusz Symonides (ed.) *Human Rights: International Protection, Monitoring, Enforcement*. London; New York: Routledge, UNESCO, 347–365.

JOSÉ KAIRE, PhD in Political Science (University of Minnesota, 2020); Assistant Professor, Arizona State University (2020 –); current interests: human rights and authoritarian politics.

Appendix

Table III presents the complete list of terms (tokens) included in the dictionary used to build the monitoring measure. Excluded terms are either negations of included ones or those that, during the hand-coding process, produced a high number of false positives. Table IV presents replications using alternative measures of human rights.

Table III. List of tokens included and excluded from the dictionary

Included terms	*Included organizations*	*Excluded terms*
Auditor	Americas Watch	Any international attention
Blogger	Article 19	Any international groups
Bulletin	Asia Watch	Any international organizations
Civil society group	Avocats Sans	Any international support
Columnist	Carter Center	Any media
Commentator	Center for Economic and Social Rights	Attitude Regarding International
Communication	Committee Against Torture	International airport
Correspondent	Committee to Protect Journalists	International Child Abductions
Document	Commonwealth Human Rights Initiative	International financial institutions
Fact-finding	European Commission on Human Rights	International financial institutions
Global	European Convention on Human Rights	International humanitarian law
International	European Court of Human Rights	International in scope
Investigation	Freedom House	International Monetary Fund
Journalist	Helsinki Committee	International norms
Media	Helsinki Watch	International Parental Child
Monitor	HRW	International Religious Freedom
Newspaper	Human Rights Watch	International scope
NGO	IACHR	International standards

(continued)

Table III. (continued)

Included terms	*Included organizations*	*Excluded terms*
Nongovernmental	ICC	International trafficking
Observer	Inter-American Commission on Human Rights	No information
Press	Kennedy Memorial	No international attention
Publication	Norwegian Refugee Council	No international groups
Pundit	Reporters without Borders	No international organizations
Record	UN	No international support
Report	United Nations	No publications
Research		No reports
Reviewer		Not a member of the United Nations
Rights group		Press censure
Rights organization		
Statement		
Transcript		
Witness		

Table IV. Replications using alternative measures of human rights

	Model 1a	*Model 2a*
Dependent variable	*Changes in civil liberties index*	*Changes in human rights*
Monitoring (logged)	0.06 (0.01) *	0.11 (0.03)*
Shaming	-0.01 (0.03)	0.07 (0.06)
Level of democracy	0.02 (0.01) *	-0.01 (0.01)
Number of protests	-0.02 (0.01)	-0.01 (0.03)
GDP	-0.02 (0.01)*	0.01 (0.02)
Population	0.01 (0.01)	-0.04 (0.03)
Foreign investment	-0.01 (0.02)	0.00 (0.05)

All variables range from 0 to 1. *p<.05. N=2,281 and 1,741, respectively. Outcome for Model 1a is V-Dem's civil liberties and physical integrity index. Model 2a has Bagozzi, Berliner, Welch (2021) event measure of human rights as the outcome. Both variables are differentiated and are estimated with random effects models.