(Semi)independent Human Rights Advocates: National Prevention Mechanisms and Human Rights Monitoring and Reporting

Dagmar Heintze

October 29, 2025

Introduction

On October 22, 2020, Peter Roth, who is serving a life sentence in a German prison, was awarded 12.000 Euros of compensation to be paid by the German government by the European Court of Human Rights (ECtHR) for non-pecuniary damages he suffered while being repeatedly strip-searched by detention personnel before and after receiving visitors. The court found that the repeated, random searches violated Roth's human dignity and constituted a violation of the prohibition of degrading treatment according to Article 3 of the European Convention on Human Rights. The ECtHR's sentencing was in line with previous rulings on degrading treatment through strip searches, and in itself did not constitute a legal innovation. What is special about the Roth v. Germany case, however, is a piece of evidence the claimant and his legal representation submitted as evidence — a report by Germany's National Prevention Mechanism (NPM), detailing and criticizing the frequent occurrence of random strip searches in Straubing prison, the prison where Roth is being held. By submitting the NPM report, Roth was not only able to substantiate his claims of abuse, but to further demonstrated that an independent human rights monitoring body that was created by the German government in accordance with the Optional Protocol to one of the nine core United Nations human rights treaties[1], the Convention against Torture (OPCAT), found the practice of random searches to be violating detainees' human rights and dignity. This episode raises an intriguing question: If NPMs, as independent monitoring bodies, can detect, document, and publicize human rights violations in detention and their reports can serve as instruments to criticize

[1] The core human rights treaties are: The International Covenant on Civil and Political Rights (ICCPR), the Convention on Elimination of All Forms of Discrimination against Women (CEDAW), the Convention against Torture and Other Cruel, Inhumane, or Degrading Punishment (CAT), the International Covenant on the Elimination of All Forms of Racial Discrimination (CERD), the Convention on the Rights of Persons with Disabilities (CRPD), the International Convention for the Protection of All Persons from Enforced Disappearance (CED), the International Covenant on Economic, Social, and Cultural Rights (ICESCR), the Convention on the Rights of the Child (CRC), and the International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families (ICPRMW).

the governments that created them — why would OPCAT ratifying governments create NPMs even though they can expose the governments' own human rights abuses in detention?

Since the Universal Declaration of Human Rights in 1948, nine core UN human rights treaties have been signed and ratified by many countries around the world. These treaties define specific rights and obligations, and by signing and ratifying them, countries formally accept the obligation to guarantee and uphold these provisions. In addition to the treaties themselves, all treaty regimes have introduced one or more optional protocols, which impose additional obligations, such as complaint and inquiry procedures, on the countries that choose to sign and ratify them.[2] In addition to the complaint and inquiry procedure, the Convention Against Torture (CAT), provides for a third instrument that may further increase a ratifying country's human rights accountability. The OPCAT, adopted in 2002, calls for national monitoring bodies within the ratifying states, NPMs, to be created within one year after ratification, and introduced an additional treaty body, the Subcommittee on Prevention of Torture (SPT), which directly communicates with, assists, and advises the NPMs and is further granted the right to visit places of detention in the signatory states itself (OHCHR, 2024b). [3].

According to the OPCAT, the purpose of NPMs is to work towards the prevention of torture and other cruel, inhuman, or degrading treatment or punishment from occurring within the ratifier countries. To achieve this objective, NPMs are to be provided the right to conduct monitoring visits to places of detention by their government to be able to document existing human rights violations and improvements in human rights standards in the places they visit. [4] NPMs are to be designed to monitor independently and are expected to be provided access to all areas of the place of detention, to view records, and, ideally, to conduct private conversations with detained persons to assess whether torture, inhuman or degrading treatment, or punishment occurs within the visited institution. After each visit, NPMs are expected to compile a visit report on their findings, which is submitted to the government to engage in a constructive dialogue, with the NPM

[2] While complaint procedures permit individual victims of human rights abuse to directly initiate proceedings against their government at the UN treaty bodies through 'individual communications', a country's acceptance of the inquiry procedure permits the respective treaty bodies to inquire directly with the ratifier country's government following allegations of systematic human rights abuse within the country.

[3] Complaint procedures are available for all nine core treaty regimes, and inquiry procedures can be initiated by six of the UN treaty bodies: The Committee Against Torture (CAT), the Committee for the Elimination of Discrimination Against Women (CEDAW), the Committee on Enforced Disappearance (CED), the Committee on the Rights of Persons with Disabilities (CRPD), the Committee on Economic, Social and Cultural Rights (CESCR), and the Committee on the Rights of the Child (CRC) can all initiate inquiry procedures.

[4] Per the definition of the OPCAT, places of detention are places where persons are deprived of their liberty and cannot leave at will based on the decision of a judicial, administrative, or other authority (OPCAT, 2024).

making recommendations on which changes the government can implement to improve human rights standards and prevent human rights violations from occurring in the future. The government can respond to the NPM to clarify potential misunderstandings, describe a plan for action for improvements, or justify why it perceives the findings to be unjustified or established practices to be necessary. Within the OPCAT framework, NPMs are further expected to submit yearly reports on their visits and findings regarding the treatment of detained persons to the SPT (OHCHR, 2024a).

Considering that ratifying the OPCAT is voluntary, which countries choose to ratify the protocol and accept the obligation to create an additional, domestic human rights oversight body and why? How do ratifying countries that create an NPM differ from those that do not, and how do NPM design choices influence the institutions' operations? Why do some countries ratify but do not subsequently create an NPM And finally — does a country's interaction with the SPT and its willingness to publish SPT visit reports indicate whether it is committed to improve human rights standards in detention?

After the OPCAT was adopted by the UN General Assembly in 2002, parties to the CAT began ratifying the protocol (United Nations, OPCAT, 2024). Some of the initial ratifier countries began creating NPMs, with some of the institutions taking up monitoring and reporting even before the OPCAT's entry into force. After reaching the required minimum number of 20 accession or ratifier countries, the protocol entered into force in 2006 (United Nations, OPCAT, 2024). Despite the formal requirement to create an NPM within one year of their treaty accession, many of the 94 OPCAT state parties to date have delayed their NPM creation or assignment by several years, and some failed to create an NPM altogether until now. Among the treaty parties that installed an NPM, some countries created a new institution while others designated an existing institution as the country's NPM (OHCHR, 2024a). In line with their inconsistent creation, monitoring, and reporting standards, not all existing NPMs adhere to the treaty obligation of submitting annual reports to the SPT. Some NPMs submit their reports annually, others submit their reports infrequently but consistently over the years since NPM creation (United Nations, SPT, 2024), or submit infrequently or only a few reports over the years.

Previous scholarship has explored why countries ratify international human rights agreements, as joining the treaty regime is voluntary and imposes binding human rights provisions on the ratifier country. This research has argued that states may ratify human rights treaties to signal human rights compliance to the international community, with some countries fulfilling treaty requirements even before ratification and ratifying as a formality (Von Stein, 2005; von Stein, 2016), while others use treaty ratification as a means to show their transition to improved democratic governance(Simmons, 2009; Hafner-Burton, 2013), or ratify as a sign of good will while not changing human rights practices (Goldsmith and Posner, 2005; Hill, 2010; Neumayer, 2005). Other research has assessed whether treaty ratification itself can improve human rights compliance, and has found mixed evidence of improved compliance (Downs and Jones, 2002; Hafner-Burton and Tsutsui, 2005; Hafner-Burton, 2008; Hathaway, 2002), while a government's ratifying of an optional protocol and accepting additional accountability measures appears to indicate its willingness to improve domestic human rights standards (Anaya-Muñoz and Murdie, 2021). Furthermore, scholars have explored how international naming and shaming (Hafner-Burton, 2008; Hendrix et al., 2013; Murdie and Davis, 2012; Keck and Sikkink, 2018), domestic human rights advocates (Risse et al., 1999; Keck and Sikkink, 1998; Simmons, 2009), and domestic human rights institutions (Cingranelli and Filippov, 2010; Cole and Ramirez, 2013; Welch et al., 2021) among other actors and factors can affect respect for human rights within human rights treaty state parties.

Despite the extensive bodies of research on human rights treaty accession and compliance, little previous scholarship explores under which conditions ratifier governments create imperfectly independent human rights agents, such as NPMs, and whether their design and working methods can influence their likelihood of driving changes in human rights respect within the ratifier country. Building on previous scholarship, a country's OPCAT ratification and NPM creation or designation can hold the potential to contribute to domestic policy changes that improve human rights practices in detention, thereby improving the respective country's human rights treaty compliance. Through their monitoring visits, NPMs are likely to gain a more detailed understanding of human rights standards in detention and can identify persistent human rights abuses at the location where they occur. Consequently, NPM reports on their visits to places of detention hold the potential to permit a more detailed understanding of human rights respect or the lack thereof within the countries that created them. To date, there exists limited research on the role NPMs play in a country's human rights compliance, and existing scholarship relies primarily on survey methods approaches focusing on a selected number of countries (Carver and Handley, 2016, 2020).

Whether NPMs can help improve human rights standards within a country will, however, be dependent on whether the institution is empowered by the government to carry out its monitoring and reporting task, thereby providing reliable and detailed evidence of human rights respect in detention. A government's choice to create a powerful or a restrained NPM will consequently condition the respective institution's ability to carry out its tasks. Despite the extensive bodies of human rights research, scholars still lack understanding of which states create empowered or incapacitated domestic human rights-promoting institutions, such as NPMs, and whether these NPMs can contribute to improved human rights standards in places of detention within the respective countries. Furthermore, little scholarship explores how the interplay between international organizations, domestic governments, and domestic institutions can indicate which countries will improve their human rights compliance. I begin to address these research gaps by assessing which factors drive NPM design choices and their subsequent monitoring and reporting frequency and depth, and whether domestic governments' interactions with the SPT and their willingness to publish SPT visit reports provide additional indicators of countries' likelihood to commit to improving human rights standards.

This article makes three main contributions. First, it assesses which domestic conditions and government preferences influence the likelihood of a government committing to create domestic human rights-promoting institutions after formally accepting the to do so after treaty ratification. Furthermore, this scholarship advances our understanding of when governments choose to create costly domestic human rights promoting agents and how national and international interactions influence this decision-making process. Second, this scholarship extends the existing body of research on domestic human rights institutions, such as National Human Rights Institutes (NHRIs), by focusing specifically on NPMs and their creation and monitoring and reporting power within the treaty parties. Third, this research explores the interactions between international regulatory bodies, ratifier governments, and domestic human rights-promoting institutions by assessing whether a country's choice to cooperate with and facilitate the SPT's country visit and publish the visit report can serve as an additional indicator of sincere commitment to human rights improvements.

Dynamics of human rights treaty compliance and sustained commitments

With this study, I assess the underlying domestic and international dynamics of NPM creation — specifically, which governments create NPMs and how a government's commitment to the OPCAT treaty obligations influences its NPM design choices. Furthermore, I evaluate whether a government's interaction with the UN treaty body serves as an indicator of its true intent to comply with the treaty provisions. In situating my theory, I build on previous research on (a) treaty ratification and compliance, (b) naming and shaming, and (c) domestic human rights promotion.

Treaty ratification and compliance

Previous research has assessed whether a country's accession to a human rights regime improves its domestic human rights conditions and has found mixed evidence. This scholarship has evaluated whether human rights treaty ratification can (a) itself improve human rights standards, or whether improved compliance (b) depends on the ratifying country's regime type.

When countries ratify an international human rights treaty, the ratification in itself appears to have no effect or may even lead to worsening human rights conditions within ratifier countries (Hafner-Burton, 2008; Hathaway, 2002; Hafner-Burton and Tsutsui, 2005). This finding raises the question whether human rights treaty ratification can improve human rights within the country or only serves as cheap talk to gain benefits within the international community (Downs and Jones, 2002; Hafner-Burton, 2013).

While treaty ratification alone does not lead to improved human rights standards in all ratifier countries, the regime type can be indicative of the likelihood of a country's human rights treaty compliance. For autocracies, the effect of treaty ratification on human rights respect appears to be low, with a country joining a human rights treaty regime potentially even resulting in worse human rights standards, depending on the country's level of citizen participation (Vreeland, 2008).

In semi-democracies, however, human rights treaty ratification can lead to improved human rights practices (Simmons, 2009). Countries transitioning to higher levels of democracy may ratify both as a representation of this transition and with the intent to signal increasing compliance with international norms (Hafner-Burton, 2013).

For democracies, treaty ratification can prompt domestic audiences to demand compliance (de Mesquita et al., 1999), or to punish non-compliance at the polls (Cordell, 2021), or through domestic courts (Powell and Staton, 2007), thereby leading to a higher likelihood of compliance with international human rights treaty regimes. Beyond enforcement mechanisms, democracies' higher institutional capacity (Cole, 2015; Anaya-Muñoz and Murdie, 2021), or previous domestic conditions that are conducive to compliance can result in high levels of treaty compliance, with countries joining human rights treaty regimes after a screening process (von Stein, 2005, 2016), and only if they intend to comply.

Naming and shaming and treaty compliance

While treaty ratification itself cannot reliably lead to improved human rights treaty compliance, naming and shaming, public criticism of non-compliant states by IOs, NGOs, or the international community, can contribute to improved human rights compliance. However, similar to the assessment of whether ratification leads to improved compliance, naming and shaming cannot reliably improve on human rights standards within all targeted countries.

While naming and shaming can worsen the human rights situation in repressive regimes (Hafner-Burton, 2008), democracies and hybrid regimes are commonly assumed to be more likely to improve their human rights standard when domestic and international actors pressure their governments for improved compliance. However, countries with a higher human rights standard appear to be less receptive to IO naming and shaming and less likely to change their human rights compliance than autocracies Hendrix et al. (2013), directly countering earlier arguments. While naming and shaming by NGOs can lead to improved human rights practices, it requires the presence of the NGOs within the country and the involvement of other actors such as third-party states, individuals, or organizations (Murdie and Davis, 2012; Keck and Sikkink, 2018; Risse et al., 1999; Jetschke and Liese, 2013). Domestic naming and shaming of the government by IOs, NGOs, or third-party states is expected to engage and motivate civil society actors to demand human rights compliance, providing them with knowledge about their government's human rights record (Risse et al., 1999; Hafner-Burton, 2013; Clark, 2013; McGaughey, 2021). However, IO naming and shaming can be counteracted by a contrary government narrative, thereby eliciting no effect or even prompting a backlash effect among constituents (Greenhill et al., 2022). When it comes to the specific actors that are naming and shaming non-compliant governments the sender matters, and some actors, such as the US State Department, are more successful in eliciting improved human rights compliance than others, such as Amnesty International (Zhou et al., 2022).

When considering specific mechanisms permitting the naming and shaming of non-compliant governments, government self-reporting and thereby providing a record of domestic human rights standards can lead to improved treaty compliance in ratifying countries when countries fulfill their reporting obligation and reports possess a certain level of depth and quality. All nine core human rights treaties provide for the system of self-reporting, requiring states to submit regular reports of human rights treaty compliance within the country (UN Treaty bodies, 2024). Governments are more likely to provide reports on their compliance when neighboring countries do so, and reporting quality depends on the regime type (Creamer and Simmons, 2015). When governments provide detailed reports, these reports engage domestic audiences and NGOs, thereby leading to increased pressure on the government to improve human rights standards (Creamer and Simmons, 2019, 2020; Risse et al., 1999). Consequently, while naming and shaming do not automatically improve human rights respect within the country, it can lead to better human rights respect in regimes that provide for some domestic participation.

Domestic human rights promotion

When it comes to the participation of domestic actors to improve human rights compliance, (a) NGOs can actively promote human rights in the ratifier countries and support the domestic civil society in its quest to improve human rights standards (Risse et al., 1999; Keck and Sikkink, 1998; Simmons, 2009), (b) the domestic judiciary can hold governments to account for human rights compliance (Dancy and Sikkink, 2012; Kim and Sikkink, 2010; Powell and Staton, 2007; Simmons, 2010), and (c) domestic institutions, such as NHRIs can contribute to improving human rights treaty compliance within ratifier countries (Cingranelli and Filippov, 2010; Cole and Ramirez, 2013; Welch, 2017, 2019; Welch et al., 2021).

Human rights NGOs can rally the opposition and incite a discourse on human rights standards within the country, thereby raising awareness about human rights issues and creating a more human-rights conducive environment (Risse et al., 1999). However, for this mechanism to lead to growing support for improved human rights respect, opposition groups must be perceived as peaceful and accepted as representatives of a movement to improve human rights (Keck and Sikkink, 1998; Jetschke and Liese, 2013). In more democratic regimes, NGOs can directly exert pressure on governments for better human rights compliance by raising awareness about existing human rights violations and becoming influential actors for change (Simmons, 2009).

Relying on a combination of normative pressure and deterrence, domestic courts can further contribute to improved human rights compliance by relying on a combination of domestic and international law to try human rights violators. However, not all human rights treaty provisions are equally relevant in domestic litigation, and the respective legal system influences the likelihood of domestic human rights litigation (Dancy and Sikkink, 2012). Furthermore, the timing of domestic human rights proceedings matters, with human rights trials after regime transition holding the potential to have a positive effect on respect for human rights within a country (Kim and Sikkink, 2010). In established democracies, effective judiciaries holding violators accountable can further improve human rights compliance (Powell and Staton, 2007; Simmons, 2010).

Government-created human rights bodies, such as National Human Rights Institutions (NHRIs), within a country can further lead to improved human rights compliance regardless of the regime type (Welch, 2019; Cole and Ramirez, 2013). While NHRIs typically rely on reports of abuses or support domestic human rights litigation, among other tasks, NPMs take on different roles by proactively visiting places of detention, interviewing detained persons, assessing medical records, and providing detailed reports on detention conditions within the ratifying countries (OPCAT, 2024). Consequently, NPMs focus on uncovering human rights abuses and working on improving human rights practices in detention without the need for a previous allegation.[5]

If a country's acceptance of the complaint and inquiry procedures can be a sign of its willingness to comply with human rights treaty provisions (Anaya-Muñoz and Murdie, 2021), a country's OP-CAT ratification can be assumed to signal a similar commitment. By creating an NPM, countries introduce a domestic agency for human rights promotion, which has the potential to exert pressure for improved compliance (Risse et al., 1999), and can provide evidence of ongoing human rights abuse to the SPT, thereby permitting naming and shaming (Risse et al., 1999; Hafner-Burton, 2013), or even prompting SPT monitoring visits. If NHRIs can improve respect for physical integrity rights, creating an NPM within an OPCAT-ratifying country can be assumed to be a state's additional signal to the international community that it intends to comply with human rights treaty provisions. The previous research suggests that a country's NPM creation can affect its human rights compliance, but for the institution to contribute to improvements in human rights standards, NPMs have to be empowered to take on the role of a domestic human rights promoter and to be able to conduct their monitoring and reporting work. In the next section, I build on the previous bodies of research on treaty compliance, naming and shaming, and domestic human rights promotion to assess how NPM composition and monitoring and reporting standards affect human rights compliance within NPM-creating countries.

[5] NHRIs were created after the adoption of the Paris Principles by the UN General Assembly in 1993 (Welch et al., 2021). They are domestic institutions that are founded by the ratifier states and are tasked with investigating human rights abuse allegations, integrating international into national law, advising the legislator on national legislation, interacting with the domestic public and international organizations, and educating the public about the existence of human rights (Welch et al., 2021). While the state created them, they are expected to be operating independently of state influence and can either be assumed to be a more costly signal of a state's willingness to comply with international human rights norms or serve as a mechanism for autocratic leaders to appease the public and secure tenure (Gandhi and Przeworski, 2007).

Argument

While countries may choose to introduce an NPM without requiring many substantial changes in their domestic human rights practices to achieve treaty compliance due to already high standards, other countries may face significant structural and reputational costs when doing so, because the NPM can highlight existing human rights abuses in places of detention.

High human rights standard countries (HRSCs) are likely to empower the NPM to conduct its monitoring and reporting tasks uninterrupted to achieve a considerable level of oversight over the current state of detention conditions. For these countries, creating an empowered NPM aligns with their policy preference of complying with international human rights provisions and can contribute to their maintaining a reputation of maintaining a high standard of international treaty compliance. While NPM-creation may introduce initial implementation costs due to domestic approval requirements, such as the passing of implementation legislation or required parliamentary approval for HRSCs, subsequent NPM monitoring and reporting can be expected to be relatively costless for implementing governments, because monitoring and reporting are unlikely to reveal severe human rights abuses, and the domestic costs of treaty compliance can be expected to be low. Consequently, HRSCs with high pre-ratification compliance can be assumed to incur initial NPM creation costs, while NPM monitoring and reporting are unlikely to elicit backlash from domestic audiences or opposition from competing political parties. After the initial implementation cost, HRSCs will benefit from assigning the detention standards-monitoring task to a specialized agent, the NPM. Furthermore, through their empowered NPM-creation, these countries can maintain a reputation of international treaty compliance, potentially benefiting them in negotiations with international partners, and can fulfill international agreement requirements (Hafner-Burton, 2013), among other benefits. NPMs in HRSCs are likely to consist of staff of human rights advocates who have a true interest in promoting and improving human rights. The NPM, consequently, may face the cost of occasional pushback when reporting on human rights abuses, but will mostly work in a cooperative environment with the government. For the NPM, maintaining a high visit and reporting standard provides the benefit of creating a basis for cooperation with the government and allows NPM staff to see the satisfaction of maintaining existing human rights standards and making progress toward additional improvements, which will outweigh the cost of occasional disagreements with the government.

Countries in which NPM creation or designation can prove costly for the government can be assumed to be ratifier states in which common practices in detention continue to violate human rights. In countries with low human rights standards (LRSCs), abuse may be widespread, and changing practices will require time and effort, may require holding abusive government forces to account, or require rebuilding and reforming the entire detention system. Creating an NPM and empowering it to conduct monitoring visits and to publish reports carries considerable risks for the governments, because reports of abuse can lead to domestic and international reputational costs and domestic backlash, both by the general public and by opposition parties. LRSCs that create an NPM face the uncertainty whether NPM monitoring and reporting can truly provide the intended benefits or will prove prohibitively costly once the NPM takes up its work. In these countries, governments may create an empowered or partially empowered NPM that is designed to serve their purpose of signaling compliance and contributing to human rights improvements in detention, but may be faced with the NPM developing its own, even stronger human rights-promoting agenda, resulting in the NPM publicizing and emphasizing abuse levels the government did not anticipate.

The NPMs in LRSCs are faced with a level of uncertainty that is comparable to the government's. While a human rights-promoting NPM may benefit from engaging in a critical, human rights-promoting dialogue with the government and see some level of improvement, developing a strong human rights agency can lead to significant costs for the NPM. Governments that perceive the NPM to be too costly may begin restricting the NPM's reach, implement institutional changes, or develop other strategies to rein in the NPM, which will make human rights promotion very costly or even prevent the NPM from conducting its work to its own standards altogether.

While the initial implementation cost of NPM creation may be lower in LRSCs than in high human rights standards countries due to lower domestic implementation requirements, countries with low human rights compliance can consequently expect to incur significantly higher costs if NPM monitoring and reporting publicizes ongoing human rights abuses in detention. NPM reports that document human rights abuses in detention and emphasize the frequency and severity of these abuses can prompt domestic audiences, such as the general public, to engage in protest or similar domestic opposition, thereby decreasing government approval rates or potentially jeopardizing future election outcomes. Furthermore, evidence of human rights abuses in detention may be utilized by domestic oppositions to weaken the government and change future election outcomes (Cordell et al., 2020). In addition to these costs imposed by domestic audiences, international partners may introduce additional costs for human rights-abusing countries. International trading partners may refrain from dealing with LRSCs, or international funding may be unavailable due to their lack of satisfactory human rights standards. A government creating an empowered NPM, therefore, faces considerable uncertainty about whether its treaty ratification and implementation benefits will outweigh the costs of the NPM publicizing ongoing human rights abuses. Creating an NPM, consequently, can prove very costly for low human rights standard countries, while the benefits may not be able to outweigh these costs.

LRSCs will design their NPM to fulfill their expected policy goals. Consequently, in countries in which the NPM is likely to report consistent abuse, the governments may (a) empower the NPM to conduct its monitoring and reporting tasks uninterrupted to achieve a considerable level of oversight over the current state of detention conditions, or (b) create NPMs that are ill-equipped to conduct frequent monitoring visits and may not report truthfully for fear of government repression to attempt to reduce domestic and international implementation costs. A third category of LRSCs may (c) take an extended period of time to create an NPM, or not create one at all. While the initial NPM design may align with government preferences at the time of the NPM designation or creation, preferences can change once governments gain a better understanding of the costs and benefits of NPM creation. Consequently, NPM empowerment levels can change after information updating, and NPMs may gain or lose some of their ability to independently monitor and report conditional on the perceived benefits of maintaining an independent and empowered NPM.

Governments are likely to create an initially empowered the NPM if they are willing and determined to improve their human rights standards and create an NPM to be a specialized agent to achieve this goal. Countries that do so perceive the increased transparency about domestic human rights standards as being beneficial to their long-term policy goals, as the implementation of a costly human rights monitoring institution signals progress towards improved compliance with international partners. Furthermore, the creation of an empowered NPM can fulfill the requirement of introducing accountability measures required for gaining access to benefits such as trade agreements, market access, EU accession, or similar benefits. For governments interested in gaining access to long-term benefits, the costs incurred by NPM creation and through NPM monitoring and reporting may be acceptable because the benefits of creating an empowered NPM warrant their acceptance. While these countries cannot be certain that their NPM-creation will result in the expected benefits outweighing the costs of introducing the institution and having it publicize human rights abuses in detention, these LRSCs will create an empowered or a partially empowered

NPM, and permit NPM monitoring and reporting to gain access to the domestic and international benefits of fulfilling their ratification requirement.

Empowered or partially empowered NPMs in LRSCs are likely to incur some costs of government criticism or push back when publishing findings of high levels of human rights abuses, but the benefits of engaging in a national and international human rights promoting dialogue and seeing some improvements of human rights standards can be expected to outweigh the costs. This cost-benefit consideration, however, relies on the assumption that the NPM does not develop a level of agency that proves prohibitively costly for the government, which could result in the government reconsidering its cost-benefit calculation and implementing more drastic restrictions on the NPM's monitoring work or even restructuring or dissolving the NPM altogether.

In cases in which the governments can expect to face reputational costs within the country and in the international community that are too costly to address, or if the governments are not capable to implement the required changes to achieve better human rights respect in detention, NPMs are likely to be designed to be ill-equipped to conduct frequent monitoring visits and may not report truthfully for fear of government repression to attempt to reduce domestic and international implementation costs. These countries will continue to seek the short to medium-term benefits of treaty ratification and implementation, such as the ability to acquire funds from development programs or trade agreements with international partner countries requiring human rights improvements. However, by creating a restrained or incapacitated NPM, these countries seek to minimize the costs of continued human rights abuse. While the initial NPM creation may be geared towards restraining the NPM's monitoring and reporting abilities, these governments cannot be certain that the NPM will stay within the defined bounds and not develop its own human rights-promoting agenda despite existing government restraints. While the NPM may consider promoting a human rights agenda to be too costly, it may also be willing to face the costs of shirking its initial responsibilities to push for stronger human rights. In cases in which the NPM develops its own agenda and works beyond its intended realm, NPM operations can be expected to prove prohibitively costly for the governments, resulting in additional restraint, restructuring, or even the dissolution of the NPM.

Governments that do not create an NPM after ratification can be expected to be facing prohibitively high costs preventing them from fulfilling their treaty obligation. These costs may be imposed by capacity constraints, preventing the government from setting up the institution, finding qualified personnel, or facing a fractured coalition with opposition parties preventing them from passing the laws necessary to create the institution. Furthermore, domestic instability, such as regime changes, conflict or war may stop governments from creating an NPM. Alternatively, countries may ratify the OPCAT without the intention to create an NPM, only seeking to acquire the short-time benefits of feigning human rights progress. While these conditions prevent governments from fulfilling their treaty obligation to create an NPM after ratification, changes in domestic conditions may result in a delayed NPM creation after government preferences and cost-benefit considerations change.

By creating an NPM, OPCAT-ratifying countries formally accept additional scrutiny regarding their treaty compliance by permitting a national monitoring body to visit places of detention and assess their adherence to the treaty's provisions. However, not all NPMs are monitoring and reporting at an equally frequent rate and in a comparable depth. While some countries appear to be equipping their NPMs with the personnel, resources, and access to places of detention required for effective monitoring, others seem to be creating NPMs that do not reach an equal level of monitoring and reporting effectiveness. Depending on the design of the NPM that countries create or designate, countries can be expected to pursue different goals through their creation, and the design of the NPMs that ratifier states create or assign, and their powers consequently may differ greatly based on the underlying motivation for their creation.

Building on the observation that NPM monitoring and reporting efficacy can differ greatly between countries, I argue that two categories of countries will create different types of NPMs, depending on the outcome they intend to achieve by creating the institution. (a) *Human Rights Champions* can be expected to empower their NPM and seek to improve efficiency in monitoring and reporting by creating a specialist body. For these countries, NPM creation is initially costly, but NPM monitoring and reporting will likely not create high ongoing costs but rather provide reputational and co-operational benefits. In contrast, (b) *Human Rights Contenders* can be expected to create an NPM as a strategic signal of increasing democratization and willingness to reform, seeking medium and long-term benefits through their NPM creation. In these countries, the initial implementation cost is likely lower, but NPM monitoring and reporting can be expected to impose higher, continuous costs on the governments. Depending on the governments' policy goals and the expected benefits of maintaining an empowered NPM, these countries will nonetheless permit monitoring and reporting if the benefits outweigh the costs. If the costs of implementation and NPM monitoring and reporting are larger than the expected long-term benefits, Human Rights Contender governments may create an immediately restrained NPM or reassess their NPMs' empowerment by implementing post-creation restraints if the NPM assumes too much independence or if government priorities change once the government gains insight into the expected costs and benefits of NPM operations. By restricting the NPM's power, these governments can gain the short-term to medium-term benefits while reducing the costs of implementation to a minimum. Table 1 provides an overview of the two types of NPM-creating countries.

Table 1: Typology of NPM Creation and Empowerment Based on Cost–Benefit Dynamics

Country Type	**Cost–Benefit Dynamics**	**NPM Power**	**Monitoring**	**Reporting**	**Policy Outcome**
Human Rights Champions	High initial implementation costs (e.g., legal set-up, legislation) but low ongoing costs. Benefits of empowered NPM (reputation, efficiency, credibility) outweigh risks, as violations are rare.	High	Frequent and regular	Detailed and consistent	Minor adjustments to maintain high standards
Human Rights Contenders	Lower initial implementation costs, but potentially high ongoing costs due to exposure of violations. Benefits (international reputation, democratization signals, accession incentives) must outweigh the political/reputational costs of transparency. Risk of ex-post restraint if costs rise.	Moderate–High (initially); may increase if benefits outweigh costs or decline if costs outweigh benefits	Medium frequency initially; may decline if costs outweigh benefits, or further increase over time if benefits outweigh costs	Mixed; depends on sustained commitment and whether benefits continue to outweigh costs	Substantive reforms possible if benefits outweigh costs

Hypotheses

Building on the argument, in this study I assess two sets of linked research questions. First, which political, institutional, and regime characteristics and which cost-benefit considerations determine whether a country creates an NPM that is empowered to frequently conduct monitoring visits and reports in detail? And second — can the government's interactions with the OPCAT treaty body, the SPT, and its willingness to publish SPT visit reports to places of detention serve as an additional signal of the country type and the government's willingness to incur the cost of improving human rights compliance?

As highlighted in the previous sections, little research exists on which countries create an NPM, what types of NPMs countries create, dependent on their respective regime types, political preferences, and expected cost and benefit considerations, and whether NPM visits improve the treatment of detained persons within OPCAT-ratifying countries. Furthermore, ratifier country signaling in the interactions with the SPT remains an understudied area of research. Previous scholarship has found that NPMs can have a positive effect on the prevention of torture, inhuman and degrading treatment, and punishment (Carver and Handley, 2016, 2020). However, the effect of NPMs appears to be smaller than anticipated.

Relying on the two-type typology introduced in the argument, I propose a first set of hypotheses that focus on **government cost-benefit considerations and corresponding NPM design choices**:

> **Hypothesis 1a**: *Human Rights Champions will create an empowered NPM.*

Human Rights Champions that ratify the OPCAT can be expected to create an empowered NPM because they seek to delegate the monitoring of detention conditions to a specialized institution, and the expected benefits outweigh the initial cost of NPM creation. The NPM will be carrying out its monitoring and reporting tasks unrestricted, and the country will continue to maintain a high human rights standard. Frequent monitoring and reporting can be expected to align both with the government's and the NPM's priorities and provide higher benefits than costs to both.

> **Hypothesis 1b**: *Human Rights Contenders that expect the benefits of NPM-creation to outweigh the costs will create a partially to fully-empowered NPM.*

Creating a partially to fully empowered NPM can be assumed to impose costs on Human Rights Contenders where human rights improvements may require substantial changes in domestic practices. Consequently, only countries that perceive the NPM creation to be benficial, and seek to acquire compliance-related benefits such as gaining access to the EU market, development funds, or similar benefits through signaling their commitment to improved human rights compliance will introduce this type of NPM, even if the NPM is likely to impose costs and initially detail existing human rights abuses in its visit reports. For the NPM, a higher level of empowerment is likely to align with its own cost-benefit considerations, prompting it to conduct its monitoring and reporting as frequently as organizational and financial provisions permit. NPMs may begin shirking their designated role and increase their monitoring and reporting frequency to attain even higher benefits of human rights promotion. While not intended by the government at the time of NPM creation, this increased activity is not likely to result in increased pushback from the government if it still expects the benefits to outweigh the costs of monitoring and reporting.

> **Hypothesis 1c**: *Huma Rights Contenders that are subjected to prohibitively high*

> *costs of NPM monitoring and reporting will restrain their NPM.*

Countries that consider the costs imposed by an empowered NPM to be prohibitively high due to its frequent publicizing of ongoing abuses, or when long-term benefits appear unattainable, are likely to change their policy goals, impose ex-post controls and restrain the NPM's monitoring and reporting power. Alternatively, governments that worry about an overzealous NPM will restrain it from restricting its operation. Furthermore, Human Rights Contenders that prove incapable of improving their human rights standards will equally restrain the NPM to avoid domestic and international criticism for continued human rights abuses. While the NPM is likely to prefer increased monitoring and reporting over restricted operations, it will reduce its monitoring and reporting activities to prevent government punishment and reduce the costs of non-alignment with the government's priorities.

> **Hypothesis 1d**: *Human Rights Contenders that find the benefits of NPM monitoring and reporting to outweigh the costs will maintain an empowered NPM or further empower their partially restrained NPM.*

Human Rights Contenders that reap the intended benefits of creating an NPM, such as gaining access to markets, are granted funds or acquire similar benefits, will maintain a partially empowered NPM and not restrain its activities, or will provide the NPM with additional powers to continue reaping the benefits. These countries will bear the costs of NPM monitoring and reporting as a necessary side-effect of achieving their intended policy preferences. In line with the government's priorities to continue reaping the benefits of NPM monitoring and reporting, the NPM will maintain its frequent monitoring and reporting or increase its monitoring and reporting frequency to achieve the benefit of improving human rights standards.

While the first set of hypotheses focus directly on government cost-benefit considerations and NPM design choices, a second set of hypotheses moves away from domestic dynamics and focuses on **strategic signaling and oversight dynamics through the involvement of a third actor, the SPT**, which moves the dynamics to include an international dimension. In this set of hypotheses, the SPT and its visit and reporting practices are included in the analysis to model how governments take advantage of international signaling opportunities when agreeing or refusing to publish SPT after visit reports. The choice to publish or refuse to publish can be assumed to be a representation of a country's underlying compliance type as Human Rights Champions or Human Rights Contenders.

> **Hypothesis 2a**: *Human Rights CHampions are likely to see a low number of SPT*

visits.

Human Rights Champions are likely to create an empowered NPM, which will provide regular and detailed reports about domestic human rights standards to the SPT. Consequently, the SPT will visit places of detention within these countries fairly infrequently, because these countries are perceived to maintain an NPM that conducts frequent visits and provides reliable reports, thereby not requiring extensive SPT oversight and guidance.

> **Hypothesis 2b**: *Human Rights Champions are likely to agree to the publication of SPT visit reports.*

When the SPT conducts a visit to HRSCs, Human Rights Champions will agree to the publication of the visit report, because the benefits of signaling their human rights commitment and compliance to domestic and international audiences outweigh the costs of potential SPT criticism within the report.

> **Hypothesis 2c**: *Human Rights Contenders are likely to see a higher number of SPT visits.*

Human Rights Contender countries are likely to receive a higher count and frequency of SPT visits because of their lower standard of human rights respect in places of detention. At the onset and right after the counties' OPCAT ratification, visit frequencies are expected not to differ among Human Rights Contender countries because the SPT cannot know for certain whether the respective government will create an empowered NPM and permit monitoring and reporting, and whether the NPM will report truthfully.

> **Hypothesis 2d**: *Human Rights Contenders that create an empowered NPM are likely to see more frequent SPT visits than Human Rights Champion countries, but will agree to the publication of visit reports by the SPT to signal reforms and increased compliance.*

Human Rights Contenders that create an empowered NPM are likely to see a higher number of SPT visits because of their lower human rights standards than Human Rights Champions. The NPM will have an interest in improving human rights practices and will communicate the existing level of abuse to the SPT in its annual reports. Despite the costs of cooperation, these countries will agree to the publication of SPT visit reports, despite the likelihood of the reports detailing existing instances of abuse. The publication of the report can serve as a signaling opportunity to demonstrate that these low human rights standard countries are working towards improving their human rights standards and accept international scrutiny and criticism of abuse as part of their reform process. For these countries, the benefits of truthful reporting and cooperation outweigh the costs of permitting NPM monitoring and reporting and publicizing SPT visit reports.

> **Hypothesis 2e**: *Human Rights Contenders where NPM monitoring and reporting is prohibitively costly or which prove incapable of implementing reforms will see more frequent SPT visits but will not consent to the publication of visit reports by the SPT.*

In contrast to Human Rights Contenders that perceive the benefits of human rights monitoring and reporting to outweigh the costs, weak compliance states in which the governments perceive NPM monitoring and reporting costs to be prohibitively high and outweighing the benefits are likely to restrain their NPMs with the intent of preventing them from documenting persistent human rights abuses. NPMs that face government pushback will further refrain from pushing their human rights advocacy agenda to prevent increasing costs from a repressive government. In line with this information suppression approach, governments of lower human rights respect countries that seek to reduce the costs of human rights monitoring and reporting will not consent to the publication of SPT visit reports to further prevent publication of human rights concerns within the country. Similarly, countries that prove incapable of implementing human rights-promoting reforms will not consent to SPT visit report publication, because they seek to prevent domestic and international criticism and the political cost of ongoing human rights abuse.

Primary Data Sources

As the primary data source to test Hypotheses 1a - 1d, this study utilizes the annual visit reports to the SPT to assess NPM institutional empowerment and human rights compliance in 74 OPCAT-ratifying countries. These annual reports provide a unique data source because NPMs gain direct access to places of detention to interview detained persons, view medical records, and assess detention conditions. In the reports, NPMs detail their findings on different categories of human rights standards, ranging from *detention conditions* (are detained persons housed in a human rights compliant manner, are provided sufficient food and water, have sufficient space, access to fresh air, clean water, clean facilities, and similar indicators), to the *respect for legal safeguards* (are detained persons informed about their rights, can they access legal counsel, is medical treatment available and does a doctor see them privately, is it possible for them to identify personnel of the facility by name, and others), to *incidents of torture, inhuman and degrading treatment or punishment* (are detained persons physically abused, do they receive required medical care, are they prevented from sleeping or other, similar methods of torture-like treatment) in detention.

Table 2: Summary of Available OPCAT NPM Reports by Country (2002–2024)

Country	First Year	Last Year	Total Reports	English	Non-English	Mean Length	Language Type
Albania	2004	2023	20	8	12	242.4	Mixed
Argentina	2019	2023	5	0	5	166.6	Non-English
Armenia	2004	2023	16	14	2	174.9	Mixed
Austria	2012	2023	12	12	0	177.9	English
Azerbaijan	2011	2022	12	12	0	94.4	English
Bolivia (Plurinational State of)	2022	2023	2	0	2	97.0	Non-English
Bosnia and Herzegovina	2009	2023	15	13	2	191.7	Mixed
Brazil	2017	2024	7	0	7	186.1	Non-English
Bulgaria	2012	2023	12	12	0	52.9	English
Burkina Faso	2020	2023	4	0	4	97.0	Non-English
Chile	2020	2023	3	0	3	258.7	Non-English
Costa Rica	2008	2023	16	0	16	106.1	Non-English
Croatia	2012	2023	12	12	0	50.7	English
Cyprus	2018	2023	6	5	1	66.5	Mixed
Czech Republic	2006	2023	18	8	10	82.3	Mixed
Denmark	2009	2023	14	14	0	90.0	English
Ecuador	2013	2020	8	0	8	84.3	Non-English
Estonia	2008	2023	14	13	1	53.7	Mixed
Finland	2016	2023	8	8	0	152.0	English
France	2008	2023	16	5	11	318.4	Mixed
Georgia	2009	2021	12	12	0	174.3	English
Germany	2010	2023	14	13	1	88.4	Mixed
Greece	2014	2023	8	8	0	64.1	English
Guatemala	2015	2024	8	0	8	300.8	Non-English
Honduras	2011	2022	3	0	3	50.7	Non-English
Hungary	2012	2023	12	12	0	75.0	English
Iceland	2022	2023	2	2	0	62.0	English
Italy	2015	2023	7	6	1	272.9	Mixed
Kazakhstan	2014	2023	9	5	4	142.4	Mixed
Kyrgyzstan	2013	2023	11	2	9	105.8	Mixed
Latvia	2021	2023	3	2	1	181.3	Mixed
Lebanon	2022	2023	2	2	0	105.5	English
Liechtenstein	2009	2024	6	4	2	9.7	Mixed
Lithuania	2015	2023	7	7	0	23.0	English
Luxembourg	2012	2023	6	0	6	82.0	Non-English
Maldives	2014	2023	10	0	10	66.2	Non-English
Mali	2018	2023	6	0	6	74.5	Non-English
Malta	2014	2019	2	2	0	21.0	English
Mauritania	2017	2024	4	0	4	74.2	Non-English
Mauritius	2023	2023	1	1	0	147.0	English
Mexico	2008	2024	16	0	16	148.4	Non-English
Mongolia	2022	2023	2	2	0	184.5	English
Montenegro	2013	2023	11	1	10	62.9	Mixed
Morocco	2019	2023	5	1	4	67.0	Mixed
Netherlands	2011	2021	10	10	0	27.7	English
New Zealand	2008	2023	14	14	0	44.1	English
Nigeria	2014	2022	7	7	0	154.7	English
North Macedonia	2008	2023	13	6	7	188.8	Mixed
Norway	2014	2023	10	10	0	75.3	English
Panama	2019	2023	5	0	5	318.2	Non-English
Paraguay	2013	2023	10	0	10	116.3	Non-English
Peru	2016	2023	7	0	7	133.0	Non-English
Poland	2008	2022	14	14	0	105.7	English
Portugal	2014	2023	10	3	7	100.2	Mixed
Republic of Moldova	2009	2023	14	10	4	223.9	Mixed
Romania	2016	2023	6	6	0	99.7	English
Rwanda	2016	2024	9	9	0	140.9	English
Senegal	2017	2023	4	0	4	94.0	Non-English
Serbia	2011	2023	13	12	1	86.8	Mixed
Slovakia	2023	2023	1	1	0	27.0	English
Slovenia	2008	2023	16	16	0	149.3	English
South Africa	2020	2023	4	4	0	54.5	English
Spain	2010	2023	14	7	7	222.6	Mixed
Sri Lanka	2019	2022	4	4	0	161.0	English
Sweden	2014	2023	8	8	0	86.5	English
Switzerland	2010	2023	14	6	8	53.7	Mixed
Togo	2012	2023	11	0	11	146.5	Non-English
Tunisia	2017	2021	2	0	2	325.0	Non-English
Türkiye	2018	2023	5	2	3	113.6	Mixed
Ukraine	2012	2023	8	6	2	239.9	Mixed
United Kingdom and Northern Ireland	2010	2024	14	14	0	67.4	English
Uruguay	2015	2023	7	0	7	105.7	Non-English

While not all NPMs submit these reports to the SPT yearly, many of the reports are available on the domestic websites of the NPMs in the respective native languages of the ratifier countries. To permit a comprehensive assessment of *NPM empowerment* within the treaty parties, this study combines annual reports submitted to the SPT and published on the SPT website with those only available on the domestic NPM websites to extract a yearly documentation of human rights monitoring of places of detention in the signatory countries and NPM reporting practices. Table 2 provides summary statistics of the annual reports collected for this study.

Utilizing the dataset of yearly visit reports permits an assessment of NPM empowerment through measuring behavioral indicators of empowerment, such as the visit frequencies (how often do NPMs visit places of detention in a given year, how many different institutions were visited) and monitoring and operating details (did NPMs face barriers when conducting monitoring visits, are they provided a sufficient budget for their operations) of the reports, with the measure being complimented by legal empowerment indicators (are NPMs formally legally independent, are NPMs provided the legal right to visit places of detention) in cases where annual reports reference specific legal safeguards that were put in place to permit the NPMs operational ability.

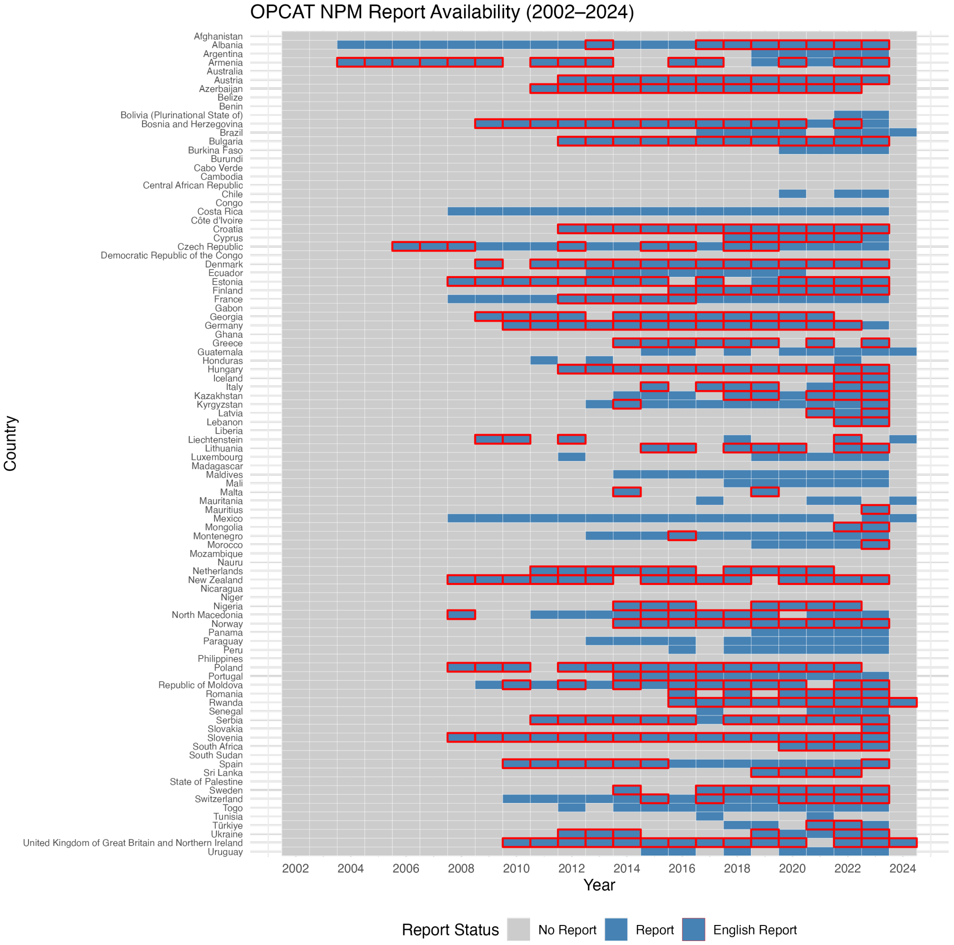

Figure 1: Heatmap of available annual reports from all 94 OPCAT ratifier countries. Tiles framed in red indicate English reports.

By utilizing the annual visit reports, this study assesses the institutional design and empowerment choices ratifier countries make when creating an NPM (Hypotheses 1a- 1d). Figure 1 shows a heatmap of available reports per country since the adoption of the OPCAT in 2002.

To test Hypotheses 2a - 2e, I include additional data collected directly from the SPT (OPCAT, 2025), detailing which countries received SPT visits in which years, and whether the SPT report was published after the monitoring visit occurred. Visit reports are further classified as initial visit, repeat visit, or detail if the visit's thematic focus was on the NPM.

Measurement

To assess Hypotheses 1a-1d, this study analyzes a subset of the OPCAT ratifier countries that have created an NPM, using original data collected from the OPCAT (2024) ratification database, and supplemented by data from the APT (2025), detailing which countries created an NPM and when.

As independent variables, I utilize Fariss et al. (2020)'s latent human rights scores as indicators of human rights standards in the respective countries prior to their OPCAT ratification and NPM creation. In addition to these human rights indicators, I utilize Coppedge et al. (2023) data on regime types to indicate whether an OPCAT ratifying country can be expected to respect its citizens' human rights in general. To assess cost-benefit considerations of the ratifier countries, I utilize data on trade from the IMF (IMF, 2025) and World Bank Development Indicators (World Bank, 2025), data on aid from the OECD Creditor Reporting System (OECD, 2025), and EU accession candidates and their subsequent accession status for countries within Europe (EU, 2025). The dependent variables to assess Hypotheses 1a- 1d consist of the empowerment indices that are extracted directly from the annual reports to the SPT. NPM empowerment is measured using data on monitoring visit frequency, monitoring and reporting details, and legal safeguards for NPM access and independence, which are all extracted from the annual reports.

Hypotheses 2a to 2e are subsequently assessed using the aforementioned data on human rights standards within the ratifier countries using and data on regime types Coppedge et al. (2023) and Fariss et al. (2020) data on human rights respect. For Hypotheses 2d and 2e, I further use the previously created aggregated NPM empowerment scores as additional independent variables. The dependent variables, SPT visits and visit report publication, are created using original data from the OPCAT (2025) database.

In all analyses, I control for GDP per capita, population size (World Bank, 2025), internal and external conflicts (UCDP), protest events (ACLED, SCAD), and regime changes (V-Dem), and regional fixed effects to account for potential confounders.

Data Processing

To extract the dependent variables, NPM empowerment indices, from the multilingual corpus of 718 reports, I adopted a multi-step approach to process the textual data for statistical modeling. As a first step, I fit an agentic Retrieval-augmented generative (RAG) model using LlamaIndex (Liu, 2022) and Mistral AI (Jiang et al., 2023), to process the multilingual reports from one county, Albania. I selected Albania as a use case because Albanian is considered a low-resource language for Natural Language Processing, with limited availability of language models capable of processing Albanian text. The RAG model extraction queries resulted in the finding that Mistral AI's Largest Latest model was capable of processing Albanian text, but provided very little detail from the

Albanian reports compared to English reports. Consequently, I opted to machine translate all available non-English reports to facilitate data extraction. Building on previous research (Osorio et al., 2025), I conducted a comparison of five machine translation engines: DeepL (DeepL, 2018), Deep Translator (2020)'s Google Translate, (Tiedemann and Thottingal, 2020), Meta's NLLB NLLB Team et al. (2022), and ChatGPT's 4.1 mini (OpenAI, 2023). The translation output was validated by a computational linguist and Albanian native speaker, and based on its best performance, I selected ChatGPT 4.1 mini as the translation engine.

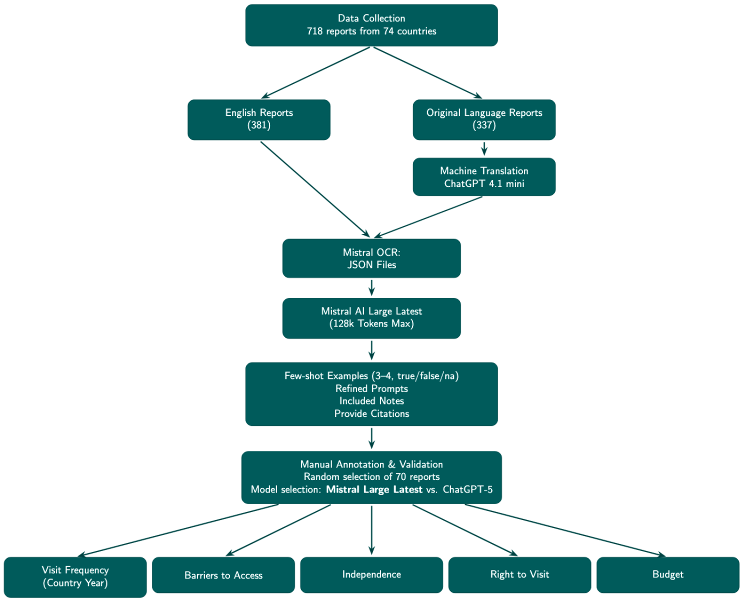

Figure 2: Data Processing and Validation Pipeline

After machine-translating all reports, I processed them through Mistral AI's OCR model to extract JSON files for more efficient processing. I then selected 70 random reports for manual annotation and processing validation. Through prompt engineering, I improved Mistral Largest Latest's extraction performance to a satisfactory level of around or over 80% accuracy for all empowerment indicators except visit frequencies, when processing the validation report corpus via Mistral AI's Largest Latest model and validating the output through comparison with the manually, human-annotated reports. For best model selection, I compared Mistral Largest Latest's extraction accuracy with ChatGPT-5's model performance and selected Mistral Largest Latest as the better-performing model. In cases where Mistral AI is unable to extract complete information from the report, the model outputs that the information was not extracted due to incomplete tables or other processing problems, allowing manual post-processing and supplementation for incomplete data. All extracted data is further documented through citations, and the model temperature is restrained to 0.2 to prevent hallucination. Figure 2 shows the processing and validation pipeline. Detailed prompts and model performance metrics are included in the Appendix.

References

Anaya-Muñoz, Alejandro and Amanda Murdie. 2021. The will and the way: How state capacity and willingness jointly affect human rights improvement. *23*(1), 127–154. Publisher: Springer Verlag.

APT. 2025. Mapping torture prevention.

Carver, Richard and Lisa Handley. 2016. *Does torture prevention work?* Cambridge, England: Liverpool University Press.

Carver, Richard and Lisa Handley. 2020. Evaluating national preventive mechanisms: A conceptual model. *Journal of human rights practice 12*(2), 387–408.

Cingranelli, David and Mikhail Filippov. 2010. Electoral rules and incentives to protect human rights. *The Journal of politics 72*(1), 243–257.

Clark, Ann Marie. 2013. *The normative context of human rights criticism: treaty ratification and UN mechanisms.* Cambridge University Press.

Cole, Wade M.. 2015. Mind the gap: State capacity and the implementation of human rights treaties. *69*(2), 405–441. Publisher: [The MIT Press, University of Wisconsin Press, Cambridge University Press, International Organization Foundation].

Cole, Wade M. and Francisco O. Ramirez. 2013. Conditional decoupling: Assessing the impact of national human rights institutions, 1981 to 2004. *American sociological review 78*(4), 702–725.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Joshua Krusell, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Natalia Natsika, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Oskar Ryden, Johannes von

Romer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundstrom, Eitan Tzelgov, Yi ting Wang, Tore Wig, Steven Wilson, and Daniel Ziblatt. 2023. V-dem [country-year/country-date] dataset v13" varieties of democracy (v-dem) project. *V-Dem*.

Cordell, Rebecca. 2021. The political costs of abusing human rights: International cooperation in extraordinary rendition. *65*(2), 255–282. Publisher: SAGE Publications Inc.

Cordell, Rebecca, K. Chad Clay, Christopher J. Fariss, Reed M. Wood, and Thorin M. Wright. 2020. Changing standards or political whim? evaluating changes in the content of US state department human rights reports following presidential transitions. *19*(1), 3–18. Publisher: Routledge _eprint: https://doi.org/10.1080/14754835.2019.1671175.

Creamer, Cossette D. and Beth A. Simmons. 2015. Ratification, reporting, and rights: Quality of participation in the convention against torture. *Human Rights Quarterly*.

Creamer, Cosette D. and Beth A. Simmons. 2019. Do self-reporting regimes matter? evidence from the convention against torture. *International Studies Quarterly*.

Creamer, Cosette D. and Beth A. Simmons. 2020. The proof is in the process: Self-reporting under international human rights treaties. *114*(1), 1–50. Publisher: Cambridge University Press.

Dancy, Geoff and Kathryn Sikkink. 2012. Ratification and human rights prosecutions: Toward a transnational theory of treaty compliance. *New York University journal of international law politics 44*(3), 751–790.

de Mesquita, Bruce Bueno, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. An institutional explanation of the democratic peace. *The American political science review 93*(4), 791–807.

Deep Translator. 2020. deep-translator: A flexible free and unlimited python tool to translate between different languages in a simple way using multiple translators. `https://github.com/nidhaloff/deep-translator`. GitHub.

DeepL. 2018. Deepl translator. `https://www.deepl.com/translator`.

Downs, George W. and Michael A. Jones. 2002. Reputation, compliance, and international law. *The Journal of legal studies 31*(S1), S95–S114.

EU. 2025. Eu enlargement.

Fariss, Christopher J, Michael R Kenwick, and Kevin Reuning. 2020, Nov). Estimating one-sided-killings from a robust measurement model of human rights. *Journal of Peace Research 57*(6), 801–814.

Gandhi, Jennifer and Adam Przeworski. 2007. Authoritarian institutions and the survival of autocrats. *Comparative political studies 40*(11), 1279–1301.

Goldsmith, Jack L. and Eric A. Posner. 2005. *The Limits of International Law.* Oxford University Press.

Greenhill, Brian, Brian Greenhill, Dan Reiter, Dan Reiter, and Dan Reiter. 2022. Naming and shaming, government messaging, and backlash effects: Experimental evidence from the convention against torture. *Journal of Human Rights*.

Hafner-Burton, Emilie. 2013. *Making human rights a reality.* De Gruyter eBooks. Princeton, New Jersey: Princeton University Press.

Hafner-Burton, Emilie M.. 2008. Sticks and stones: Naming and shaming the human rights enforcement problem. *International Organization*.

Hafner-Burton, Emilie M.. 2013. Making human rights a reality | princeton university press. ISBN: 9780691155364.

Hafner-Burton, Emilie M. and Kiyoteru Tsutsui. 2005. Human rights in a globalizing world: The paradox of empty promises. *110*(5), 1373–1411. Publisher: The University of Chicago Press.

Hathaway, Oona. 2002. Do human rights treaties make a difference? *111.*

Hendrix, Cullen S., Cullen S. Hendrix, Wendy H. Wong, Wendy H. Wong, 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*.

Hill, Daniel W.. 2010. Estimating the effects of human rights treaties on state behavior. *72*(4), 1161–1174. Publisher: [The University of Chicago Press, Southern Political Science Association].

IMF. 2025. Imf data.

Jetschke, Anja and Andrea Liese. 2013. *The power of human rights a decade after: from eurphoria to contestation?* Cambridge University Press.

Jiang, Albert Q., Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Keck, Margaret E. and Kathryn Sikkink. 1998. *Activists beyond borders: Advocacy networks in international politics.* Cornell University Press.

Keck, Margaret E. and Kathryn Sikkink. 2018. Transnational advocacy networks in international and regional politics. *International social science journal 68*(227-228), 65–76.

Kim, Hunjoon and Kathryn Sikkink. 2010. Explaining the deterrence effect of human rights prosecutions for transitional countries. *International studies quarterly 54*(4), 939–963.

Liu, Jerry. 2022, 11). LlamaIndex.

McGaughey, Fiona. 2021. NGOs and the human rights council. In *Non-Governmental Organisations and the United Nations Human Rights System.* Routledge. Num Pages: 20.

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

Neumayer, Eric. 2005. Do international human rights treaties improve respect for human rights? *49*(6), 925–953. Publisher: Sage Publications, Inc.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia-Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation. *arxiv*.

OECD. 2025. Crs: Creditor reporting system (flows).

OHCHR, SPT. 2024a. Annual reports received by the subcommittee from national preventive mechanisms.

OHCHR, SPT. 2024b. Treaty bodies.

OPCAT. 2024. Optional protocol to the convention against torture and other cruel, inhuman or degrading treatment or punishment.

OPCAT, SPT. 2025. Ohchr.

OpenAI. 2023. Chatgpt: Optimizing language models for dialogue. `https://openai.com/chatgpt`.

Osorio, Javier, Afraa Alshammari, Naif Alatrush, Dagmar Heintze, Amber Converse, Sultan Alsarra, Latifur Khan, Patrick T. Brandt, and Vito D'Orazio. 2025. The devil is in the details: Assessing the effects of machine translation on llm performance in domain-specific texts. In *Proceedings of the 20th Machine Translation Summit (MT Summit 2025)*.

Powell, Emilia Justyna and Jeffrey K. Staton. 2007. Domestic judicial institutions and human rights treaty violation.

Risse, Thomas, Steve C. Ropp, and Kathryn Sikkink. 1999. *The Power of Human Rights: International Norms and Domestic Change.* Cambridge University Press.

Simmons, Beth. 2010. Treaty compliance and violation. *Annual review of political science 13*(1), 273–296.

Simmons, Beth A.. 2009. *Mobilizing for human rights : international law in domestic politics.* Cambridge books online. Cambridge ;: Cambridge University Press.

Tiedemann, Jörg and Santhosh Thottingal. 2020, November). OPUS-MT – building open translation services for the world. In André Martins, Helena Moniz, Sara Fumega, Bruno Martins, Fernando Batista, Luisa Coheur, Carla Parra, Isabel Trancoso, Marco Turchi, Arianna Bisazza, Joss Moorkens, Ana Guerberof, Mary Nurminen, Lena Marg, and Mikel L. Forcada (Eds.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, Lisboa, Portugal, pp. 479–480. European Association for Machine Translation.

UN Treaty bodies. 2024. Self reporting procedure.

United Nations, OPCAT. 2024.

United Nations, SPT. 2024. Annual reports received by the subcommittee from National Preventive Mechanisms — Ohchr.

Von Stein, Jana. 2005. Do treaties constrain or screen? selection bias and treaty compliance. *The American political science review 99*(4), 611–622.

von Stein, Jana. 2005. Do treaties constrain or screen? selection bias and treaty compliance. *99*(4), 611–622. Publisher: Cambridge University Press.

von Stein, Jana. 2016. Making promises, keeping promises: Democracy, ratification and compliance in international human rights law. *British Journal of Political Science 46*(3), 655–679.

Vreeland, James Raymond. 2008. Political institutions and human rights: Why dictatorships enter into the united nations convention against torture. *62*(1), 65–101. Publisher: [MIT Press, University of Wisconsin Press, Cambridge University Press, International Organization Foundation].

Welch, Ryan M.. 2017. National human rights institutions: Domestic implementation of international human rights law. *Journal of human rights 16*(1), 96–116.

Welch, Ryan M.. 2019. Domestic politics and the power to punish: The case of national human rights institutions. *Conflict management and peace science 36*(4), 385–404.

Welch, Ryan M., Jacqueline H. R. DeMeritt, and Courtenay R. Conrad. 2021. Conceptualizing and measuring institutional variation in national human rights institutions (nhris). *The Journal of conflict resolution 65*(5), 1010–1033.

World Bank. 2025. World development indicators.

Zhou, Yuan, Ghashia Kiyani, and Charles Crabtree. 2022. New evidence that naming and shaming influences state human rights practices. *Journal of Human Rights*.