The Review of International Organizations (2022) 17:89–115
https://doi.org/10.1007/s11558-021-09413-5

Instrumental or intrinsic? Human rights alignment in intergovernmental organizations

David Benjamin Weyrauch[1] · Christoph Valentin Steinert[2]

Accepted: 7 January 2021 / Published online: 1 June 2021

Abstract

Why do states' human rights records converge with co-members in intergovernmental organizations (IGOs)? This study provides new insights on whether interactions in IGOs have the capacity to genuinely transform state preferences or whether norm diffusion is a consequence of instrumental processes. We leverage information about the timing of human rights alignment to disentangle intrinsic from instrumental motives. We hypothesize that instrumental motives find expression in pre-membership alignment and reversions to original normative standards after IGO exits. Intrinsic motives lead to gradual alignment during IGO membership and result in stable normative changes beyond IGO exits. Using varying-slopes, varying intercepts models, we investigate the distance on human rights indices between individual states and IGO means. While we find evidence for systematic convergence during IGO membership, no significant changes occur before and after IGO membership. Testing alignment of different physical integrity rights, we find no evidence for instrumental shifts to clandestine repression during IGO membership. Overall, the results suggest that norm alignment in IGOs is at least not exclusively instrumentally motivated. Our findings support constructivist arguments on state interests and suggest that IGOs are capable of transforming states' human rights related preferences.

Responsible editor: Axel Dreher

Both authors contributed equally to all parts. Both contributed approximately 50% of the research design, 50% of the quantitative analysis and drafted approximately 50% of the prose. The order of authors does not reflect the significance of authors' contributions but was decided via coin toss.

✉ David Benjamin Weyrauch
dweyrauc@mail.uni–mannheim.de

[1] Department of Political Science II, European Politics, University of Mannheim, A5, 6 Room 351, Mannheim, Germany

[2] Department of Political Science IV, University of Mannheim, A5, 6 B124, 68159 Mannheim, Germany

Keywords Intergovernmental organizations · Human rights · Norm diffusion · Socialization · Strategic repression

1 Introduction

Constructivism has greatly enriched the scholarly understanding of political processes by emphasizing that state interests are no fixed entities but fundamentally shaped by the types of interactions that states are involved in (e.g., Klotz, 1995; Wendt, 1999). Building on this premise, recent research has demonstrated that intergovernmental organizations (IGOs) provide forums for the transmission of human rights norms (Greenhill, 2010, 2016).[1] Empirically, these studies identify a strong correlation between the human rights records of a country's fellow IGO members and a state's own human rights performance in subsequent periods. This finding is primarily explained by socializing effects pointing to the transmission of group norms to individual states through repeated interactions.

While appreciating their substantial contribution to our understanding of the global diffusion of human rights norms, we contend that these studies lack clarity when it comes to the causal mechanisms driving human rights alignment in IGOs. Two opposing mechanisms are potentially at play: states could alter their repressive outputs in order to conform to prevalent group norms and to derive benefits from membership (instrumental norm alignment).[2] Alternatively, states could internalize different norms as a consequence of their repeated interaction with other states resulting in genuine preference transformation (intrinsic norm alignment). At this stage, empirical scholarship is unable to answer the question of *why* states adapt their human rights records as members of IGOs.

We argue that the timing of norm convergence contains valuable insights into the underlying motives for human rights alignment. Our key argument is that instrumental and intrinsic mechanisms materialize at different points in time. We argue that instrumental motives result in immediate alignment processes at the time of IGO accession whereas norm internalization is a gradual process evolving during the time of IGO membership. Hence, the length of IGO membership is expected to be systematically related to norm alignment in the case of intrinsic motives for human rights alignment. We further suggest that post-membership human rights records enable us to draw inferences about the genuineness of norm adaption. If states' post-membership human rights records revert to pre-accession levels, norm alignment was likely instrumentally motivated. In contrast, if human rights records change during the time of membership and remain close to IGO averages beyond, norm adoption is likely intrinsically motivated.

[1] We apply the standard definition of intergovernmental organizations capturing formal organizations consisting of at least three states that hold regular plenary sessions and that possess a permanent secretariat (Pevehouse et al., 2004).

[2] Goodman and Jinks (2013) differentiate between material pressures for norm alignment, termed as 'inducement', and social pressures for norm alignment, labeled as 'acculturation'. We focus solely on the intrinsic vs. instrumental dichotomy disregarding differences between materially induced and socially induced instrumental motives.

Furthermore, we exploit information about the type of physical integrity rights violations committed by states to investigate the causal mechanisms underlying norm convergence.[3] Physical integrity rights violations differ in their level of intensity and, consequently, in the associated reputation costs (e.g., Greitens, 2016; Hafner-Burton, 2008). In response, states have strategically altered their repressive mix to avoid sanctions. Disappearances are more difficult to link to the government than other human rights violations since perpetrators tend to remain hidden. Kinzelbach and Spannagel (n.d.), highlight, for instance, how Argentina shifted in the 1970s from political imprisonments to disappearances in light of international pressure for the release of political prisoners. Against this backdrop, we assume that instrumental norm adaption in IGOs leads governments to shift to clandestine physical integrity rights violations that are less attributable to the state (Payne and Abouharb, 2016). In contrast, states that internalize human rights norms improve their records consistently across different types of physical integrity rights violations.

We find that new member states' human rights scores systematically align with IGO means during IGO membership. The longer states are members of IGOs, the closer their human rights scores align with group averages. In line with the instrumental logic, we find weak evidence for pre-membership alignment. However, in contradiction of the instrumental logic, states tend to keep altered human rights standards when they exit IGOs instead of reverting to pre-accession levels. By studying different types of physical integrity rights violations, we also find no evidence that states strategically shift to repression that is less attributable to governments. Most notably, respect for the right not to be disappeared systematically aligns with group standards over the time of IGO membership. Overall, our results suggest that human rights alignment in IGOs is at least not exclusively driven by instrumental motives as suggested by prior research (Greenhill, 2010; Greenhill, 2016). In contrast, the findings suggest that interactions in IGOs might result in genuine preference transformation in the context of human rights.

Our findings hold across a wide range of IGOs with different policy scopes and with varying degrees of authority. We demonstrate that the alignment effect is not driven by the subset of human rights IGOs but shaped by the average human rights norms of the respective member states in an IGO. We contribute to the norm diffusion literature by providing a new perspective on the mechanisms driving norm alignment (e.g., Bearce and Bondanella, 2007; Gilardi and Wasserfallen, 2019; Taninchev, 2015). Specifically, we contribute to the body of scholarship dedicated to the alignment of human rights norms in IGOs (Greenhill, 2010; Greenhill, 2016; Lindemann and Petiteville, 2019). We also contribute to the literature dedicated to the effectiveness of external means to promote human rights (e.g., Donno and Neureiter, 2018; Kelley and Simmons, 2015; Terman and Voeten, 2018). We show that co-membership in IGOs provides an important channel to affect human rights practices of other states and that new normative contents might be transmitted through repeated interactions.

This article proceeds as follows: first, we discuss previous research on human rights alignment in IGOs. Second, we present our theoretical framework explaining the logic underlying human rights alignment in IGOs. Third, we introduce our empirical strategy

[3] Physical integrity rights violations are a subset of human rights violations including the right not be tortured, to be disappeared, extra-judicially killed, or imprisoned for political beliefs (Cingranelli et al., 2014).

exploiting both the timing and duration of IGO membership and information about the type of physical integrity rights violation to disentangle instrumental and intrinsic motives for norm alignment in IGOs. Fourth, we introduce our empirical models accounting for the hierarchical data structure by controlling for state and IGO characteristics. Lastly, we discuss the empirical findings and elaborate on the limitations of our study.

2 Norm alignment in intergovernmental organizations

The social network perspective on state interests, which emphasizes interactions in IGOs, has a long scholarly tradition. Finnemore (1993) famously described international organizations as "teachers of norms" and the "world society"-school claims that international organization socialize states into a universal set of norms (Boli and Thomas, 1999; Beckfield, 2003; Meyer et al., 1997).[4] The assumption underlying these studies is that international norms diffuse into domestic politics through repeated interactions of state representatives in IGOs. State representatives exchange ideas, obtain new information, and develop a sense of cohesion and loyalty to colleagues in IGOs (Chelotti et al., 2018).[5] As a consequence, membership in IGOs leads to a convergence in states' policy preferences (Bearce and Bondanella, 2007; Cao, 2009; Checkel, 2005; Chelotti et al., 2018). Reflecting a logic of appropriateness (March and Olsen, 1998), these convergence processes have been explained with social pressure (Goodman and Jinks, 2013), concerns for prestige (Paul et al., 2014), or desires for conformity (Johnston, 2001). Others contend that norm diffusion processes genuinely shape state interests emphasizing mechanisms of teaching and persuasion (Gheciu, 2005; Lewis, 2005).

While IGO-driven convergence processes have been studied in several policy areas such as domestic economic policies (Cao, 2009), public-sector downsizing (Lee and Strang, 2006), or economic liberalism (Simmons et al., 2006), scholars have only recently turned to human rights norms. Greenhill pioneered this field of research with a foundational article in 2010 and a more extensive analysis of his theory in his 2016 monograph. His key findings are that IGO networks provide a conduit for the diffusion of human rights norms leading to convergence processes over time. The impact of IGOs goes beyond their formal mandates since diffusion processes are not sensitive to the mandate of IGOs. More specifically, Greenhill discovers that the convergence effect is not only restricted to those IGOs with a direct connection to human rights issues. He explains this finding with pressures for social conformity in IGOs leading states to strategically adapt their human rights records to group standards.

While appreciating Greenhill's substantial contribution to our understanding of human rights alignment in IGOs, we contend that his analysis is unable to explain whether intrinsic or instrumental motives lead states to align their human rights records

[4] These studies tend to use the term "international organization" instead of "international governmental organizations" including also non-governmental organizations and multinational corporations. We stick to the term IGOs because direct encounters of government representatives are most likely to have an impact on state preferences.

[5] Lewis (2005) provides a powerful illustration of this mechanism using the example of the EU's Committee of Permanent Representatives.

with co-members in IGOs. Greenhill suggests that intrinsic motives would imply that norm alignment is stronger in the subset of IGOs specifically dedicated to human rights.[6] That is, because human rights IGOs (e.g. UN Human Rights Council) are more likely to shape human rights related preferences. Since he finds no evidence for stronger alignment processes in human rights IGOs, he rejects intrinsic motives as drivers of norm alignment. However, this test cannot convince since it is likely that group pressure related to human rights violations is also stronger in the subset of IGOs specifically dedicated to human rights. In other words, while human rights IGOs might be more likely to shape human rights-related preferences, they also make human rights violations more salient leading to enhanced human rights-related group pressure. Thus, it is unclear whether his finding refutes the intrinsic or the instrumental logic. Hence, we argue that a subgroup analysis of human rights IGOs offers no compelling evidence for the underlying motives of norm convergence across IGOs.

Furthermore, Greenhill (2010, 2016) argues that policy learning is less relevant in the realm of physical integrity rights since a lack of information is unlikely to be responsible for violations of these rights. However, given that weak state organization characterized by agency slack have been identified as a key driver of human rights abuse (Butler et al., 2007; Hafner-Burton, 2014), we deem information-sharing with regard to effective capacity-building as essential in shaping human rights compliance. For instance, it has been shown that effective security sector reforms contribute to improved human rights records (e.g., Holm and Eide, 2000). Beyond that, we argue that it is problematic to reduce non-instrumental motives for norm alignment entirely to information-sharing. As Greenhill conceives persuasion himself as a process involving a "genuine change in [states'] beliefs" (p. 44), it includes more complex dynamics such as the diffusion of values and the transformation of preferences.

All things considered, we argue that it remains unclear why convergence processes of human rights norms take place in IGOs. In the following, we first present our theoretical framework depicting IGOs as social environments. Subsequently, we introduce our new empirical approaches to disentangle the mechanisms of instrumental and intrinsic norm alignment in IGOs.

3 Intergovernmental organizations as social environments

We consider intergovernmental organizations as social environments that create channels of international influence (see Goodman and Jinks, 2013). Beyond their formal purposes, IGOs are venues of social interaction and international exchange. Building on Greenhill's (2016) 'three-stage model of social influence in IGOs', we argue that there is a reciprocal influence between IGOs and states. Individual states shape the social environment of IGOs and at the same time, they are affected by this environment themselves. In the following sections, we explain the key steps in the causal chain explaining the impact of IGOs on states: (1) IGOs develop 'human rights cultures', (2)

[6] We use the term instrumental norm alignment to capture both social conformity pressures (acculturation) and material conformity pressures (inducement) and intrinsic norm alignment to describe the mechanism termed 'persuasion' in Greenhill's study (see Greenhill, 2016; Goodman and Jinks, 2013).

states are subject to different alignment pressures in IGOs, and (3) preferences in IGOs shape domestic policy preferences.

3.1 'Human rights cultures' in IGOs

IGOs are created to promote international cooperation and standard-setting in diverse policy fields (Hooghe and Marks, 2015). To fulfil these goals, IGOs provide physical spaces where state representatives can come together, deliberate, and compromise. While the key objective of these encounters is defined by the policy scope of the IGO, we argue that these encounters cannot be reduced to technical decisions about pre-defined topics. The deliberation of common policies in IGOs is a profoundly social process that relies on sharing ideas, on persuading and conceding, and on forming alliances.

Representatives, engaging in IGOs, are deeply aware of the social nature of policy-making. Therefore, they place a strong emphasis on forming networks and establishing social ties. Official deliberations in IGOs are complemented by various social events and informal meetings (Greenhill, 2016). By repeatedly interacting in different social settings, state representatives tend to develop a sense of cohesion and loyalty to other group members (Hofstede, 2004). Such social ties might be deemed especially valuable during shared expatriate experiences at the headquarters of IGOs that are far from home for the majority of state representatives.

Social ties between individuals tend to form around some form of shared identity (Kossinets and Watts, 2009). The combination of national institutions, history, and social environments shape the identities of states and, by association, of their diplomats. Human rights norms are likely to lie at the core of these identities. Various states perceive human rights norms as 'universal' or 'unalienable' principles that form the backbone of their constitutional laws. Role theory describes how such national identities may affect social interactions in international relations (Harnisch et al., 2011). Hence, it is plausible that human rights norms, as key elements of state identities, affect the social interactions of states. In practical terms, state representatives are likely to form social ties with other diplomats that share a core set of principles and ideas. However, they might be hesitant to form ties with representatives of states that tolerate torture or arbitrary killings.

Based on these insights, it has been argued that IGOs develop their own 'human rights cultures' (see Greenhill, 2016). These IGO-specific human rights cultures reflect the aggregate views on human rights norms of all member states in an IGO. Thus, not the policy scope of an IGO but the aggregate norms of its group members are constitutive for its human rights culture. For instance, major human rights organizations such as the UN Human Rights Council may have comparatively weak human rights cultures given that various human rights abusers are among its members (Freedman and Houghton, 2017). In contrast, an IGO that is unrelated to human rights such as the International Organisation of Vine and Wine might have a stronger human rights culture given that its members have comparatively positive human rights records.

While member states shape the human rights culture of IGOs, human rights cultures likewise affect individual states. Greenhill's (2016) study provides powerful evidence that the human rights norms of states converge with co-members in IGOs. He demonstrates that alignment patterns are unrelated to the policy scopes of IGOs but driven by

the aggregate human rights norms of their group members. States tend to align with the prevailing human rights norms in IGOs constituted of the average human rights scores of their member states. Two different types of group influence could explain this empirical pattern.

3.2 Two types of group influence

States could align their human rights records with co-members in IGOs due to *instrumental* or *intrinsic* motives.

Instrumental norm alignment describes the strategic adaption to prevalent group norms in response to social or material incentives. States could have incentives to align with prevalent human rights norms since co-members in IGOs might link cooperation on specific policy issues to improved human rights records. Issue-linkage has been demonstrated to be a powerful tool of inter-state influence (e.g., Haas, 1980; McGinnis, 1986). States might derive material benefits from standard-setting in IGOs that aligns with their favored policy positions. Hence, it might be strategically beneficial to adapt to prevalent human rights norms if compliance is linked to desired policy outputs.

Further, instrumental alignment could be driven by social pressure. States are generally susceptible to group pressure since they value status and prestige (see Paul et al., 2014).[7] IGOs provide a social environment where group pressure is especially salient (Greenhill, 2016; Johnston, 2001). Co-membership in IGOs gives states a platform to express their discontent over norm violations and allows them to form alliances against pariah states.[8] Representatives of human rights abusing states must bear the immediate social cost of international contempt. In response, state representatives might have incentives to mimic their social environment driven by the desire to "fit it" (Turner et al., 1987). Previous research suggests that human rights alignment in IGOs is primarily the product of such instrumental motives (see Greenhill, 2010; Greenhill, 2016).

Alternatively, norm alignment in IGOs could be driven by intrinsic processes. Intrinsic norm alignment occurs when actors redefine their interests and identities in line with prevalent group norms (see Goodman and Jinks, 2013). The precondition for intrinsic norm alignment is the spread of norms through repeated interactions between actors (Finnemore and Sikkink, 1998). However, norm exposure does not necessarily translate into norm internalization. Norm internalization occurs only if norm receivers start to sincerely believe in the intrinsic value of transmitted norms. Therefore, the substantive content of norms is decisive in the case of intrinsic norm alignment.[9]

Since intrinsic norm alignment relies on the practice of deliberation, IGOs provide an ideal setting for this process (see Risse, 2000). Intrinsic alignment is most likely to occur when IGOs make an explicit effort to educate their members. For instance, Gheciu's (2005) study describes how the NATO educated elites from new member states which led to long-lasting policy changes even after the membership conditions

[7] On the reverse side, state policies might be subject to public shaming (DeMeritt, 2012; Kelley and Simmons, 2015).

[8] Dorussen and Hard (2008) demonstrate how the IGO network provides communication channels between states allowing them to intervene more effectively in state affairs.

[9] The process of intrinsic alignment reflects fundamental insights of constructivist scholarship (e.g. Keck and Sikkink, 2014; Tannenwald, 2007).

were fulfilled. Intrinsic alignment may also occur beyond institutionalized platforms resulting from day-to-day interactions between state representatives. While it may be at times the product of extensive deliberation processes, it might also occur from self-reflection in light of an external reference point.

Further, intrinsic norm alignment includes learning processes of effective means to promote certain norms. As members of IGOs, state representatives might learn about effective policies to promote human rights leading to a redefinition of their policy preferences. For instance, co-members in IGOs might share intelligence on effective security sector reforms that enhance human rights (e.g., Holm and Eide, 2000). If the acquired knowledge of such policies leads to redefined policy preferences, norm alignment is intrinsically motivated.

To summarize, instrumental norm alignment is strategic in nature being agnostic to the substantive contents of group norms. Actors seek to conform for the sake of deriving benefits from group membership and not because of their inherent appreciation of specific norms.[10] In contrast, intrinsic norm alignment captures the genuine endorsement of new normative contests leading to a redefinition of state preferences and identities.

3.3 The impact of transmitted norms on domestic policies

Finally, updated preferences of state representatives in IGOs must affect domestic policy-making to shape states' human rights records. This is the final link in the causal chain between human rights cultures in IGOs and national human rights records. The link is straightforward in the case of strategic (instrumental) alignment in response to issue-linkage. State representatives will communicate the conditions of favored policy compromises to policy-makers that decide whether to align their policies accordingly.

Three mechanisms could explain how intrinsic norm alignment processes in IGOs could shape domestic policy-making (see Greenhill 2010, 2016). First, state representatives with altered preferences might seek to convince their policy-makers to adopt different policies. Second, there are various examples of state representatives that become policy-makers themselves at later stages of their careers. Third, several policy-makers directly attend IGO meetings leading to direct impacts on their policy preferences.[11]

4 Disentangling instrumental and intrinsic motives

How can we identify which of those two mechanisms - instrumental or intrinsic norm alignment is decisive for norm convergence in IGOs? In practice, the boundary lines are far from clear-cut. States might be motivated by an interplay of instrumental and intrinsic motives when adapting to prevalent standards in IGOs. Adding further complexity, it might be the case that states are originally motivated by instrumental motives but begin to internalize norms over time. Officially stated motives are likely to deviate from actual

[10] Reflecting this instrumental motive, Goodman and Jinks (2013) describe this process also as "incomplete internalization".

[11] A detailed discussion of these micro-level processes goes beyond the scope of our study and can be found in Greenhill (2016: 48–52).

motivations. State representatives might publicly convey a rhetoric of norm endorsement while being motivated by instrumental motives. Being aware of these difficulties, we seek to disentangle the respective impact of those mechanisms by leveraging information about the timing of norm alignment and strategic shifts of the types of human rights violations.

4.1 Exploiting the timing of norm convergence

First, we argue that motives for norm convergence can be identified by studying the timing of norm convergence. The underlying assumption is that instrumental and intrinsic mechanisms materialize at different points in time. The idea to leverage the timing of norm convergence has been applied in different contexts such as in research about the protection of labor rights (Kim, 2012) or the signing of international treaties (von Stein, 2005). Building on these studies, we distinguish between norm convergence (a) *prior to* IGO accession, norm convergence (b) *during* IGO membership, and the development of normative standards (c) *after* IGO exit to draw inferences about the impact of instrumental and intrinsic motives on norm alignment.

(a) Beginning with the time prior to IGO membership, we argue that systematic norm convergence shortly before membership is indicative of instrumental motives. States might instrumentally adapt their human rights records to enhance their chances to be admitted to an IGO. This process could occur voluntarily or due to formal conditionality requirements. For instance, states might strategically adapt their domestic policies in the realm of human rights to be admitted to IGOs. Membership in IGOs could be linked to diverse material benefits such as access to markets or influence on standard-setting. If the anticipated material benefits of cooperation are the underlying motivation for human rights improvement, norm alignment is driven by instrumental motives.

In contrast, we argue that sudden norm alignment in the pre-membership phase is unlikely to be driven by intrinsic motives. Certainly, norm diffusion also takes place outside of IGOs (Gilardi, 2016; Volden et al., 2008). However, it is implausible that such general diffusion processes between states systematically increase shortly before IGO accession. Further, it has been shown that norm diffusion in IGOs goes beyond general regular norm diffusion between states (Freyburg, 2015). It could be objected that intrinsic norm alignment before accession leads states to self-select into IGOs with similar normative standards. However, it has been demonstrated that IGOs tend to form on a regional basis and not due to self-selection related to human rights standards (see Greenhill, 2016). Consequently, we argue that sudden norm convergence in the immediate run-up to IGO membership is indicative of instrumental motives. Building on these insights, we test the impact of instrumental motives on norm alignment with the following hypothesis:

> H_1: States adapt their human rights records to group standards shortly before joining an intergovernmental organization.

(b) Both mechanisms are likely at play during IGO membership. States could adjust their human rights records for instrumental motives to reduce group pressure and evade the risk of expulsion. States could also align their human rights records for intrinsic motives as a product of repeated interactions with other group members. Hence, both

motives are expected to be observationally equivalent if membership is regarded in its entirety. Therefore, we suggest a dynamic perspective on the time of membership focusing on the pace of norm alignment.

Given that normative standards and identities are characterized by a certain degree of stability, we expect intrinsic alignment processes to evolve gradually over the course of membership time. As implied by the influential 'norm cascade' (Finnemore and Sikkink, 1998), actors must be confronted with new normative contents over a prolonged time involving demonstration and persuasion processes before these contents might be internalized. Weyland (2012) states that "new values and principles do not diffuse instantaneously, persuasion takes time" (p. 919). Prior research also demonstrated that persuasion is enhanced in targeted, structured, and interpersonal settings such as IGOs (Freyburg, 2015). Hence, we expect, on average, gradually increasing alignment during IGO membership in the case of intrinsic alignment. In contrast, we would expect immediate norm convergence and no systematic trends over membership time in the case of instrumental alignment. Against this backdrop, we suggest the following observable implication of intrinsically motivated norm alignment:

> H_2: The longer states are members of intergovernmental organizations, the closer their human rights record align with the average of the other members.

(c) We further argue that human rights records after IGO exits are telling for the genuineness of norm adoption. More specifically, we contend that intrinsic norm diffusion leaves a visible trace beyond the membership time. That is because if states adopt group norms due to intrinsic commitments, it is unlikely that they abandon these norms once membership is terminated. The same applies if intrinsic alignment operates through a learning-mechanism resulting in superior levels of expertise about effective human rights policies. States are unlikely to jettison their acquired knowledge once IGO membership is terminated. Therefore, we expect that intrinsically motivated norm alignment creates a normative legacy being reflected in stable human rights records after IGO exists. If norm adaption was primarily driven by instrumental reasons, we would expect that states diverge from group norms once the incentives for alignment disappear.[12]More specifically, we would expect a state to revert to its original human rights record before the IGO might have exerted influence on its norms. We test the instrumental logic with the following hypothesis:

> H_3: States return to their pre-membership human rights records after IGO exits.

The theoretical expectations are summarized in Fig. 1. While the stylized pathways suggest two clearly delimited processes, it is important to emphasize that the pathways

[12] IGO exits result from self-selection dynamics that likely reflect substantial disagreements with remaining members. If a state leaves to escape IGO policy constraints, it might be likely to use its newly-gained independence to increase policy distance (e.g., the UK could be expected to de-regulate its economy in the case of Brexit). This biases the setup in favor of instrumental alignment. Conversely, the absence of distance increases after exits offers a strong case for intrinsic norm alignment.

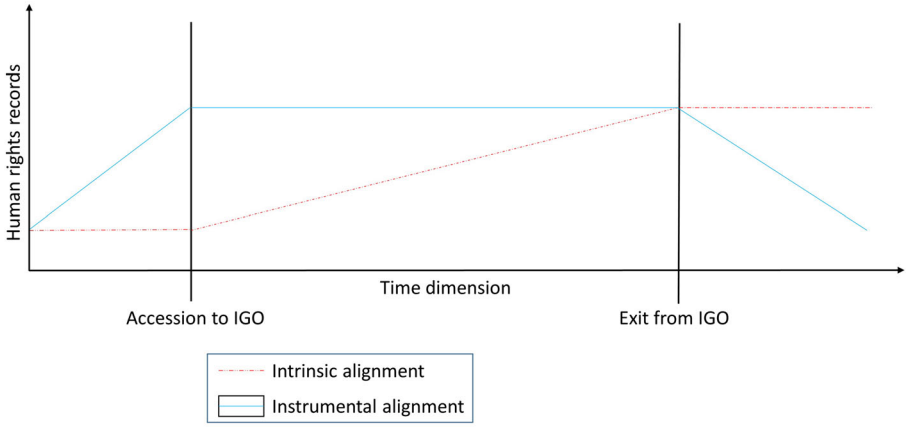

Fig. 1 Stylized pathways of new IGO members' human rights records. The Figure refers only to the subset of states that are originally more repressive than IGO means. We also focus on this subset of states in our empirical analyses

are neither mutually exclusive nor deterministic. It might be the case that alignment was originally a product of instrumental reasons transforming into intrinsic norm adoption over time. We further acknowledge that motives for convergence are unlikely to be purely instrumental or intrinsic but rather a mixture of both. Being aware of these caveats, we suggest that the timing of norm alignment across IGOs provides suggestive evidence about which type of motive dominates. To enhance the credibility of our findings, we further leverage information about the types of physical integrity rights violations.

4.2 Strategic shifts to 'clandestine' repression

Governments make strategic decisions about the repressive tools they use to forward their goals (see Tilly, 2003). To avoid reputation costs, governments strategically use certain types of human rights violations to offset restraint on other repressive means (Hafner-Burton, 2008). Yet, previous studies of human rights alignment utilize an aggregate measure of physical integrity violations including extrajudicial killings, torture, political imprisonments, and disappearances to study human rights alignment in IGOs.[13] This additive index is drawn from the Cingranelli & Richards Human Rights Data Project and assigns equal weight to each physical integrity right violation (Cingranelli et al., 2014). The underlying assumption is that they are equally costly in terms of human rights shaming.

Drawing on the human rights literature, we argue that this assumption is problematic. Political imprisonments are associated with high reputation costs as they are frequently perceived to be the direct result of deliberate policy choices of governments (Bell et al., 2013; Bell and Chernykh, 2019). In contrast, it has been argued that

[13] Physical integrity rights are a subset of civil and political human rights, also called first-generation human rights, as protected in the International Covenant on Civil and Political Rights.

disappearances are less costly for governments since it is more difficult to tie them to the regime (Payne and Abouharb, 2016).

Disappearances allow governments to reduce accountability since their secret nature makes it difficult to identify the perpetrators. In several cases disappearances have been blamed on organized crime even though governments had been involved (see Muñoz, 2011). This illustrates that repressive governments have incentives to strategically shift to disappearances when they fear external pressure. Kinzelbach and Spannagel (n.d.) highlight, for instance, how Argentina's repressive apparatus shifted in the 1970s from political imprisonments to disappearances in light of international pressure for the release of political prisoners.

Against this backdrop, we argue that human rights alignment in IGOs might come at the expense of rising disappearances if it is instrumentally motivated. Superior repressive records might obscure a strategic shift from salient violence to clandestine violence. Such a substitution might allow governments to balance strategic interests in repression with instrumentally motivated norm alignment. If human rights alignment derives from intrinsic motives, governments have no reason to prioritize certain physical integrity violations over others. Therefore, we expect consistent reductions of physical integrity rights violations instead of systematic substitution dynamics.[14] Within this context, we argue that increases in disappearances suggest instrumental norm alignment in IGOs, whereas an improvement across all types of physical integrity rights violations would cast doubts upon a mechanism reliant on strategic motives. We test instrumental norm alignment with the following hypothesis:

> H_4: If states align human rights norms for instrumental reasons, we expect increased levels of disappearances.

To summarize, we expect that instrumental motives find expression in norm alignment shortly before IGO membership, in reversion processes once IGO membership is terminated, and in strategic shifts toward disappearances. Intrinsic motives are expected to be reflected in gradual norm convergence over the time of IGO membership and in elevated human rights records beyond IGO membership. We recognize that none of the suggested hypotheses provides conclusive evidence for instrumental or intrinsic motives as drivers of norm alignment. Ultimately, these mechanisms remain inherently unobservable. However, an interplay of these arguments provides suggestive evidence which motive dominates on average.

5 Research design

We study the impact of IGO membership on the human rights records of states in a sample of 73 IGOs between 1965 and 2014. Our unit of analysis is an 'IGO-country-year' capturing member states in IGOs on a yearly basis. Information on state

[14] We acknowledge that some violations are more difficult to tackle than others, particularly if they are agent-centered (see Mitchell, 2004). While this could imply staggered improvement processes, we have no reason to expect systematic increases of disappearances in the case of intrinsic norm adoption.

membership in IGOs is derived from the Correlates of War International Governmental Organizations Dataset (Pevehouse et al., 2004).

Our sample of IGOs builds on the study from Hooghe and Marks (2015) that codes systematic differences between IGOs. They select their sample of IGOs based on the following criteria: IGOs are selected if they have a distinct physical location or website, a formal structure, at least 50 permanent staff, a decision body that meets at least once a year, and a written constitution or convention. Hence, the scope conditions of our analysis apply to major IGOs with a certain degree of institutionalization. The sample covers a broad range of both regional and global organizations from various different policy areas.[15]

Our sample of countries includes only those states that (a) became IGO members in the observation period and that (b) had worse human rights records before membership than the average of the respective IGO. Hence, we focus our empirical analysis exclusively on the alignment process of new member states that were originally more repressive than IGO averages. While it is also possible that originally less repressive states adapt to inferior human rights standards, we leave the study of this reversed process up to future research.

5.1 Dependent variables

Our main dependent variable is the relative distance in human rights scores between new member states and IGO means. Per definition, new member states are only considered if they had worse human rights records before membership than the IGO mean. IGO means proxy for IGO-specific 'human rights cultures' and they are defined as the average human rights scores across all members in an IGO.[16] Since we focus on the subset of states that had originally worse human rights scores than IGO averages, decreasing distances imply convergence to superior human rights standards.[17] Hence, we expect alignment toward IGO means during IGO membership.

We rely on Fariss' (2014) human rights data that combines various standards-based and events-based datasets and accounts for changes in human rights monitoring and standards of accountability over time. In so doing, Fariss (2014) provides a continuous measure of human rights compliance whereby rising values indicate increasing levels of respect for human rights.

As a robustness test, we re-run our analyses with the physical integrity rights index from the CIRI Human Rights data project (Cingranelli et al., 2014). The physical integrity rights index quantifies human rights compliance on an eight-point scale based on annual country reports from Amnesty International and the U.S. State Department.

To test potential differences in alignment patterns across different physical integrity rights, we rely on the constituent indicators of the physical integrity rights index. These

[15] We present an overview of all included IGOs in the online Appendix.

[16] We exclude the respective states whose distance is studied from the calculation of the group mean.

[17] Note that originally more repressive states might also improve beyond IGO means. Since this also implies that they adopt superior human rights norms, we measure relative improvement in relation to the group mean instead of absolute distances.

19 Additionally, we account for reversed causality by testing our models with differentiated dependent variable defined as the yearly change in human rights distances for each state-IGO combination. The results remain robust.

are extrajudicial killings, torture, disappearances, and political imprisonments. We are especially interested in disappearances defined as "cases in which people have been disappeared, agents of the state are likely responsible, and political motivation may be likely" (Cingranelli et al., 2014, p. 12). State respect for the right not to be disappeared is compared with the rights not to be killed, tortured, or imprisoned for political reasons. All four physical integrity rights are measured with three-step ordinal variables, whereby higher values indicate more respect for the respective right. To avoid reversed causality, we lagged all our outcome variables by one year.[19]

5.2 Explanatory variables

Our key explanatory variable is the duration of state membership in IGOs. Since we are interested in the temporal dimension of IGO membership, we created a count variable that indicates the number of membership years for each state in an IGO. We analyze, for each individual state-IGO combination, how the duration of membership is related to the relative distance in human rights scores to IGO means.

In addition, we are interested in the time periods before and after IGO membership. Therefore, we created two additional variables indicating the number of years before or respectively after IGO membership for each state. These variables allow us to investigate whether human rights distances between states and IGOs systematically changed in the pre-membership or post-membership period.[18]

5.3 Control variables

We include two types of control variables that could confound the relationship between IGO membership and respect for human rights: (a) characteristics of states and (b) characteristics of IGOs.

We account for states' formal commitments to human rights as indicated by the ratification of the International Covenant on Civil and Political Rights (ICCPR). While the causal mechanisms remain contested, previous research suggests that ICCPR ratification is linked to reduced respect for physical integrity rights (Hill Jr., 2010; Neumayer, 2005). Further, ICCPR ratification could correlate with IGO membership since it might capture variation in the willingness of governments to comply with international standards. We control for *ICCPR ratification* with a binary variable coded for each state-year with data drawn from the United Nations High Commissioner for Human Rights (UN OHCHR, 2020). We control for ongoing *civil wars* with data from the UCDP Armed Conflict Dataset (Harbom et al., 2008) since conflicts tend to be linked to human rights violations and correlated with membership in IGOs (Poe and Tate, 1994; E. J. Wood, 2006). Previous studies have identified a lack of state capacity as a key driver of human rights abuse (Englehart, 2009; Hafner-Burton, 2014). Further, states might differ in their capacity to fulfil membership criteria in IGOs. We control for

[18] IGO exits are relatively rare. In total, our sample includes 36 cases of IGO exits as presented in the Online Appendix.

state capacity with GDP per capita obtained from the World Development Indicators as provided by the World Bank (The World Bank, 2018).[19]

Turning to IGO characteristics, we control for the level of *delegation* as classified by Hooghe and Marks (2015). They define delegation as the ability of the secretariat of an IGO to take action in different issue areas. The delegation measure is a proxy for IGOs' degree of authority and accounts for systematic differences in the capacity of IGOs to exert pressure on member states in the context of human rights compliance. We account for the *membership size* of IGOs since socializing dynamics could systematically differ depending on the size of IGOs. We control for *human rights IGOs* to test whether human rights alignment is driven by this subset of IGOs. Further, we control for the annual *change in the average human rights scores* of IGOs. This allows us to preclude that potential effects are due to moving averages in IGOs instead of improving human rights records of new members. Additionally, we control for *years* to control for systematic changes in human rights over time. Summary statistics for all variables included in our models are presented in the online Appendix.

5.4 Model specification

We use varying-slopes, varying-intercepts models to test our hypotheses. These models are especially well-suited for our data structure because they allow for variation in the intercept (i.e., the IGO-state combination) and the slope (i.e., the duration of state membership in an IGO) (Gelman and Hill, 2006). Further, they allow us to control for confounders at the IGO-level and at the country-level.

The dependent variable is continuous since it calculated by subtracting IGO means from states' individual human rights scores. IGO means are calculated by averaging the human rights scores of all members measured in integers leading to values with decimals. We estimate varying-intercept and varying-slopes OLS regressions. The varying slope in these models is an interaction between the continuous predictor time and the grouping indicator "case ID" that consists of every state-IGO combination in the dataset.

Our key explanatory variable are the years of membership in an IGO. We treat the years of membership as a continuously increasing count, the first year of membership corresponds to a 1 in the dataset, the second membership year to a 2, and so forth. We formulate our model as proposed in Gelman and Hill (2006):

$$Distance_i = \alpha_{caseID[i]} + \beta_{caseID[i]} duration_i + \beta_n X_i + \epsilon_i$$

6 Results

We present the results in the following order: first, we show the results of our analysis modelling human rights alignment during IGO membership. Second, we present the

[19] Hendrix (2010) suggests that taxes/GDP is theoretically more suitable to capture state capacity. Hence, we additionally run our models with this measure. Due to substantial numbers of missing observation, we stick to GDP per capita in our main models acknowledging that it represents only a distant proxy for state capacity.

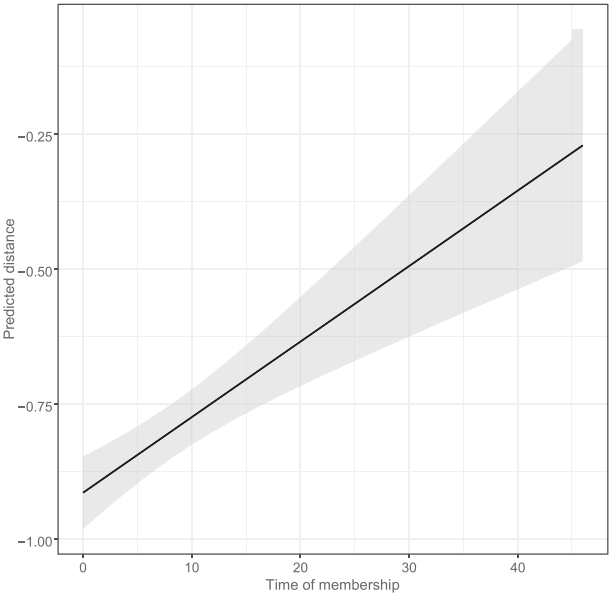

Fig. 2 Predicted distance to the IGO mean during membership

average trends for the human rights distances before and after IGO membership. Subsequently, we show how different physical integrity rights align during IGO membership.

6.1 Human rights alignment during IGO membership

The results pertaining to human rights alignment during IGO membership are presented in Table 2 and summarized in Fig. 2. The key explanatory variable 'duration of membership' is statistically significant at the $p = 0.001$ level. As demonstrated in Fig. 2, the duration of IGO membership is associated with systematic convergence of human rights distances between new member states and IGO means. The longer states are members of IGOs, the closer their human rights scores align with IGO averages. The systematic distance reduction is not due to moving averages of IGOs, since we control for annual changes in IGO means. Hence, the findings imply that states that used to be more repressive improve their human rights records as members of IGOs suggesting socialization into 'human rights cultures'.

The substantive effect sizes are comparatively small which is not surprising given that human rights records are multi-factorial phenomena and IGO membership is but one of many influences.[20] Fariss' human rights scores are constructed from several human rights indices that tend to be relatively stable over time and it has been argued that substantial changes in human rights conditions are necessary for a country to move from one category to another (see Gohdes and Carey, 2017; R. M. Wood and Gibney, 2010). Consequently, the setup tends to be biased against large effects and we are confident that the findings are notable despite relatively small effects.

[20] Fariss' (2014) human rights scores range from −3.1 to 4.7 in our sample.

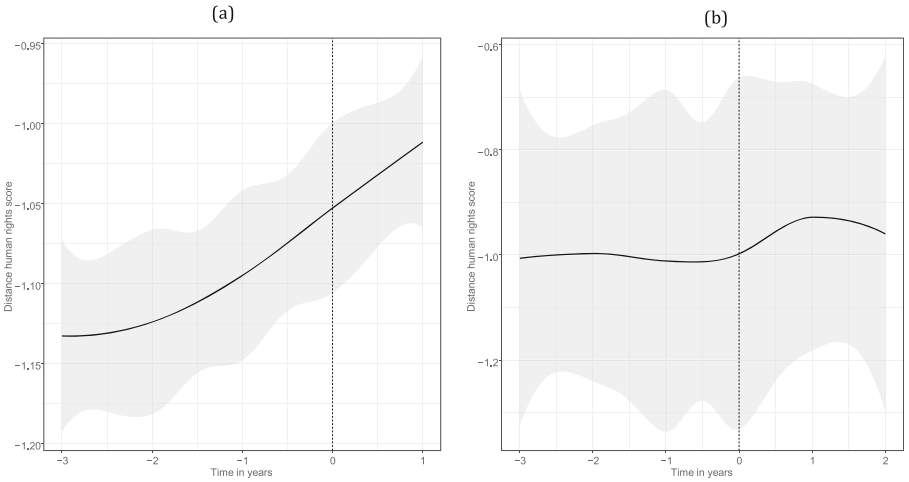

Fig. 3 Distances in human rights scores (LOESS)

6.2 Pre-membership and post-membership alignment

Next, we test whether there is evidence for instrumental alignment by analyzing human rights distances in the times before and after IGO membership. Instrumental motives would suggest that states systematically improve their human rights records in the pre-accession period (hypothesis 1). Further, instrumental motives would imply that states return to their pre-membership human rights records after IGO exits (hypothesis 3).

Figure 3a illustrates the trend of human rights distances before IGO membership averaging across all state-IGO combinations in the sample. The zero-line indicates the time point of IGO accession. The figure demonstrates that human rights distances tend to become smaller before IGO membership. This might reflect an instrumental adaptation of human rights records in response to membership requirements. However, the pre-membership alignment effect remains substantively small. Hence, our findings lend only weak support to the instrumental logic.

Figure 3b shows the trend of human rights distances after IGO membership averaging across all state-IGO combinations in our sample. The zero-line indicates the time point of IGO exit. The Figure demonstrates that there are no systematic changes of human rights distances after IGO exits. On average, states that leave IGOs tend not to use their new independence to increase policy distance in the realm of human rights. The empirical evidence rather suggests that new human rights standards become enshrined in domestic policy and practice. Due to the small sample sizes (in total, 37 cases of IGO exits) resulting in large confidence intervals, this finding must be taken with a grain of salt. Generally, the setup tends to be biased in favor of instrumental alignment. IGO exits result from self-selection dynamics that likely reflect substantial disagreements with remaining members. If a state leaves to escape IGO policy constraints, it is likely to use its newly-gained independence to increase policy distance. Since we do not find evidence for changes of human rights records after IGO exits, this casts doubt to the claim that human rights alignment was only strategically motivated.

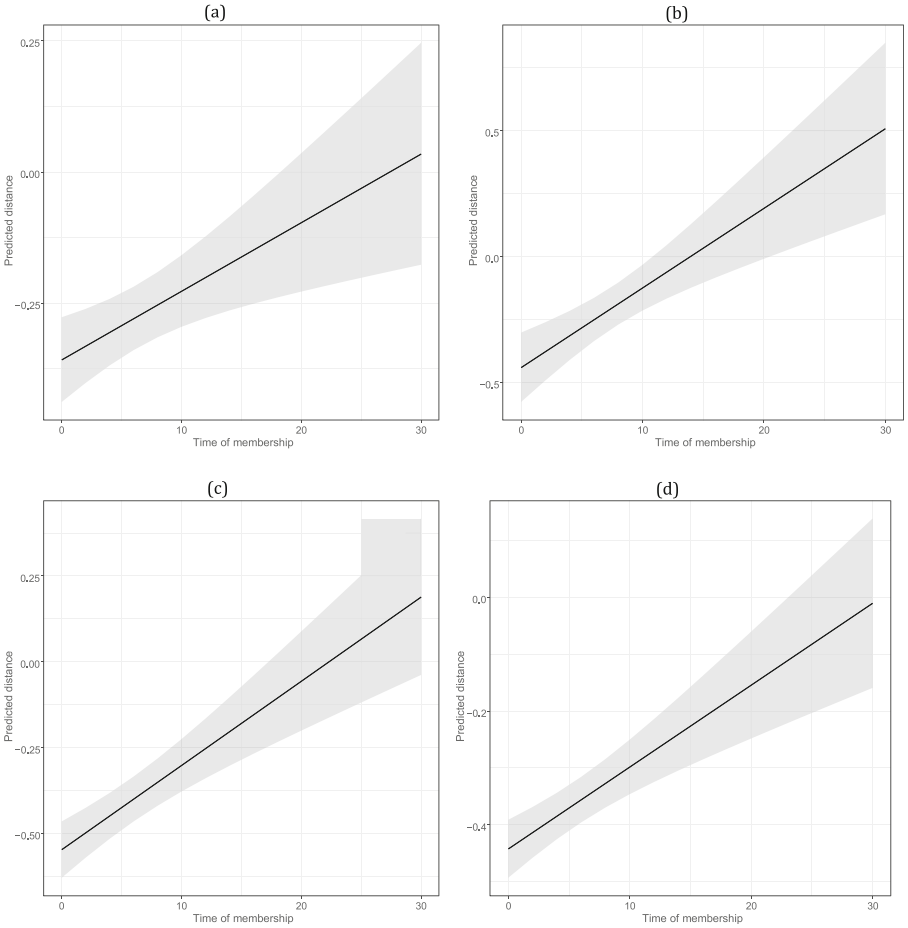

Fig. 4 Alignment of sub-indices during IGO membership, (a) Extrajudicial killings, (b) Disappearances, (c) Political Imprisonment, (d) Torture

6.3 Alignment of different physical integrity rights

Finally, we investigate whether human rights alignment during IGO membership differs systematically across different physical integrity rights. The instrumental logic suggests that states strategically shift to disappearances (hypothesis 4). The results are presented in Table 3 and visualized in Fig. 4.

The findings indicate that the distances on the indices of all physical integrity rights violations tend to decrease during IGO membership. Most importantly, state respect for the right not to be disappeared systematically aligns with IGO group means as a function of membership years in IGOs. Thus, our empirical evidence runs counter to the argument that states shift to disappearances to balance strategic interests in repression with instrumentally motivated norm alignment. Originally more repressive states tend to improve all physical integrity rights as they become members of IGOs with superior 'human rights cultures'. This casts doubt to the

Table 1 Overview of hypotheses

Hypotheses supporting:	
Instrumental motives	Intrinsic motives
H1	H2
H3	
H4	

claim that human rights alignment in IGOs comes at the expense of clandestine repression.

Table 2 Varying-slopes, varying-intercepts OLS regression model

	Distance in Fariss' Human Rights Scores
Intercept	−0.76***
	(0.04)
Duration of Membership	0.01***
	(0.00)
Year	−0.06*
	(0.03)
GDP per Capita	0.38***
	(0.04)
Ratification ICCPR	0.07***
	(0.02)
Civil War	−0.27***
	(0.01)
Delegation	−0.06***
	(0.02)
Human Rights IGO	−0.04
	(0.04)
Membership Size	0.11***
	(0.02)
Change in Mean HR – Fariss	−1.12***
	(0.07)
AIC	19,152.45
BIC	19,260.87
Log Likelihood	−9562.22
Num. obs.	17,056
Num. groups: IGO-Country	848
Var: IGO-Country (Intercept)	0.60
Var: IGO-Country Duration	0.00
Cov: IGO-Country (Intercept) Duration	−0.02
Var: Residual	0.13

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3 Varying-slopes, varying-intercepts models

Distance in:	Disappearances	Extrajudicial Killings	Torture	Pol. Imprisonment
Intercept	−0.54***	−0.31***	−0.25***	−0.42***
	(0.12)	(0.06)	(0.03)	(0.05)
Duration of Membership	0.03***	0.01**	0.01***	0.02***
	(0.01)	(0 00)	(0 00)	(0.00)
Year	−0.10	−0.12*	−0.13***	−0.28***
	(0.08)	(0.05)	(0.03)	(0.05)
GDP per Capita	−0.12	0.26**	0.31***	0.07
	(0.17)	(0.09)	(0.04)	(0.06)
Ratification ICCPR	0.11*	0.17***	0.04	0.15***
	(0.05)	(0.03)	(0.03)	(0.04)
Civil War	−0.25***	−0.19***	−0.11***	−0.15***
	(0.04)	(0.03)	(0.02)	(0.03)
Delegation	0.01	−0.03	−0.01	0.01
	(0.03)	(0.02)	(0.01)	(0.03)
Human Rights IGO	0.21	0.15	−0.00	0.04
	(0.16)	(0.10)	(0.06)	(0.09)
Membership Size	0.15***	0.14***	0.08***	0.16***
	(0.04)	(0.03)	(0.02)	(0.03)
Change in Mean HR - Disappearances	−0.50*** (0.09)			
Change in Mean HR – Killings		−0.50*** (0.06)		
Change in Mean HR - Torture			−0.43*** (0.06)	
Change in Mean HR - Imprisonment				−0.52*** (0.06)
AIC	5058.94	7828.56	6088.46	7789.27
BIC	5141.82	7919.05	6179.33	7879.69
Log Likelihood	−2515.47	−3900.28	−3030.23	−3880.63
Num. obs.	2751	4742	4869	4715
Num. groups: IGO-Country	190	362	391	363
Var: IGO-Country (Intercept)	0.38	0.26	0.10	0.25
Var: IGO-Country Duration	0.00	0.00	0.00	0.00
Cov: IGO-Country Duration	−0.03	−0.01	−0.00	−0.00
Var: Residual	0.28	0.24	0.17	0.23

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

7 Robustness tests

We conduct several robustness tests to demonstrate that our results hold across diverse model specifications. First, we re-run the main analysis of human rights alignment during IGO membership replacing Fariss' human rights scores with the CIRI physical

integrity rights index. The results presented in Table 1 in the Online Appendix show that our findings are robust to this different measurement of our main dependent variable. Subsequently, we split the sample of IGOs in human rights and non-human rights IGOs and re-run our analysis for both subgroups as presented in Table 2 in the Online Appendix. The results show that membership in human rights IGOs has no significant impact on human rights distances while membership in non-human rights IGOs is associated with significant alignment processes. This supports our key assumption that human rights cultures are shaped by membership structures in IGOs and not by their policy scopes. This also aligns with our additional analysis presented in Table 3 in the Online Appendix in which we compare the effect in political, economic, and social IGOs. While the duration of membership seems to be related to reducing human rights distances in all types of IGOs, the effect fails to reach significance in political IGOs. Finally, we test whether regionalism confounds our findings. IGOs tend to form on a regional basis and geographic proximity could likewise affect the diffusion of human rights norms (Edwards et al., 2018). Hence, we run another model where we control for geographic proximity with a binary measures for different regional groupings in the United Nations (UN DGACM, 2020). The findings remain robust as demonstrated in Table 4 in the Online Appendix.

8 Discussion

Our results support the previously established finding that IGOs provide a conduit for the diffusion of human rights norms (Greenhill, 2010, 2016). We are able to show that human rights alignment takes place during membership in IGOs and that the effect is consistent across different types of physical integrity rights. Confirming hypothesis 2, we find that the longer states are members of IGOs, the closer their human rights records align with the average of the other IGO members. This finding applies across a broad range of IGOs and is not driven by the subset of human rights IGOs. This supports our central argument that 'human rights cultures' in IGOs are determined by the membership structures of IGOs and not by their policy scopes.

The empirical findings tend to contradict our theoretical expectations related to instrumentally motivated norm alignment. We find only weak evidence for alignment in the years before IGO membership. The instrumental logic would suggest that human rights alignment occurs primarily in the pre-membership period while no systematic trends occur during IGO membership. Further, we do not find evidence that states revert to original repressive records once they exit IGOs. While exit cases are relatively rare, our results offer tentative support that repressive records remain largely unaffected.[21] Finally, the absence of strategic shifts to clandestine forms of repression suggests that states do not only adjust high-salience human rights with co-members in IGOs.

Does this imply that human rights alignment is driven by intrinsic motivations? The underlying motivations of human rights alignment remain unobservable and our study cannot provide conclusive evidence for intrinsic motivations. Being aware of this

[21] Arguably, our test is lenient toward instrumental mechanisms since we only expect states to revert to pre-membership human rights records. The European Union literature even suggests backsliding processes once the incentives for compliance are gone (e.g., Rupnik, 2007).

inherent limitation, our results cast doubt on claims of prior studies that primarily strategic motivations drive human rights (see Greenhill, 2010, 2016). Several observable implications that might indicate instrumental motives find no empirical support in the aggregate of state-IGO dyads. While instrumental motives might lead in some cases to alignment during IGO membership, the empirical pattern rather fits to the logic of intrinsic preference transformations that result from repeated interactions (Gheciu, 2005; Finnemore, 1996). Hence, it is likely that human rights alignment in IGOs is at least not exclusively driven by instrumental motives.

An alternative explanation for the alignment of human rights records in IGOs could be that states with similar identities self-select into the same organizations. In other words, norm alignment would not occur due to influences exerted by other group members but because of similar identities driving group formations. Such a homophily-effect would bias the findings of our study since in that case none of the mechanisms would causally explain norm convergence. If self-selection would be the key driver of norm convergence, factors endogenous to individual states would shape human rights alignment. Being aware of this caveat, we are confident that homophily dynamics do not bias our findings. Greenhill (2016: chapter 6) provides extensive empirical evidence that IGOs tend to form on the basis of regional proximity. We account for the confounding effect of regional proximity with controls for the different regional groups and our effects remain constant. This suggests that our results are not a product of self-selection but driven by network effects that tend to take intrinsic shapes.

9 Conclusion

Do countries change their domestic human rights records when they become members of IGOs and if so, why? Our empirical analysis has demonstrated that states tend to align their human rights records with other member states in IGOs during their membership. By levering the timing of human rights alignment and variation across physical integrity rights, we developed tentative empirical support that this process is at least partially motivated by intrinsic factors. Our evidence takes the form of an *argumento a contrario* since we rule out the empirical implications of the common proposition that instrumental motives drive norm alignment (Goodman and Jinks, 2008, Greenhill, 2016). Neither do states significantly align their repressive records before IGO membership nor do they shift to repression that is less attributable to the government. Instead, we find evidence that states gradually align their human rights records during IGO membership and that the exposure to different human rights norms leaves traces beyond membership in IGOs.

Our research contributes to the literature on norm diffusion in IGOs (e.g., Bearce and Bondanella, 2007; Greenhill and Lupu, 2017; Taninchev, 2015). First, we generate observable implications allowing us to disentangle intrinsic from instrumental motives for norm alignment. This has important implications for constructivist scholarship that has been frequently criticized because it cannot be tested empirically (e.g., Kacowicz, 2005: 181). Even constructivist thinkers themselves have attested constructivist scholarship a certain degree of "empirical ad-hocism" (Checkel, 1998: 325). Our study provides a new way to test constructivist arguments against observable empirical

patterns. We hope to inspire similar research approaches that develop creative empirical tests for constructivist tenets.

Second, we apply a new modeling strategy to analyze the effects of state membership in multiple IGOs on state behavior. State membership in IGOs over time warrants a triadic data structure that captures each individual state-IGO combination in a given year. By modeling individual triads and accounting for IGO-level and state-level confounders, we offer a new empirical approach to study human rights alignment in IGOs. Further, we contribute to International Relations theory more generally by identifying a key mechanism for norm diffusion between states. We offer empirical evidence that rising similarity or 'isomorphism' (e.g., Finnemore, 1996; Meyer et al., 1997) between states and societies is at least partially shaped by co-membership in IGOs. Our study also contributes to the literature studying the effectiveness of external means to promote human rights (e.g., Kelley and Simmons, 2015; Peksen, 2009). Human rights compliance tends to be regarded as a high-salience domestic issue that is not easily susceptible to 'soft' international influence. We show that international influence in form of co-membership in IGO has the potential to shape domestic human rights records.

It is important to highlight some further limitations of our study. We offer a large-N perspective on the timing of inter-state alignment of different physical integrity rights as a function of IGO membership. Our study does not provide empirical evidence on the micro-level processes explaining how exactly altered preferences of state representatives in IGOs affect domestic policy-making. Given that an extensive analysis of these complex micro-level dynamics goes beyond the scope of our study, we suggest that our findings are evaluated in interplay with previous studies that shed light on these processes (e.g., Gheciu, 2005; Greenhill, 2016). We also acknowledge that our study cannot provide conclusive evidence for intrinsic motivations. While the large-N pattern speaks rather in favor of intrinsic motives, we cannot prove the internalization of new normative contents but only contradict certain observable implications of instrumental norm alignment. Finally, it must be emphasized that our study assumes that governments are able to reduce repression if they are willing to do so. However, government control over the repressive apparatus is no determinism in light of self-interests of coercive agents for human rights violations (Butler et al., 2007).

Being aware of these caveats, we are confident that our study adds an important facet to our understanding of norm alignment processes in IGOs. It is well-established that peer groups have a substantial influence on the behavior of its members (e.g., Fortuin et al., 2016; Simons-Morton and Farhat, 2010). Our study demonstrates that this insight from everyday social interactions also applies to the international arena. This finding has important policy implications. Given that co-membership in IGOs seems to have the capacity to shape governments' repressive actions, inclusive membership criteria might represent an effective means to exert leverage on repressive governments. Beyond material benefits of cooperation, policy-makers should interpret co-membership in IGOs as a channel of dialogue and influence. While affecting human rights practices of isolated states is chronically challenging, governments possess a powerful channel of influence if they share membership in IGOs.

This study demonstrated that IGOs provide a platform for alignment processes of physical integrity rights and suggested that these dynamics are driven by the internalization of new normative contents. In essence, the results mirror key constructivist arguments suggesting that states are fundamentally social actors with malleable identities.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11558-021-09413-5.

Funding This work was supported by the the University of Mannheim's Graduate School of Economic and Social Sciences (GESS). Open Access funding enabled and organized by Projekt DEAL.

References

Bearce, D. H., & Bondanella, S. (2007). Intergovernmental organizations, socialization, and member-state interest convergence. *International Organization, 61*(4), 703–733.

Beckfield, J. (2003). Inequality in the world polity: The structure of international organization. *American Sociological Review 68*, (3), 401–424.

Bell, S. R., & Chernykh, S. (2019). Human rights violations and post-election protest. *Political Research Quarterly, 72*(2), 460–472.

Bell, S. R., Cingranelli, D., Murdie, A., & Caglayan, A. (2013). Coercion, capacity, and coordination: Predictors of political violence. *Conflict Management and Peace Science, 30*(3), 240–262.

Boli, J., & Thomas, G. M. (1999). *Constructing world culture: International nongovernmental organizations since 1875*. Palo Alto: Stanford University Press.

Butler, C. K., Gluch, T., & Mitchell, N. (2007). Security forces and sexual violence: A cross-national analysis of a principal—Agent argument. *Journal of Peace Research, 44*(6), 669–687.

Cao, X. (2009). Networks of intergovernmental organizations and convergence in domestic economic policies. *International Studies Quarterly, 53*(4), 1095–1130.

Checkel, J. T. (1998). The constructivist turn in international relations theory. *World Politics, 50*(2), 324–348.

Checkel, J. T. (2005). International institutions and socialization in Europe: Introduction and framework. *International Organization, 59*(4), 801–826.

Chelotti, N., Dasandi, N., & Mikhaylov, S. J. (2018). Do intergovernmental organizations have a socialization effect on member state preferences? Evidence from the UN general debate. Unpublished Manuscript.

Cingranelli, D. L., Richards, D. L., Clay, & Chad, K. (2014). The CIRI human rights dataset. http://www.humanrightsdata.com

DeMeritt, J. H. R. (2012). International organizations and government killing: Does naming and shaming save lives? *International Interactions, 38*(5), 597–621.

Donno, D., & Neureiter, M. (2018). Can human rights conditionality reduce repression? Examining the European Union's economic agreements. *The Review of International Organizations, 13*(3), 335–357.

Dorussen, H., & Ward, H. (2008). Intergovernmental organizations and the Kantian peace: A network perspective. *Journal of Conflict Resolution, 52*(2), 189–212.

Edwards, T. H., Kernohan, D., Landman, T., & Nessa, A. (2018). Good neighbours matter: Economic geography and the diffusion of human rights. *Spatial Economic Analysis, 13*(3), 319–337.

Englehart, N. A. (2009). State capacity, state failure, and human rights. *Journal of Peace Research, 46*(2), 163–180.

Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review, 108*(2), 297–318.

Finnemore, M. (1993). International organizations as teachers of norms: The United Nations educational, scientific, and cultural organization and science policy. *International Organization, 47*(4), 565–597.

Finnemore, M. (1996). Norms, culture, and world politics: Insights from sociology's institutionalism. *International Organization, 50*(2), 325–347.

Finnemore, M., & Sikkink, K. (1998). International norm dynamics and political change. *International Organization, 52*(4), 887–917.

Fortuin, J., van Geel, M., & Vedder, P. (2016). Peers and academic achievement: A longitudinal study on selection and socialization effects of in-class friends. *The Journal of Educational Research, 109*(1), 1–6.

Freedman, R., & Houghton, R. (2017). Two steps forward, one step back: Politicisation of the human rights council. *Human Rights Law Review, 17*(4), 753–769.

Freyburg, T. (2015). Transgovernmental networks as an apprenticeship in democracy? Socialization into democratic governance through cross-national activities. *International Studies Quarterly, 59*(1), 59–72.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gheciu, A. (2005). Security institutions as agents of socialization? NATO and the 'new Europe'. *International Organization, 59*(4), 973–1012.

Gilardi, F. (2016). Four ways we can improve policy diffusion research. *State Politics & Policy Quarterly, 16*(1), 8–21.

Gilardi, F., & Wasserfallen, F. (2019). The politics of policy diffusion. *European Journal of Political Research, 58*(4), 1245–1256.

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

Goodman, R., & Jinks, D. (2008). Incomplete internalization and compliance with human rights law. *European Journal of International Law, 19*(4), 725–748.

Goodman, R., & Jinks, D. (2013). *Socializing states: Promoting human rights through international law*. Oxford: Oxford University Press.

Greenhill, B. (2010). The company you keep: International socialization and the diffusion of human rights norms. *International Studies Quarterly, 54*(1), 127–145.

Greenhill, B. (2016). *Transmitting rights: International organizations and the diffusion of human rights practices*. Oxford: Oxford University Press.

Greenhill, B., & Lupu, Y. (2017). Clubs of clubs: Fragmentation in the network of intergovernmental organizations. *International Studies Quarterly, 61*(1), 181–195.

Greitens, S. C. (2016). *Dictators and their secret police: Coercive institutions and state violence*. Cambridge: Cambridge University Press.

Haas, E. B. (1980). Why collaborate? Issue-linkage and international regimes. *World Politics, 32*(3), 357–405.

Hafner-Burton, E. M. (2008). Sticks and stones: Naming and shaming the human rights enforcement problem. *International Organization, 62*(4), 689–716.

Hafner-Burton, E. M. (2014). A social science of human rights. *Journal of Peace Research, 51*(2), 273–286.

Harbom, L., Melander, E., & Wallensteen, P. (2008). Dyadic dimensions of armed conflict, 1946—2007. *Journal of Peace Research, 45*(5), 697–710.

Harnisch, S., Frank, C., & Maull, H. W. (2011). *Role theory in international relations*. Milton Park: Taylor & Francis.

Hendrix, C. S. (2010). Measuring state capacity: Theoretical and empirical implications for the study of civil conflict. *Journal of Peace Research, 47*(3), 273–285.

Hill Jr., D. W. (2010). Estimating the effects of human rights treaties on state behavior. *The Journal of Politics, 72*(4), 1161–1174.

Hofstede, G. (2004). Diplomats as cultural bridge builders. In H. Slavik (Ed.), *Intercultural communication and diplomacy* (pp. 25–38). Geneva: Diplo Foundation.

Holm, T. T., & Eide, E. B. (2000). *Peacebuilding and police reform*. London: Psychology Press.

Hooghe, L., & Marks, G. (2015). Delegation and pooling in international organizations. *The Review of International Organizations, 10*(3), 305–328.

Johnston, A. I. (2001). Treating international institutions as social environments. *International Studies Quarterly, 45*(4), 487–515.

Kacowicz, A. M. (2005). *The impact of norms in international society: The Latin American experience, 1881–2001*. Notre Dame, Indiana: University of Notre Dame Press.

Keck, M. E., & Sikkink, K. (2014). *Activists beyond borders: Advocacy networks in international politics*. Ithaca, New York: Cornell University Press.

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

Kim, M. (2012). Ex ante due diligence: Formation of PTAS and protection of labor rights. *International Studies Quarterly, 56*(4), 704–719.

Kinzelbach, K., & Spannagel, J. (n.d.). *Human rights advocacy data revisited: Why and how amnesty international selectively underreports cases of political imprisonment*. Unpublished Manuscript.

Klotz, A. (1995). Norms reconstituting interests: Global racial equality and US sanctions against South Africa. *International Organization, 49*(3), 451–478.

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology, 115*(2), 405–450.

Lee, C. K., & Strang, D. (2006). The international diffusion of public-sector downsizing: Network emulation and theory-driven learning. *International Organization, 60*(4), 883–909.

Lewis, J. (2005). The Janus face of Brussels: Socialization and everyday decision making in the European Union. *International Organization, 59*(4), 937–971.

Lindemann, T., & Petiteville, F. (2019). Norm diffusion in international relations: The case of human rights and humanitarian norms. In L. Delcour & E. Tulmets (Eds.), *Policy transfer and norm circulation* (pp. 185–204). London: Routledge.

March, J. G., & Olsen, J. P. (1998). The institutional dynamics of international political orders. *International Organization, 52*(4), 943–969.

McGinnis, M. D. (1986). Issue linkage and the evolution of international cooperation. *Journal of Conflict Resolution, 30*(1), 141–170.

Meyer, J. W., Boli, J., Thomas, G. M., & Ramirez, F. O. (1997). World society and the nation-state. *American Journal of Sociology, 103*(1), 144–181.

Mitchell, N. (2004). *Agents of atrocity: Leaders, followers, and the violation of human rights in civil war*. London: Palgrave Macmillan.

Muñoz, A. A. (2011). Explaining high levels of transnational pressure over Mexico: The case of the disappearances and killings of women in Ciudad Juárez. *The International Journal of Human Rights, 15*(3), 339–358.

Neumayer, E. (2005). Do international human rights treaties improve respect for human rights? *Journal of Conflict Resolution, 49*(6), 925–953.

Paul, T. V., Larson, D. W., & Wohlforth, W. C. (2014). *Status in world politics*. Cambridge: Cambridge University Press.

Payne, C. L., & Abouharb, M. R. (2016). The international covenant on civil and political rights and the strategic shift to forced disappearance. *Journal of Human Rights, 15*(2), 163–188.

Peksen, D. (2009). Better or worse? The effect of economic sanctions on human rights. *Journal of Peace Research, 46*(1), 59–77.

Pevehouse, J., Nordstrom, T., & Warnke, K. (2004). The correlates of war 2 international governmental organizations data version 2.0. *Conflict Management and Peace Science, 21*(2), 101–119.

Poe, S. C., & Tate, C. N. (1994). Repression of human rights to personal integrity in the 1980s: A global analysis. *American Political Science Review, 88*(4), 853–872.

Risse, T. (2000). "Let's argue!": Communicative action in world politics. *International Organization, 54*(1), 1–39.

Rupnik, J. (2007). Is east-Central Europe backsliding? From democracy fatigue to populist backlash. *Journal of Democracy, 18*(4), 17–25.

Simmons, B. A., Dobbin, F., & Garrett, G. (2006). Introduction: The international diffusion of liberalism. *International Organization, 60*(4), 781–810.

Simons-Morton, B. G., & Farhat, T. (2010). Recent findings on peer group influences on adolescent smoking. *The Journal of Primary Prevention, 31*(4), 191–208.

Taninchev, S. B. (2015). Intergovernmental organizations, interaction, and member state interest convergence. *International Interactions, 41*(1), 133–157.

Tannenwald, N. (2007). *The nuclear taboo: The United States and the non-use of nuclear weapons since 1945*. Cambridge: Cambridge University Press.

Terman, R., & Voeten, E. (2018). The relational politics of shame: Evidence from the universal periodic review. *The Review of International Organizations, 13*(1), 1–23.

The World Bank. (2018). World development indicators: GDP per capita (in constant 2000 US dollars). https://data.worldbank.org/indicator/ny.gdp.pcap.cd

Tilly, C. (2003). *The politics of collective violence*. Cambridge: Cambridge University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford: Basil Blackwell.
UN DGACM. (2020). United Nations department for general assembly and conference management: United Nations regional groups of member states. https://www.un.org/depts/DGACM/RegionalGroups.shtml
UN OHCHR. (2020). UN treaty body database. https://tinyurl.com/y3pw2p9o
Volden, C., Ting, M. M., & Carpenter, D. P. (2008). A formal model of learning and policy diffusion. *American Political Science Review, 102*(3), 319–332.
von Stein, J. (2005). Do treaties constrain or screen? Selection bias and treaty compliance. *American Political Science Review, 99*(4), 611–622.
Wendt, A. (1999). *Social theory of international politics*. Cambridge: Cambridge University Press.
Weyland, K. (2012). The Arab spring: Why the surprising similarities with the revolutionary wave of 1848? *Perspectives on Politics, 10*(4), 917–934.
Wood, E. J. (2006). Variation in sexual violence during war. *Politics and Society, 34*(3), 307–342.
Wood, R. M., & Gibney, M. (2010). The political terror scale (PTS): A re-introduction and a comparison to CIRI. *Human Rights Quarterly, 32*, 367–400.