

Hidden figures: how legal experts influence the design of international institutions

European Journal of
International Relations
1–26

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13540661231210931

journals.sagepub.com/home/ejt



Nicole De Silva 
Concordia University, Canada

Anne Holthoefer
Saint Anselm College, USA

Abstract

Whose preferences influence the design of international institutions? Scholarship on the legalization of international politics and creation of international legal institutions largely adopts a state-centric perspective. Existing accounts, however, fail to recognize how states often delegate authority over institutional design tasks to independent legal experts whose preferences may diverge from those of states. We develop a principal–agent (PA) framework for theorizing relations between states (collective principals) and legal actors (agents) in the design process, and for explaining how legal actors influence the design of international institutions. The legal dimensions of the PA relationship increase the likelihood of preference divergence between the collective principal and the agent, but also create conditions that enable the agent to opportunistically advance its own design preferences. We argue that the more information on states' preferences the agent has, the more effectively it can exploit its legal expertise to strategically select and justify design choices that maximize its own preferences and the likelihood of states' acceptance. Our analysis of two cases of delegated institutional design concerning international criminal law at the United Nations and the African Union supports our theoretical expectations. Extensive archival and interview data elucidate how agents' variable information on states' preferences affects their ability to effectively advance their design preferences. Our theory reveals how independent legal experts with delegated authority over design tasks influence institutional design processes and outcomes, which has practical and normative implications for the legalization of international politics.

Corresponding author:

Nicole De Silva, Department of Political Science, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, Quebec, H3G 1M8, Canada.

Email: nicole.desilva@concordia.ca

Keywords

Institutional design, international law, non-state actors, principal–agent theory, delegation, international criminal law

Introduction

The broad trend toward the legalization of international politics has prompted questions about factors influencing the creation of international legal institutions. These institutions range from non-binding “soft law” to binding “hard law” in the form of treaties that may have third-party monitoring and dispute settlement mechanisms (Abbott and Snidal, 2001; Koremenos, 2016). Existing scholarship considers the design of international legal agreements, or “international legal design” (Abbott and Snidal, 2013), a largely state-driven enterprise (Voeten, 2019). This state-centrism reflects international law’s dependence on state consent, where intergovernmental processes are typically the most visible aspects of the institutional design process. Recent scholarship recognizes the participation and influence of non-state actors in the design process. Intergovernmental organization (IGO) secretariats (Johnson, 2014; Johnson and Urpelainen, 2014; Johnstone, 2013; Pollack, 2003), nongovernmental organizations (NGOs) (Charnovitz, 2006; Glasius, 2002; Struett, 2008), epistemic communities (Rietig, 2014), and other non-state actors may set the agenda for states to develop an international institution and may advocate for particular design choices. We assert that states’ *delegation* of authority over design tasks to non-state actors with legal expertise is significantly different from such non-state actor *participation* in design (Boyle and Chinkin, 2007; Bradley and Kelley, 2008; Grant and Keohane, 2005; Liese et al., 2021). This article investigates how these non-state legal actors’ distinct preferences matter for institutional design processes and outcomes.

States’ use of non-state legal actors for drafting international agreements is a widespread but understudied phenomenon. Major IGOs, such as the United Nations (UN), Organization of American States, and African Union (AU), have institutionalized bodies of legal experts that are mandated to provide member states guidance on the development of international law and design of international institutions.¹ States may also seek guidance on a more ad hoc basis from, for example, special rapporteurs or expert working groups. These actors are “non-state,” as they do not represent a particular state’s interests. They are “legal,” as they have recognized expertise in international law. They may be corporate actors (e.g. a body of legal experts) or individual legal professionals (e.g. special rapporteurs). States assign legal actors design tasks ranging from researching issues surrounding the prospective institutional design to drafting a proposed legal agreement. This phenomenon deserves attention because the design preferences of these “hidden figures” may diverge from those of states and influence institutional design.

Existing interdisciplinary scholarship on international relations (IR) and international law (IL)—whether state-centric or inclusive of non-state actors—has not systematically investigated how non-state legal actors with delegated advisory roles influence international legal design. We fill this gap by theorizing the interactions between states and legal actors in the design process as a principal–agent (PA) relationship. We argue that legal dimensions of the PA relationship not only increase the likelihood of preference

divergence between states as the collective principal (Lyne et al., 2006: 44–45) and legal actors as their design agents, but also enable legal actors to opportunistically advance their own design preferences. When the agent perceives preference divergence with the collective principal, the agent can draw on its legal expertise to strategically select and justify design choices that *narrow* or *shift* the collective principal's design choices as close as possible to the agent's ideal design preferences. The more information on states' design preferences—as revealed within and beyond the PA relationship—the agent has, the more effectively it can exploit its legal expertise to propose design choices that maximize the agent's design preferences and the likelihood of the collective principal's acceptance.

We test our theoretical expectations through two case studies of states delegating the design of international criminal legal institutions to non-state legal actors: the UN General Assembly's (UNGA's) delegation to the International Law Commission (ILC), and the AU Assembly of Heads of State and Government and AU Commission's delegation to the Pan African Lawyers Union (PALU). Drawing on new archival and interview data, we show how the design agents' different levels of information regarding states' preferences (low in the ILC case and high in the PALU case) affected the agents' capacities to strategically advance their design preferences. Compared with the ILC, PALU could more effectively formulate design strategies that maximized its design preferences and states' acceptance of them.

Our analysis proceeds in three steps. First, we build on scholarship on delegation and PA relations to theorize how legal experts can influence institutional design. Second, we evaluate our theoretical conjectures about design agent strategies through our two case studies. Finally, in the conclusion we summarize the article's contributions to the literature on institutional design, and highlight its practical and normative implications for international cooperation and governance.

Designing international legal institutions: from state-centrism to the inclusion of non-state legal actors

Within IR, the rational–functional approach to explaining the design of international legal institutions theorizes institutional designs as equilibrium outcomes reflecting states' preferences at the time of contracting (Koremenos, 2005, 2016; Koremenos et al., 2001; Thompson, 2009). States retain decision-making authority and control over institutional design. Some scholarship challenges this state-centrism by theorizing, from rationalist and constructivist perspectives, how non-state actors' participation in intergovernmental processes (Bexell et al., 2010; Tallberg et al., 2013) can influence the design of international institutions (Johnson, 2014; Johnson and Urpelainen, 2014; Johnstone, 2013), including international legal institutions like the International Criminal Court (ICC) (Deitelhoff, 2009; Glasius, 2002; Struett, 2008). Scholars have demonstrated how transnational advocacy networks (Carpenter, 2007), epistemic communities (Haas, 1992; Rietig, 2014), and private sector actors (see Sell, 2003: 30 on institutional design generally; Durkee, 2016 on treaty design) influence states' cooperation in specific issue areas

by mobilizing around their preferred institutional or policy choice. Technical experts are recognized as influential in developing private sector regulatory standards (Büthe and Mattli, 2011).

However, the non-state actors discussed in the existing literature lack the authority to directly influence institutional design choices that affect public international law and institutions. Instead, they leverage their moral and expert authority to indirectly shape institutional design through agenda-setting and norm advocacy (Finnemore and Sikkink, 1998). The delegation of design tasks from states to non-state actors (e.g. researching and proposing design choices, drafting agreements) departs from other modes of non-state actor participation, because it involves a formal grant of authority over aspects of the design process. This creates distinct opportunities for non-state legal actors to influence the drafting of legal agreements, a process that requires further theorizing.

Our argument builds on existing scholarship on delegation (Bradley and Kelley, 2008; Koremenos, 2008; Pollack, 2003) and PA relations in IR (Hawkins and Jacoby, 2006; Johnson and Urpelainen, 2014). The PA approach focuses on how dynamic interactions between principals and agents influence outcomes, and accommodates a range of assumptions (e.g. from rationalist or constructivist perspectives) regarding the potentially distinct preferences principals and agents bring to the delegation relationship (Snidal and Tamm, 2018: 139). Considering IGO secretariats' frequent involvement in institutional design (Johnstone, 2013: 271; Tieku, 2021: 257), scholars have theorized how IGO bureaucrats' preferences can impact the design of international institutions, where they can insulate IGO progenies from state control (Johnson, 2014; Johnson and Urpelainen, 2014). There can be similar but distinctly "legalized" PA dynamics when states delegate authority to non-state legal actors to design international institutions. Scholars have engaged PA theory when explaining the behavior and independent influence of international judges (e.g. Elsig and Pollack, 2014), but not non-state legal actors in the context of institutional design. We further develop these lines of PA theorizing by accounting for the particularities of legal institutionalization and expertise, which contribute to legal actors having unique opportunities for influencing design outcomes according to their preferences.

Delegating institutional design to non-state legal actors

When pursuing a new international legal institution, a group of states (collective principal) may choose to delegate design-related tasks to a non-state actor with legal expertise. By centralizing design tasks in a single actor, states may aim to increase the efficiency of the design process and overcome collective action problems, such as those driven by preference heterogeneity (Green and Colgan, 2013: 491; Koremenos, 2008). States can seek expert information on the availability and implications of potential design choices to reduce uncertainty and facilitate cooperation, particularly for complex cooperation problems (Koremenos, 2008: 168; Pollack, 2003). States may also draw on a legal actor's expert and moral authority to legitimize the design process and outcome (Johnstone, 2013). Legal experts can, for example, clarify the normative consistency of prospective design choices within existing institutional and normative frameworks (Barnett and Finnemore, 2004: 24ff.; Copelovitch and Putnam, 2014; Haas, 1992). As independent

legal professionals rather than representatives of particular states, these legal actors may be perceived to more legitimately pursue the development of international law, rather than advance particular political interests. This can be advantageous for states seeking to increase the credibility of their international legal commitments (Pollack, 2003: 30).

Preference divergence in this PA relationship is likely. Legal actors' professional norms and reputational interests (Alter, 2008) within the "global community of law" (Helfer and Slaughter, 2005: 907, 953) mean that their values and interests differ from those of states. Central to legal actors' preferences is an aversion to violating principles of legality (Brunnée and Toope, 2010: 352–356). They also may be more concerned with some legal aspects of design (e.g. developing a more coherent system of rules) than states as political entities. Legal actors also will likely favor more ambitious legal commitments, such as new international legal rights, obligations, and accountability mechanisms. Unlike states, they do not face the costs (e.g. sovereignty costs) of such commitments but enjoy reputational benefits from leading the development of international law.

In conventional PA accounts, principal control mechanisms *ex ante* (e.g. selecting the agent, defining its mandate) and *ex post* (e.g. monitoring the agent, revoking its authority) serve to mitigate preference divergence by aligning the agent's behavior with principal preferences (Bradley and Kelley, 2008: 25ff.; Hawkins and Jacoby, 2006: 24). However, when states delegate design tasks to a legal actor, the PA relationship's legal aspects may limit the collective principal's ability to exercise control beyond standard PA assumptions (Abbott et al., 2020: 5–6; Pollack, 2003: 45). The relatively narrow pool of potential agents with relevant legal expertise, especially in highly specialized issue-areas, limits control through agent selection (Hawkins et al., 2006: 25). The collective principal may need to trade-off gaining expertise with accepting divergent or ambiguous agent preferences (Johnson and Urpelainen, 2014: 10). Generally, monitoring an agent with specialized legal expertise is more costly (Hawkins et al., 2006: 25). Individual state-principals' variable access to legal expertise can contribute to legal uncertainty, preference heterogeneity, and collective action problems within the collective principal (Liese et al., 2021: 370). These issues can undermine the collective principal's ability to select the agent and specify its mandate *ex ante*, and to effectively monitor and mobilize to control the agent *ex post*. When monitoring the agent, state-principals with lower legal expertise may defer to the agent's expertise (in specifying its mandate and carrying out its design tasks), rather than risk deferring (and giving negotiation advantages) to state-principals with higher levels of legal expertise within the collective principal.

These control problems related to legal expertise unintentionally facilitate agent autonomy, unlike PA relationships where principals accept the agent requires autonomy to perform its mandate (Abbott et al., 2020: 5–6; Alter, 2008). They create an opportunity structure for the agent to advance its distinct preferences for the development of international law when fulfilling its design mandate. Given the design agent's role is advisory, however, the collective principal retains ultimate *ex post* control to accept, modify, or reject the agent's design proposals.

Design agent strategies

We assume the agent will strategically propose institutional design choices that maximize both its design preferences and the likelihood the collective principal will accept

them. To do so, the agent needs to consider the level of preference divergence in the PA relationship, which can be conceptualized based on a spectrum of potential institutional design options. The agent will have its “ideal point” and acceptable range for design choices, based on its preferences. These are shaped by its professional norms (e.g. a commitment to legality), reputational interests, and ideas for how international law should develop in the cooperation area. It will estimate the collective principal’s acceptable range of design choices—the design options the agent expects the collective principal would accept—based on available information on state-principals’ preferences. This information needs to be timely, considering state preferences can change over the course of the design process based on new information on the cooperation problem, norm advocacy (Hawkins and Jacoby, 2006: 208; Finnemore and Sikkink, 1998; Keck and Sikkink, 1998), and changes in “the state of the world” (Koremenos et al., 2001: 778), among other factors. We argue that the more information on state-principals’ preferences the agent has, the more effectively it can exploit its legal expertise to select and justify design choices that maximize its design preferences and gain the collective principal’s acceptance.

Sources of information on states’ preferences. An agent’s level of information depends on whether and how state-principals reveal their design preferences within and beyond the PA relationship. Within the PA relationship, the collective principal’s soft controls, such as precisely defining the agent’s mandate *ex ante* and monitoring and feedback *ex post*, provide the agent with direct signals of the collective principal’s design preferences. Soft *ex post* controls allow the agent to expand and update its information on the collective principal’s preferences. The availability of this information depends on the principal’s ability and willingness to exercise soft control, which may be hindered by preference heterogeneity and collective action problems within the collective principal.

This provides a distinct perspective on the role of principal controls. Typical PA accounts assume that principal controls *constrain* agent behavior, aligning it with principal preferences (Bradley and Kelley, 2008: 21; Hawkins and Jacoby, 2006; Koremenos, 2008: 165). We highlight how soft controls can provide the agent with valuable information on principal preferences and *enable* the agent’s strategies for maximizing its own preferences. In addition, a lack of (soft) controls is generally assumed to increase agent discretion/autonomy (Hawkins et al., 2006: 11, 20–21; Johnson, 2014: 32–34), but we note that it also reduces the agent’s information on principal preferences and increases its uncertainty on how to exploit its autonomy without provoking harder controls that threaten the agent’s influence (e.g. reducing or revoking its mandate).

The agent can also infer the collective principal’s design preferences using information on individual state-principals’ design preferences. The agent estimates how these aggregate within the collective principal, considering collective decision-making procedures (e.g. consensus, majoritarian voting) and potential collective action problems (e.g. distributional or normative conflicts). PA frameworks in IR emphasize the influence of direct collective PA interactions, but agents can look beyond these to better estimate the collective principal’s preferences. They can unpack individual state-principals’ preferences based on bargaining dynamics and cooperative behavior within the collective principal. The availability of such information varies. Individual state-principals differ in

how much they signal their design preferences (e.g. “ideal point preferences”) to each other, and by extension to the design agent, in negotiations (Morrow and Cope, 2021). Design agents, like non-state actors more generally (Tallberg et al., 2013), will also have variable access to intergovernmental negotiations.

In addition, the agent can discern individual state-principals’ design preferences based on their cooperative behavior in the broader normative and institutional environment in the same issue area or related issue areas. This is particularly relevant in the international legal realm, considering legalization proceeds through iterative processes (Abbott and Snidal, 2013: 35). In densely institutionalized cooperation areas, the design agent can infer a state’s preferences for a prospective institution based on its contestation of existing norms (Stimmer and Wisken, 2019) and institutions (Morse and Keohane, 2014). The agent may also infer a state’s design preferences based on its legalized cooperation behavior, regardless of the issue area. For example, if a state consistently avoids delegating authority to international courts, the agent may assume the state prefers softer monitoring and enforcement mechanisms for the prospective institution. The relevance of this source of information will depend on the level of existing normative and institutional development, and state-principals’ engagement with it.

PA scholarship emphasizes how the principal’s level of information on the agent’s preferences and behavior affects the principal’s ability to control the agent, deterring opportunism and slack. We highlight how agents’ level of information on principals’ preferences affects agents’ ability to opportunistically maximize their own preferences. The more information the agent has on the collective principal’s preferences, the more accurately the agent can assess the level of preference divergence with the collective principal and the more effectively it can, drawing on its legal expertise, strategize how its design choices can overcome it.

Selecting and justifying design choices. The agent’s design strategies involve selecting and justifying design choices that maximize its design preferences and the likelihood of the collective principal’s acceptance. The agent’s level of information on the collective principal’s preferences affects its ability to strategically calibrate both its selection and justification of design choices. The more information the agent has, the more accurately it can estimate its preference divergence with the collective principal and the design options from which it can select. The agent may develop and select design options based on three distinct preference divergence scenarios (Figure 1): (1) the design agent’s ideal point overlaps with the collective principal’s acceptable range; (2) the agent’s acceptable range, but not its ideal point, overlaps with the collective principal’s acceptable range; and (3) there is no overlap at all. In addition, the more information the agent has on the collective principal’s preferences, the more effectively the agent can formulate justifications for potential design options. The agent can devise justifications that appeal to state-principals’ existing preferences or that change state-principals’ preferences to better align with the agent’s.

In short, the agent’s level of information guides its strategies for gaining the collective principal’s acceptance of design choices that are as close as possible to the agent’s ideal point and at least within its acceptable range. Thus, the agent’s design strategies may aim to *narrow* or *shift* the collective principal’s acceptable range of design choices (Figure 1).

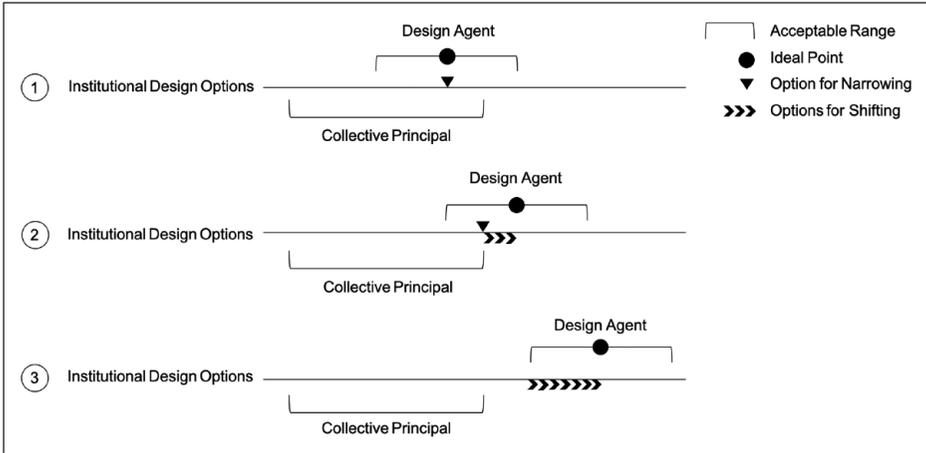


Figure 1. Three scenarios of preference divergence and potential agent design strategy.

A narrowing strategy is suitable when available information indicates preference divergence is lower and the collective principal's acceptable range overlaps with the agent's acceptable range (scenario 2) and, especially, its ideal point (scenario 1). The agent's design proposal narrows the collective principal's acceptable range to the design choice closest to its ideal point by taking design options that some states (or another design agent) might have preferred and otherwise pursued off the table. For example, when the collective principal's acceptable range of design choices includes any form of third-party monitoring or enforcement, and the agent prefers third-party enforcement, it will propose that design choice. Narrowing aids the collective principal in choosing among multiple design options (i.e. equilibrium selection, in rational choice terms) (Koremenos et al., 2001: 765, 795). This may be states' objective with delegation to the legal actor, but the agent's preferences matter because it narrows the design choices states consider as close as possible to the agent's ideal point. Since these design choices are within the collective principal's acceptable range the agent's justifications may appeal to state-principals' existing preferences and facilitate collective action for adoption.

The agent can use a shifting strategy to maximize its preferences when it estimates higher preference divergence with the collective principal, where its ideal point (scenario 2) or whole acceptable range (scenario 3) does not overlap with the collective principal's acceptable range. To maximize its preferences, the agent can aim to shift the collective principal's acceptable range as close as possible to the agent's ideal point. For instance, the agent may assess that states prefer a narrow set of trade-focused legal obligations for a prospective trade agreement, but the design agent justifies including human rights conditionalities in the agreement to maximize its preference for linking human rights and trade obligations. Justifications for shifting may appeal to state-principals' existing preferences or seek to change them to better align with the agent's. To appeal to states' existing preferences, for example, the agent can provide information emphasizing the design

choice's benefits and withhold information on its costs. The collective principal's limited "permeability" (Hawkins and Jacoby, 2006: 202) to information from third parties (e.g. other legal experts) and, as discussed earlier, deference to the agent's expertise could facilitate this agent strategy. To change states' preferences and acceptable range, the agent can engage in normative suasion for its preferred design choices (e.g. Deitelhoff, 2009; Struett, 2008).

There is the risk of the collective principal perceiving the agent's attempt to shift its acceptable range as agency slack (i.e. independent action by the agent that is undesired by the principal) (Hawkins et al., 2006: 7–8). In response, the collective principal may exercise *ex post* control. But shifting, as a riskier strategy, may be preferable over narrowing to maximize the agent's preferences (scenario 2), or may be necessary for the agent to overcome preference divergence (scenario 3). To effectively shift the collective principal's acceptable range, the agent uses available information to calibrate the design option closest to its ideal point that it can justify in a way the collective principal will likely accept.

Overall, the legal dimensions of the PA relationship create an opportunity structure for the agent to advance its distinct design preferences. The more information the agent has on states' preferences, the more effectively it can develop design strategies that maximize its preferences and the likelihood of state acceptance. Higher levels of information enable the agent to more precisely identify the collective principal's range of acceptable design choices, and clarify the agent's design options for narrowing or shifting the collective principal's acceptable range as close as possible to the agent's ideal point. In addition, higher levels of information increase the likelihood the agent can formulate justifications for design options that states will accept. Conversely, with lower levels of information, the agent is more prone to miscalculating its selection and/or justification of design choices, inhibiting it from maximizing its preferences and lowering the likelihood of states' acceptance.

Evidence from the design of international criminal legal institutions in the United Nations and African Union

We test our theoretical expectations regarding design agents' ability to influence institutional design through two case studies of delegated international legal design concerning international crime. In our first case, the UNGA mandated the ILC to design a code of international crimes. In the second case, the AU Assembly of Heads of State and Government tasked the AU Commission with hiring a consultant, PALU, to design a protocol establishing a criminal chamber for the African Court.

Our approach combines controlled comparison (cross-case analysis) of carefully matched cases with congruence testing (within-case analysis) to maximize causal inference (Beach and Pedersen, 2016: 269; George and Bennett, 2005: 181ff.). Providing a rare opportunity for controlled comparison, these cases are similar in many aspects of the PA relationship and delegated mandate, yet differ in the design agent's level of information on their collective principal's design preferences. In both cases, the collective principal is a formally organized intergovernmental body which delegates a design mandate

for researching and drafting a treaty concerning international crime. Both collective principals exercise minimal *ex ante* control by delegating imprecise design mandates and selecting agents based on a newly institutionalized relationship with the intergovernmental body. The design agents are corporate actors in which an individual leads the decision-making underpinning the legal actor's preferences and design proposals to the collective principal. These agents have similar design preferences—which diverge from their collective principals' preferences—for institutionalizing strong international criminal legal obligations and judicial enforcement. However, the two design agents' level of information regarding state-principals' preferences for the prospective treaty—based on revealed design preferences within and beyond the PA relationship—differ substantially. According to our theoretical expectations, the design agents' divergent levels of information on their collective principals' preferences (low for the ILC and high for PALU) would affect their capacity to develop design strategies that maximize their preferences, gain states' acceptance, and influence institutional design.

Our analysis of agents' design preferences, information on collective principals' preferences, and design strategies draws on extensive archival and interview data (see Supplemental Appendix for all referenced data sources, such as treaties, documents from the AU, ILC, and UNGA, and interviews with PALU). The historical ILC case study relies on archival data—considering the ILC meticulously documents its work in its yearbook (e.g. ILC, 1949), including its internal debates on how to carry out its activities—and secondary literature on the drafting and negotiation processes. AU treaties do not have equivalent *travaux préparatoires*. We therefore draw data from interviews with PALU to complement the available documentation on AU decisions and discussions during the design process.

Case 1: the UNGA and the ILC

Delegation. After World War II, UN member states agreed to institutionalize stronger legal accountability for atrocity crimes, including aggression, genocide, war crimes, and crimes against humanity. However, they lacked well-defined preferences on how to develop an international criminal legal regime. The United States primarily wanted an official endorsement of the legal principles of the Nuremberg Charter (Schabas, 2011: 772–773). Other states aimed to create legal instruments to prosecute genocide and other political violence during war and peacetime (UNGA, 1947a A_PV-123: 1295–1290). Existing UN bodies (i.e. the Secretariat and the Economic and Social Council) felt ill-prepared to tackle these problems and preferred that states consolidate responsibility for addressing atrocity crimes in a specialized agency (Schabas, 2007: 38–40). In 1947, the UNGA created the ILC as a subsidiary body with expert authority (UNGA, 1947b Res. 174(II)) to fulfill the UN Charter mandate of progressively developing and codifying international law (Charter of the United Nations, Article 13(1)(a)). As its first task, the UNGA authorized the ILC to (1) formulate the principles of international law recognized in the Nuremberg Charter and judgments of the Nuremberg trials, and (2) based on these, prepare a draft code of offenses against the peace and security of mankind (UNGA, 1947c Res. 177(II)).²

The ILC operates independently but is subject to UNGA control. It is comprised of 34 legal experts elected by UN member states for their “recognized competence in international law” (ILC Statute, Article 2(1); Helfer and Meyer, 2016: 307). *Ex ante*, the UNGA influences the scope of the ILC’s work via resolutions that task the ILC with addressing specific areas of international law for progressive development. However, the ILC also asserts its authority to codify international law without explicit UNGA approval (ILC, 1949 [1956]: 279). *Ex post*, the UNGA monitors the ILC, requiring annual progress reports on its activities. The UNGA decides whether the ILC’s work is adopted as a report or draft article (soft law) or proceeds to treaty negotiations (hard law) (ILC Statute, Article 23). Ultimately, if dissatisfied with the ILC’s performance, the UNGA can revoke or suspend an ILC mandate.

Design agent strategies—initial mandate (1949–1954). The ILC began work on a draft code of crimes in 1949, with limited information regarding the UNGA’s preferences for institutionalization (ILC, 1950b: 8ff.). The UNGA’s ambiguous instructions exercised minimal *ex ante* control and left open whether the ILC should merely reproduce the Nuremberg Charter or develop international law beyond this precedent. The ILC repeatedly requested but failed to receive further information on its design mandate,³ specifically whether it should primarily codify existing law or progressively develop new legal rules (ILC, 1949 [1956]: 9, 11–23, 279). Only four states answered the questionnaire the ILC provided (ILC, 1950a: 249–253). The ILC could gain little additional information on state-principals’ preferences from their behavior within this new cooperation area. States had cooperated in the Nuremberg and Tokyo Tribunals to prosecute individuals from Germany and allied states for war crimes, crimes against peace, and crimes against humanity. However, debates within the UN over the Genocide Convention highlighted states’ preference heterogeneity on how to develop international criminal law more broadly (Schabas, 2007: 38–40; UNGA, 1947a A_PV-123: 1295–1290), undermining the ILC’s ability to determine how state preferences would aggregate in the collective principal.

With considerable uncertainty about states’ acceptable design choices, the ILC strategized how to advance its two overarching design preferences. First, the ILC preferred a comprehensive criminal code of fundamental violations of international law that went beyond a mere codification of the Nuremberg Charter, thus progressively developing international law (ILC, 1951: 58). Substantively, the ILC sought to strengthen legal obligations regarding atrocity crimes by expanding the code to cover atrocities committed within states against civilians. It also sought to insulate the draft code from criticism that the Nuremberg Charter did not reflect existing international law. Second, the ILC had a strong preference for including an integrated court as an enforcement mechanism. Although the UNGA mandate did not reference a potential enforcement mechanism, the ILC considered a code without enforcement ineffective (ILC, 1951: 134).

Concerning the comprehensive code, the ILC pursued two main design strategies. First, it estimated that its preference for organizing the code around core crimes based on the UN and Nuremberg Charters—referenced in the design mandate—overlapped with the UNGA’s acceptable range (Figure 1, scenario 1). The ILC thus adopted a narrowing strategy, focusing the draft code on the most fundamental violations of international law

with a “political element,” that is, breaches of international law sanctioned or carried out by political actors (ILC, 1951: 59). It based its definition of offenses against peace and security of mankind on the broad prohibition of the use of force in the UN Charter, as well as the prohibition of war crimes, crimes against humanity, and genocide in the Nuremberg Charter. The ILC considered these unique offenses as they challenged international order and not just the rights of individual states. The ILC’s strategy excluded potential design options based on previous institutionalization attempts from the inter-war period that covered a much broader set of offenses within the category of international crimes (war crimes, aggression but also piracy, terrorism, drug, and human trafficking, etc.) (Holthoefer, 2016; Lewis, 2014).

The ILC perceived partial overlap (Figure 1, scenario 2) concerning its second main design preference for the code to increase the protection of the civilian population from violence beyond the Nuremberg Charter and the UN Charter. The negotiations over the Genocide Convention indicated that states accepted the need for such protection but not to the extent the ILC would ideally choose. The ILC estimated it could shift the UNGA’s acceptable range to pursue its ideal design choice of extending the code to “inhumane acts” against civilian populations independent of inter-state conflict beyond the Nuremberg precedent that limited accountability to wartime atrocities (ILC, 1954: 150ff.). Its justification appealed to states’ existing preferences, emphasizing that states had already consented to legal protection for civilian populations through the Geneva Conventions (1949) and Genocide Convention (1948) (ILC, 1950b: 59ff.). The draft code would thus broaden protection by delinking “inhumane acts” from inter-state war and genocide. This was a significant normative change that rested on a broad interpretation of “offenses against peace and security,” as well as a restrictive view of the sovereign power of states in the domestic context.

The ILC assessed that its second overarching design preference for a strong judicial enforcement mechanism for the code was beyond the UNGA’s acceptable range of design choices (Figure 1, scenario 3). The ILC preferred an independent court but also considered the possible alternative of enforcement through domestic courts. The UNGA’s design mandate, however, did not reference enforcement. Moreover, the ILC’s internal debates show that it understood, through observing negotiations within the UNGA, that states were concerned over sovereignty costs and preferred using the draft code to signal their commitment to the rule of law in the international system rather than as a jurisdictional foundation for third-party (e.g. court) enforcement. To maximize its design preference, the ILC initially attempted to shift the UNGA’s acceptable range of design choices to accommodate a judicial enforcement mechanism for the code. Its justification tried to connect its design mandate for the code to its previous, separate UNGA mandate from 1948 to investigate the “desirability and possibility” (ILC, 1950a: 379–380; 1950b: 18; 1951: 134) of an international criminal tribunal envisioned in Article VI of the Genocide Convention (UNGA, 1948 Res. 260(III) B). But when the UNGA was unresponsive to this proposal indicating a low likelihood of acceptance, the ILC dropped it (ILC, 1954: 150).

Except for the enforcement mechanism, the ILC’s (1954) draft code realized its core preferences for a comprehensive code including crimes against peace, aggressive war, crimes against humanity (“inhumane acts”), and genocide, as well as foreign

intervention. The UNGA, however, tabled the proposal (UNGA, 1954 Res. 897 (IX)) based on considerations beyond the scope of the ILC's design strategies. Nominally states sought time to settle their internal disagreement over defining the crime of aggression, a core offense in the code, but as political tensions increased with the intensifying Cold War, states' preferences shifted toward strengthening sovereignty at the expense of legal accountability.

Design agent strategies—renewed mandate (1981–1996). Due to Cold War exigencies, the UNGA only re-authorized the ILC's work on the draft code in 1981. The ILC's updated mandate contained some additional information on states' preferences: the code should implement the fundamental principles of the UN Charter to strengthen "peace and security," considering relevant new legal developments (e.g. the definition of aggression the UNGA adopted in 1974 (Res. /24/3314)). The renewed mandate still failed to reference an enforcement mechanism (UNGA, 1981 Res. /36/106). The nearly 30 years of additional legal developments since the ILC's first phase of work suggested that the collective principal's preferences had not significantly changed. Notably, states (and legal experts) lacked agreement on what constituted "international offenses" beyond the established categories of aggression, war crimes, crimes against humanity, and genocide (Mégret, 2019: 77-81). Overall, the ILC still had very limited information on the UNGA's design preferences.

With a new Special Rapporteur Doudou Thiam leading this second phase of the ILC's work on the draft code, the ILC developed new design preferences for code and court that apparently did not overlap with states' preferences (Figure 1, scenario 3). Therefore, the ILC needed to pursue shifting strategies to advance its design goals. However, the lack of timely and concrete information on states' preferences undermined the ILC's ability to calibrate how to select and justify design choices for effective shifting strategies.

First, moving beyond its original approach in the 1954 draft code, the ILC sought a comprehensive code of crimes that further expanded what constituted "offenses against peace and security." The 1954 draft code's narrower set of offenses appeared to fall within the UNGA's acceptable range, based on the UNGA's original instructions and lack of active resistance to this part of the ILC's 1954 proposal, but design choices for fulfilling the ILC's updated preference seemingly did not. The ILC's shifting strategy pursued its ideal design choices for the code to cover "the most serious international offenses" recognized under international law, including colonialism, apartheid, economic aggression, use of certain weapons, and serious environmental damages (which subsequently would be labeled "treaty crimes") beyond the core crimes (aggression, genocide, crimes against humanity, war crimes). It justified this expanded scope based on legal instruments completed since the 1950s, arguing that these offenses reflected existing legal norms and were consistent with state interests (ILC, 1984a: 100; 1984b: 11). The ILC assumed this expansion could align with the interests of developing countries recovering from colonialism and economic exploitation, and that their support would enable the ILC to shift the acceptable range of the collective principal overall (UN, 1996: 10–12).

Second, the ILC continued to prefer to integrate a strong judicial enforcement mechanism into the code (ILC, 1983a: 149; 1983b: 13ff., 16). The UNGA's renewed mandate exclusively focused on the code, implying a court was likely beyond its acceptable range of design choices. To calibrate its shifting strategy, the ILC attempted to elicit UNGA feedback (soft control) to gain information on state-principals' preferences. It asked the UNGA in 1983 and 1986 for guidance on the code's enforcement mechanisms and whether its design mandate included a court. States' unresponsiveness merely indicated a court was likely outside their acceptable range of design choices, rather than clarifying how the ILC (1986: 54) should tailor its court proposal. The ILC's (1987: 4) design preferences required a court for strong enforcement of the code, so it continued its shifting strategy while risking that it lacked sufficient information on state-principals' preferences to effectively advance its ideal design choice of an integrated court.

In the 1980s, states' preferences shifted toward a court because of the changing cooperation problem, rather than the ILC's design strategy. The UNGA sought to address transnational crimes of drug trafficking and terrorism and, on Trinidad and Tobago's initiative, formally asked the ILC to "address the question of establishing an international criminal court" (UNGA, 1989 Res. 44/39) with jurisdiction over such crimes. Despite this shift, the UNGA's acceptable range still did not overlap with the ILC's acceptable design choices for an expanded set of crimes and a court as enforcement mechanism (Figure 1, scenario 3). The ILC built the draft code around a definition of international crimes as offenses to the peace and security of mankind, but transnational crimes like trafficking and terrorism did not easily fit into these established categories (ILC, 1990: 17–18; UNGA, 1981 Res. /36/106). States seemed to be interested in extending their ability to punish criminals beyond state borders but provided little additional information on enforcement mechanisms for international crimes more generally through soft controls (state feedback) or otherwise. Uncertainty regarding states' preferences hampered the ILC's efforts to target design choices and justifications. It also undermined the ILC's capacity to ally with states that could support a unified code and court, and help shift the UNGA's acceptable range (UN, 1996: 10–12).

In the 1990s, significant changes to the cooperation problem further shifted the UNGA's acceptable range, so it overlapped with the ILC's regarding an independent criminal court but not in terms of the ILC's ideal of integrating it into the draft code (Figure 1, scenario 2). After the creation of ad hoc international criminal tribunals for Yugoslavia and Rwanda, the UNGA requested the ILC prioritize its work on a permanent court for prosecuting international crimes. This resulted in more UNGA monitoring of and feedback to the ILC, which increased the ILC's information on changing collective principal preferences regarding court and code and its ability to adjust its design strategies.

Considering these changes and increased information on states' preferences, the ILC internally debated how to maximize its core preferences for a comprehensive code and court for enforcement (ILC 1992a: 11; 1994b: 12ff.; 1995: 32–50). It concluded that states were unlikely to accept its justifications for a court with broad compulsory jurisdiction over the crimes outlined in the draft code (Crawford, 1995: 410). Abandoning its ideal design option of an integrated code and court (ILC, 1992b: 58, fn. 55), the ILC strategically pivoted to a narrowing strategy that separated the draft code from the statute

for a criminal court, would be justifiable based on states' existing design preferences, and would more likely gain states' acceptance. In an environment of rapidly changing information on states' preferences, separating the code and court proposals also prevented states' rejection of one proposal from automatically threatening the other (ILC, 1994a: 76; 1996: 17).

The ILC's design proposal for an ICC, presented to the UNGA in 1994, reflected shifting strategies that the ILC adjusted based on available information on UNGA preferences. It demonstrated the ILC's design preferences for a comprehensive code by including both crimes under general international law (genocide convention, aggression, war crimes, crimes against humanity) and a provision for extending jurisdiction to so-called treaty crimes (apartheid, terrorism, drug trafficking, torture). The ILC emphasized that the statute was procedural (i.e. focused on the court's institutional features) and did not attempt to define the specific crimes, referring instead to its work on the draft code. A court with compulsory jurisdiction over all crimes would provide the strongest enforcement. Instead of pursuing this ideal, the ILC selected a design option within the overlap of its and the UNGA's acceptable ranges. The court would have compulsory jurisdiction only in the case of genocide and require state consent for jurisdiction over other crimes (ILC, 1994a: 27ff.). The ILC justified compulsory jurisdiction for genocide given the severe nature of this crime and its consistency with existing legal developments, such as the proposed enforcement mechanism in the genocide convention (ILC, 1994a: 37, 38).

The ILC had relatively less information on states' preferences to guide its shifting strategies for its final proposal for the Draft Code of Crimes, submitted to the UNGA in 1996. The ILC reorganized the draft code, which was previously broader, around a list of "core crimes" (aggression, genocide, crimes against humanity, and war crimes), as states had indicated their preference for a limited list in their discussions and feedback to the ILC (ILC, 1996: 16). However, the ILC used the category "crimes against humanity" to shift toward the ILC's ideal preference for a comprehensive code by including a broad set of offenses, including racial/ethnic discrimination, torture, mass deportation, etc. (ILC, 1995: 36; 1996: 16, 47). The ILC justified its expansive definition of crimes against humanity by drawing on the normative authority of the Nuremberg Charter and Judgment (referenced in its mandate), as well as the tribunals for Yugoslavia and Rwanda (implicitly related to its mandate). It also included a flexibility provision allowing for future expansion of the international criminal code to other issue areas (ILC, 1996: 16).

Design outcome. The UNGA did not follow the ILC's recommendation to convene a treaty-making conference to adopt the draft statute for a criminal court but appointed an ad hoc committee to review the ILC draft statute. After several revisions of the draft, the UNGA exercised *ex post* control and established a new design agent, the Preparatory Committee, tasked with consolidating a treaty text that would eventually lead to the creation of the ICC. States used the ILC's court proposal as a basis for continuing design negotiations, but they dismissed the ILC's proposal for a separate code. Receiving the final version of the draft code, the UNGA thanked the ILC for its work and invited states to provide comments to the ILC. Neither the UNGA nor any states took further action on the proposal (UNGA, 1996 Res. 51/160).

Case 2: the AU and the PALU

Delegation. AU states' interest in a new AU institution to prosecute international crimes emerged from many African states' dissatisfaction with existing norms and institutions within the highly developed international criminal legal regime. This interest initially arose in the context of African leaders' objections to non-African (frequently European) courts' application of the principle of universal jurisdiction to prosecute Africans, particularly African leaders (Deya, 2014; Murungu, 2011). According to this legal principle, international crimes are so severe that they affect the international community as a whole and therefore can be prosecuted in any state. Building on a 2008 AU Commission (AUC) report criticizing the application and implications of universal jurisdiction (AUA, 2008: para. 5 (iii)), the AU Assembly of Heads of State and Government (the AU's supreme intergovernmental decision-making body) condemned "the political nature and the abuse of the principle of universal jurisdiction by judges from non-African States against African leaders," and argued it was "a clear violation of [African states'] sovereignty and territorial integrity" (AUA, 2008: para. 5 (ii)). It ordered the AUC, in consultation with the African Court on Human and Peoples' Rights (the AU's judicial organ), to investigate the potential for empowering the Court to try international crimes and to report back to the Assembly by 2010 (AUA, 2009a: para. 9).

Fitting its common practice, the AUC hired an external consultant for this mandate. First drafts of most AU legal agreements are developed by the AUC Office of the Legal Counsel (OLC), as the AU's in-house legal expertise, or by AUC-hired consultants (Tieku, 2021: 260). Here, the OLC's significant workload and lack of expertise in international criminal law drove the decision to outsource (Deya, 2019). Although it often does open calls for consultants, the AUC directly appointed PALU for this mandate (Deya, 2019). Since 2006, PALU and the AU had a Memorandum of Understanding (MOU) to cooperate in areas such as promoting regional and international jurisdictions and strengthening African unity under the rule of law (AU, 2006: Article I (3–4)). The AU committed to make grants to PALU and use PALU experts for specific missions (AU, 2006: Article III (2–3)). PALU's President Akere Tabeng Muna, who was also President of the AU's Economic, Social and Cultural Council, had been lobbying the OLC to give PALU funded consultancy work, and the AUC finally gave PALU this mandate (Deya, 2019).

In February 2010, the OLC, on behalf of AU states (the collective principal), provided PALU (the agent) a letter of instructions to prepare a study and draft protocol for creating a criminal chamber at the African Court (Deya, 2019). PALU was to propose measures to strengthen existing African institutions for democracy, human rights, and addressing international crimes. This included drafting a protocol to extend the jurisdiction of the African Court of Justice and Human Rights to cover international crimes and establish, if appropriate, an appellate jurisdiction (AU, 2010: 1). PALU would report directly to the AU Assembly and its advisory intergovernmental bodies, including the Executive Council and the Permanent Representatives Committee (PRC).

Design agent strategies (2009–2010). Led by PALU's Chief Executive Officer Donald Deya (Murungu, 2011: 1068),⁴ PALU aimed to use its design mandate to strengthen African regional governance, continue Africans' global leadership and innovation in the

development of international law, and thus contribute to “Africa rising” (Deya, 2012b: 26; 2014). Advancing human rights, good governance, and the rule of law in Africa was central to PALU’s organizational mission and to Deya’s (2014) personal convictions and experiences as a legal professional. To strengthen international criminal justice in Africa, PALU preferred for the prospective institution to broaden international criminal legal obligations and strengthen their enforcement beyond the status quo (Deya, 2014).

PALU assessed preference divergence considering multiple sources of information on state-principals’ individual and collective preferences for the prospective “African criminal court.” The AU’s imprecise design mandate could be compatible with PALU’s design preferences, but African states’ behavior within the well-developed international criminal legal regime provided further information. By the time PALU was executing its design mandate, African states were opposing both the application of universal jurisdiction and the ICC, based on the perception that both were targeting Africans, particularly African leaders. African states had initially strongly supported the ICC, as its largest regional bloc of states parties and the first states to refer situations to the Court (Deya, 2014). However, African states were vehemently opposed to the ICC Prosecutor applying for (in July 2008) and ICC judges granting (in March 2009) an arrest warrant for Sudanese President Omar al-Bashir (ICC, 2009). The AU Assembly decided in July 2009 that AU states “shall not cooperate” with the arrest warrant and African states parties to the Rome Statute should prepare guidelines regulating the ICC Prosecutor’s discretion to initiate cases (AUA, 2009b: paras. 10–11). PALU observed that AU states’ opposition to universal jurisdiction and the “bitter divide between Africa and the ICC” (Deya, 2012a) were “running in parallel” and driving this “moment of political will” for creating an African criminal court (Deya, 2019). The AU Assembly’s consensus-based decision-making, where the AU’s “Big Five” states had disproportionate influence, guided PALU’s estimation of the collective principal’s design preferences (Deya, 2019).

PALU assessed that its design preferences overlapped with AU states’ preferences in terms of using the Rome Statute as a foundation for the design (Figure 1, scenario 1). AU states’ contestation apparently focused not on the Rome Statute’s general design—especially considering most AU states’ membership in and strong initial support for the ICC—but on the ICC’s discretion when implementing it (Deya, 2014). PALU assumed that most Rome Statute provisions, reflecting its ideal design choices, would be acceptable to AU states. As a narrowing strategy covering many design choices, PALU therefore used the Rome Statute as its design template.

PALU, however, perceived preference divergence with AU states concerning its goal of dramatically strengthening international criminal legal obligations. AU states’ behavior within the existing international criminal legal regime indicated many states were aiming to weaken their obligations via the new institution (Deya, 2014). PALU’s ideal design choices for expanding international criminal legal obligations apparently were outside AU states’ acceptable range (Figure 1, scenario 3), but PALU assessed that a shifting strategy could overcome this. PALU chose to criminalize 10 additional acts beyond the Rome Statute’s 4 core crimes of genocide, war crimes, crimes against humanity, and aggression. The new crimes included unconstitutional change of government, piracy, terrorism, mercenarism, corruption, money laundering, trafficking in persons, trafficking in drugs, trafficking in hazardous wastes, and illicit exploitation of natural

resources. Existing AU treaties regarding human rights, good governance, and the rule of law already prohibited these acts (Deya, 2014), but PALU was converting states' obligations to address these issues into international crimes subject to individual criminal accountability.⁵ Paralleling the ILC's justifications for obligations in the 1980s, PALU justified these additional crimes by emphasizing their consistency with existing AU law and downplaying their departure from the status quo (Deya, 2014). To maximize the likelihood of states' acceptance, PALU also appealed to the AU norm of finding "African solutions to African problems": in Africa, these additional crimes could underpin and escalate to the ICC's four core crimes (Deya, 2014).

PALU also perceived that its preferences for strengthening enforcement significantly diverged from states' preferences (Figure 1, scenario 3), thus necessitating shifting strategies for two design choices. First, it sought to expand the scope of international criminal legal enforcement from individuals to include corporations (Malabo Protocol, Article 46C), considering corporations often held responsibility for serious human rights abuses and even fueling conflicts in Africa (Deya, 2014). PALU recognized this was beyond the design options states would have considered and, therefore, engaged in normative suasion by again arguing its design choice adhered to the AU norm of "African solutions to African problems" (Deya, 2014).

Second, PALU carefully formulated a shifting strategy for advancing its preference for the African criminal court to work collaboratively with other courts to enforce international criminal law (Deya, 2014). While not explicit in the design mandate, many AU states evidently sought to create an African criminal court to block international criminal prosecutions of Africans, particularly African leaders, by non-African actors (Deya, 2014). If Africans held their own international criminal trials, external actors would not be able to intervene, either under universal jurisdiction or the ICC. The ICC's complementarity clause (Rome Statute, Article 17) renders cases inadmissible if under investigation or prosecution by a state that has jurisdiction, so the African criminal court presumably could help block ICC interventions in Africa. This could serve the "self-interests of African political elites," such as leaders seeking to evade criminal accountability, rather than the interests of justice for Africans (Deya, 2014).

Based on its information on states' preferences, PALU strategically did not explicitly reference complementarity with the ICC and instead included an ambiguous clause that "The Court shall be entitled to seek the cooperation or assistance of regional or international courts, non-States Parties or co-operating partners of the African Union and may conclude Agreements for that purpose" (Malabo Protocol, Article 46L). This design choice aimed to insulate the question of complementarity with the ICC from political negotiations during the design process; instead, after the court's creation, legal actors (i.e. African Court and ICC judges) could decide the relationship between the two courts (e.g. through an MOU) (Deya, 2014). This increased the likelihood of realizing PALU's preference for stronger enforcement through national, regional, and global institutions "work[ing] harmoniously together to help end impunity for international crimes in Africa" (Deya, 2012b: 26). Drawing on its legal expertise and information on states' preferences, PALU's justification for this clause and its ambiguity stressed the need for complementarity with a variety of regional courts in Africa and avoided referencing the

ICC (Deya, 2014). PALU thus strategically withheld information on the implications of this design choice and relied on African states' deference to its legal expertise.

Design agent strategies (2010–2014). After receiving PALU's study and design proposal, AU states requested further expert information on the design's implications. This indicated AU states' limited legal expertise and capacity to exercise *ex post* control through monitoring and feedback. Reluctant to defer to PALU's legal expertise, they increased the "permeability" (Hawkins and Jacoby, 2006) of the PA relationship. The AUC organized two workshops in 2010 with PALU and other external legal experts to validate PALU's design proposal (Abass, 2017: 12; Murungu, 2011: 1068). These broader expert consultations, however, did not yield any particular challenges to PALU's selection and justification of its design choices. In 2012, the AU Assembly tabled the PALU draft for adoption, but rather than adopt it as expected (Abass, 2017: 12), it requested further information on the implications of PALU's design proposal. It ordered the AUC, in collaboration with AU legal experts from the African Court and AU Commission on International Law (AUCIL), to clarify the "financial and structural implications" of the expanded jurisdiction and the "definition of the crime of unconstitutional change of government" (AUA, 2012: paras. 2–3). In December 2012, another validation workshop—convened by the AUC with representatives from the AUC, African Court, AUCIL, and PALU—again yielded no notable challenges to the PALU design and concluded that "the only additional expenses envisaged will be in the expanded structure and operation of the [African Court]" (AU, 2012: 5). Overall, the collective principal only had the capacity to flag areas of uncertainty, where it sought increased expert information, but it could not effectively exercise soft control (e.g. provide feedback).

This increased PALU's autonomy in advancing its design choices, but also necessitated PALU looking beyond PA relations to discern AU states' preferences and likelihood of accepting PALU's design choices. Successive AU Assembly decisions opposing the ICC and its increased charges against Africans and African leaders over time demonstrated the collective principal's preference for expediently creating the African criminal court. AU states' strong preference for quickly adopting a design seemed to expand the collective principal's acceptable range of design choices. In addition, based on its access to discussions in the AU Assembly and the validation workshops, PALU discerned preference heterogeneity within the collective principal and how its shifting strategies were gaining some state-principals' acceptance (Deya, 2019). Based on these developments, PALU continued to advocate for its proposal that closely reflected its ideal design choices and expected it would likely gain the collective principal's acceptance.

Design outcome. In 2014, the AU's newly-created Specialized Technical Committee on Justice and Legal Affairs—comprising states' legal representatives and led by the AU's Legal Counsel—conducted a final review of the proposed protocol for regional criminal jurisdiction. Reflecting the collective principal's overriding preference of sovereign immunity from prosecution, the committee introduced a single change to the PALU design: a subclause (Article 46A bis) providing immunity from prosecution for senior government officials while in office (AUA, 2014). Except for this specific, state-introduced enforcement gap, PALU effectively gained states' acceptance of its preferred

design choices supporting stronger international criminal legal obligations and enforcement.

Discussion

Our cases supported our core hypothesis that the more information a legal actor has on states' preferences, the more effectively it can select and justify design choices that maximize its own preferences and the likelihood of states' acceptance. Both the ILC and PALU pursued similar design preferences for strong international criminal legal obligations and judicial enforcement throughout their design mandates. However, these agents had significantly different levels of information (ILC low and PALU high), which affected their capacities to advance their design preferences.

When commencing their mandates, the two agents had limited information on their collective principals' design preferences based on *ex ante* controls within the PA relationship (i.e. vague design mandates). But the availability of information from other sources, namely state-principals' behavior within the collective principal and the broader normative and institutional environment, varied considerably. Overall, UNGA states did not extensively discuss or reveal their preferences regarding institutionalization to the ILC. There was little opportunity to observe state behavior vis-à-vis existing norms and institutions because this was a new cooperation area. AU states, conversely, were revealing their design preferences much more, both at the individual and collective (AU) levels, based on their opposition to existing institutionalization.

This disparity in the agents' information on their collective principal's preferences persisted as PA interactions progressed. Demonstrating agents' need for timely information, both agents had protracted design processes where the cooperation problem and states' preferences changed over time, independent of PA interactions. This was extreme in the ILC case, where the UNGA halted, renewed, and emphasized aspects of the ILC's mandate as the cooperation problem changed. The ILC attempted to elicit information beyond its mandate but generally had to infer state-principals' preferences based on their unresponsiveness, rather than active signals. PALU, conversely, had abundant information on state-principals' individual and collective preferences, based on PALU's direct access to negotiations and states' openly escalating opposition to the institutional status quo.

Supporting our theoretical expectations, this variation in information on the collective principal's preferences contributed to different agent capacities to develop design strategies that would maximize agents' design preferences and gain states' acceptance. The agents pursued narrowing and shifting strategies based on their estimation of their collective principals' acceptable ranges of design choices, relative to their own, but the availability of information affected the agents' ability to assess these acceptable ranges. The ILC was uncertain regarding the UNGA's acceptable range, making it difficult to calibrate its selection and justification of design choices. The ILC, with limited information from other sources, heavily relied on interpreting its design mandate to estimate the UNGA's acceptable range and formulate its proposals. Narrowing design choices (i.e. core crimes) based on its delegated mandate was relatively straightforward. However, the ILC's limited information undermined its capacity to determine which design options

and justifications could shift the UNGA's acceptable range toward the ILC's preference of integrating the code and court.

PALU, conversely, could more precisely target its design strategies to states' preferences and have greater certainty that states would accept its proposed design choices. For example, AU states' behavior regarding existing institutionalization (especially the Rome Statute) clarified where PALU's acceptable range overlapped with its collective principal's and facilitated narrowing many design choices. PALU's understanding of state-principals' preferences heightened its ability to formulate justifications (e.g. by withholding information on legal implications) that states would likely accept, which also facilitated shifting design choices toward its ideal point. Similar ILC and PALU justifications (framing expanded international criminal legal obligations as consistent with existing legal obligations) were more effectively targeted to states' preferences in the PALU case.

As expected, the collective principal's efforts to control their agent had paradoxical impacts, either constraining or enabling—based on increased information on the collective principal's preferences—the agent's design strategies. This played out differently in the two cases. For example, increased information from the UNGA's soft controls toward the end of its design process contributed to the ILC abandoning its ideal choice of an integrated code and court (shifting), instead pursuing an acceptable choice of separating the two (narrowing) to increase the likelihood of states' acceptance. However, this increased information and change in strategy also enabled the ILC's shifting vis-à-vis other design choices (e.g. a comprehensive code of crime). PALU's collective principal generally lacked the collective capacity (e.g. legal expertise) to exercise soft control, and the process of attempting control actually provided PALU information that increased its certainty about maintaining its shifting strategy.

In both cases, variable information affected the agents' ability to select and justify design choices to effectively gain states' acceptance (i.e. avoid hard *ex post* control). Except for states exercising hard *ex post* control by introducing a targeted exception to the strong enforcement PALU advanced in its design proposal, states adopted PALU's design choices as their institutional design. In contrast, states did not adopt the ILC's proposals directly (i.e. exercised hard *ex post* control). However, given its expert authority, the ILC's work indirectly influenced future design efforts by other legal actors that ultimately gained states' acceptance after the UNGA–ILC PA relationship concluded. This demonstrates how an agent's design proposals (e.g. norm entrepreneurship) may lack influence within the delegation relationship but impact institutional design processes beyond it.

This controlled comparison held constant several aspects of the PA relationship to demonstrate the influence of the agent's variable information on states' design preferences. Further analysis could explore the potential influence of other variables (e.g. scope conditions, background factors) influencing design agents' strategies and their ability to shape the design outcome. For example, there may be variation in the agent's ability to discern how preferences aggregate in formally versus informally organized collective principals. Additionally, we focused on the creation of hard law (treaty-making). Our PA dynamics should hold for developing soft law, but further analysis could explore scope conditions regarding the procedural differences surrounding creating hard versus soft law. We focused on legal experts as design agents, but states equally may

delegate design tasks to actors with other forms of expertise (e.g. scientific) when designing institutions. Potential variation in strategies by agents with different forms of expertise, therefore, is another area for further research.

Conclusion

We have elucidated the phenomenon of states formally delegating institutional design tasks to non-state legal actors. Compared with the existing scholarship's emphasis on the *participation* of networks or communities of non-state actors, *delegation* to an individual non-state legal actor constitutes a distinct opportunity structure to influence the design of international institutions. Our PA theory accounts for how legal actors' preferences, which can significantly diverge from states' preferences, influence institutional design. This increases our understanding of delegation and expert agent autonomy in IR. It also highlights an important opportunity structure for legal norm entrepreneurship. Agents' ability to exploit and increase their discretion/autonomy when carrying out their mandates depends on having accurate and timely information on principal preferences. Agents can proactively draw on different sources of information within and beyond the PA relationship to strategically maximize their preferences.

The ability of these "hidden figures" to influence design according to their distinct preferences has important practical and normative implications for whether and how international cooperation for institutionalization proceeds. From a rationalist perspective, non-state legal actors may facilitate cooperation by anticipating and addressing potential costs, distributional conflicts, and other cooperation barriers states do not foresee. However, there is a risk of states abandoning the institutional design process or not using the created institution if the institutional design reflects the legal actor's preferences more than states' preferences. For example, AU states collectively adopted PALU's treaty design, but it has not entered into force due to insufficient state ratifications (AU, 2019). Agent strategies, like withholding information on the implications of design choices, may successfully influence the institutional design, but then provoke contestation and noncompliance once the implications of design choices become clear. From a constructivist perspective, a non-state legal actor's influence could both promote or undermine the legitimacy of the institutional design. A legal actor's distinct design preferences may better align with community norms and/or the preferences of those the institution is intended to serve, increasing institutional legitimacy. But this influence of an individual non-state legal actor also raises questions concerning representation, accountability, and the democratic legitimacy of the institutional design process and outcome (e.g. Bexell et al., 2010; Grant and Keohane, 2005).

Overall, this article highlights the importance of looking beyond states to understand institutional design. Scholars need to avoid post hoc rationalizations of states' design preferences when it is other actors' preferences influencing design outcomes.

Acknowledgements

The authors thank Sam Rowan, Fredrik Söderbaum, Anette Stimmer, Henning Tamm, Alexandra Zeitz, and their anonymous reviewers for their insightful and productive comments. They also thank Holden Carroll, Eugene Danso, Aidan Pierce, and Robert Zaleski for their research assistance.

Funding

The author(s) disclosed the receipt of the following financial support for the research, authorship, and/or publication of this article: Nicole De Silva received financial support for the research for this article from the Social Sciences and Humanities Research Council of Canada.

ORCID iD

Nicole De Silva  <https://orcid.org/0000-0002-7426-3466>

Supplemental material

Supplemental material for this article is available online.

Notes

1. The International Law Commission, Inter-American Juridical Committee, and African Union Commission on International Law, respectively.
2. Based on the ILC's (1987) recommendation, the UNGA renamed the Draft Code of Offenses against Peace and Security of Mankind as Code of Crimes against Peace and Security of Mankind to reflect the criminal responsibility attached to violating fundamental community norms.
3. For example, less than 25% of member states commented on the 1950 draft code.
4. Deya (2014) led the drafting first as the PALU Secretariat's sole employee and then as more staff became involved.
5. For example, PALU criminalized "mercenarism" considering states' obligation to domestically combat mercenarism in the International Convention against the Recruitment, Use, Financing and Training of Mercenaries (1989), Article 3.

References

- Abass A (2017) Historical and political background to the Malabo Protocol. In: Werle G and Vormbaum M (eds) *The African Criminal Court: A Commentary on the Malabo Protocol*. The Hague: T.M.C. Asser Press, pp. 11–28.
- Abbott KW and Snidal D (2001) Hard and soft law in international governance. In: Goldstein J, Kahler M, Keohane R, et al. (eds) *Legalization and World Politics*. Cambridge, MA: MIT Press, pp. 37–72.
- Abbott KW and Snidal D (2013) Law, legalization, and politics: an agenda for the next generation of IR-IL scholars. In: Dunoff JL and Pollack M (eds) *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art*. New York: Cambridge University Press, pp. 33–57.
- Abbott KW, Genschel P, Snidal D, et al. (2020) Competence-control theory: the challenge of governing through intermediaries. In: Abbott KW, Zangl B, Snidal D, et al. (eds) *The Governor's Dilemma: Indirect Governance Beyond Principals and Agents*. Oxford: Oxford University Press, pp. 3–38.
- Alter KJ (2008) Agents or trustees? International courts in their political context. *European Journal of International Relations* 14(1): 33–63.
- Barnett MN and Finnemore M (2004) *Rules for the World: International Organizations in Global Politics*. Ithaca, NY: Cornell University Press.
- Beach D and Pedersen RB (2016) *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press.

- Bexell M, Tallberg J and Uhlin A (2010) Democracy in global governance: the promises and pitfalls of transnational actors. *Global Governance* 16(1): 81–101.
- Boyle A and Chinkin C (2007) *The Making of International Law*. Oxford: Oxford University Press.
- Bradley CA and Kelley JG (2008) The concept of international delegation. *Law and Contemporary Problems* 71(1): 1–36.
- Brunnée J and Toope SJ (2010) *Legitimacy and Legality in International Law: An Interactional Approach*. Cambridge: Cambridge University Press.
- Büthe T and Mattli W (2011) *The New Global Rulers: The Privatization of Regulation in the World Economy*. Princeton, NJ: Princeton University Press.
- Carpenter RC (2007) Setting the advocacy agenda: theorizing issue emergence and non-emergence in transnational advocacy networks. *International Studies Quarterly* 51(1): 99–120.
- Charnovitz S (2006) Nongovernmental organizations and international law. *American Journal of International Law* 100(2): 348–372.
- Copelovitch MS and Putnam TL (2014) Design in context: existing international agreements and new cooperation. *International Organization* 68(2): 471–493.
- Crawford J (1995) The ILC adopts a statute for an International Criminal Court. *American Journal of International Law* 89(2): 404–416.
- Deitelhoff N (2009) The discursive process of legalization: charting islands of persuasion in the ICC case. *International Organization* 63(1): 33–65.
- Deya D (2012a) Is the African Court worth the wait. *Open Society Justice Initiative for Southern Africa*, 22 March. Available at: <https://allafrica.com/stories/201203221081.html> (accessed 1 November 2023).
- Deya D (2012b) Worth the wait: pushing for the African Court to exercise jurisdiction for international crimes. *OpenSpace* 2: 22–26.
- Durkee MJ (2016) The business of treaties. *UCLA Law Review* 63(2): 264–321.
- Elsig M and Pollack MA (2014) Agents, trustees, and international courts: the politics of judicial appointment at the World Trade Organization. *European Journal of International Relations* 20(2): 391–415.
- Finnemore M and Sikkink K (1998) International norm dynamics and political change. *International Organization* 52(4): 325–347.
- George AL and Bennett A (2005) *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Glasius M (2002) Expertise in the cause of justice: global civil society influence on the statute for an International Criminal Court. In: Glasius M, Kaldor M and Anheimer H (eds) *Global Civil Society Yearbook 2002*. Oxford: Oxford University Press, pp. 137–168.
- Grant RW and Keohane RO (2005) Accountability and abuses of power in world politics. *American Political Science Review* 99(1): 29–43.
- Green JF and Colgan J (2013) Protecting sovereignty, protecting the planet: state delegation to international organizations and private actors in environmental politics. *Governance* 26(3): 473–497.
- Haas PM (1992) Introduction: epistemic communities and international policy coordination. *International Organization* 46(1): 1–35.
- Hawkins DG and Jacoby W (2006) How agents matter. In: Hawkins DG, Lake D, Nielson DL, et al. (eds) *Delegation and Agency in International Organization*. Cambridge: Cambridge University Press, pp. 199–228.
- Hawkins DG, Lake D, Nielson DL, et al. (2006) Delegation under anarchy: states, international organizations, and principal-agent theory. In: Hawkins DG, Lake D, Nielson DL, et al. (eds)

- Delegation and Agency in International Organization*. Cambridge: Cambridge University Press, pp. 3–38.
- Helfer LR and Meyer T (2016) The evolution of codification: a principal-agent theory of the International Law Commission's influence. In: Bradley CA (ed.) *Custom's Future: International Law in a Changing World*. Cambridge: Cambridge University Press, pp. 305–331.
- Helfer LR and Slaughter AM (2005) Why states create international tribunals: a response to professors Posner and Yoo. *California Law Review* 93(3): 899–956.
- Holthoefer A (2016) Constructing international crime: lawyers, states, and the origin of international criminal prosecution in the interwar period. *Law and Social Inquiry* 42(3): 711–743.
- Johnson T (2014) *Organizational Progeny: Why Governments are Losing Control Over the Proliferating Structures of Global Governance*. Oxford: Oxford University Press.
- Johnson T and Urpelainen J (2014) International bureaucrats and the formation of international organizations: institutional discretion sweetens the pot. *International Organization* 68(1): 177–209.
- Johnstone I (2013) Law-making by international organizations: perspectives from IL/IR theory. In: Dunoff JL and Pollack M (eds) *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art*. New York: Cambridge University Press, pp. 266–292.
- Keck ME and Sikkink K (1998) *Activists Beyond Borders*. Ithaca, NY and London: Cornell University Press.
- Koremenos B (2005) Contracting around international uncertainty. *American Political Science Review* 99(4): 549–565.
- Koremenos B (2008) When, what, and why do states choose to delegate? *Law and Contemporary Problems* 71(1): 151–192.
- Koremenos B (2016) *The Continent of International Law: Explaining Agreement Design*. Cambridge: Cambridge University Press.
- Koremenos B, Lipson C and Snidal D (2001) The rational design of international institutions. *International Organization* 55(4): 761–799.
- Lewis M (2014) *The Birth of the New Justice: The Internationalization of Crime and Punishment, 1919–1950*. Oxford: Oxford University Press.
- Liese A, Herold J, Feil H, et al. (2021) The heart of bureaucratic power: explaining international bureaucracies' expert authority. *Review of International Studies* 47(3): 353–376.
- Lyne MM, Nielson DL and Tierney MJ (2006) Who delegates? Alternative models of principals in development aid. In: Hawkins DG, Lake D, Nielson DL, et al. (eds) *Delegation and Agency in International Organization*. Cambridge: Cambridge University Press, pp. 41–76.
- Morrow JD and Cope KL (2021) The limits of information revelation in multilateral negotiations: a theory of treaty-making. *Journal of Theoretical Politics* 33(4): 399–429.
- Morse JC and Keohane RO (2014) Contested multilateralism. *Review of International Organizations* 9(4): 385–412.
- Murungu CB (2011) Towards a criminal chamber in the African Court of Justice and Human Rights. *Journal of International Criminal Justice* 9: 1067–1088.
- Pollack MA (2003) *The Engines of European Integration: Delegation, Agency, and Agenda Setting in the EU*. Oxford: Oxford University Press.
- Rietig K (2014) “Neutral” experts? How input of scientific expertise matters in international environmental negotiations. *Policy Sciences* 47(2): 141–160.
- Schabas WA (2007) Origins of the genocide convention: from Nuremberg to Paris. *Case Western Reserve Journal of International Law* 40: 35–56.

- Schabas WA (2011) International war crimes tribunals and the United States. *Diplomatic History* 35(5): 769–786.
- Sell SK (2003) *Private Power, Public Law: The Globalization of Intellectual Property Rights*. Cambridge: Cambridge University Press.
- Snidal D and Tamm H (2018) Rational choice: from principal-agent to orchestration theory. In: Weiss TG and Wilkinson R (eds) *International Organization and Global Governance*. New York: Routledge, pp. 135–145.
- Stimmer A and Wisken L (2019) The dynamics of dissent: when actions are louder than words. *International Affairs* 95(3): 515–533.
- Struett MJ (2008) *The Politics of Constructing the International Criminal Court: NGOs, Discourse, and Agency*. New York: Palgrave Macmillan.
- Tallberg J, Sommerer T, Squatrito T, et al. (2013) *The Opening Up of International Organizations*. Cambridge: Cambridge University Press.
- Thompson A (2009) Rational design in motion: uncertainty and flexibility in the climate regime. *European Journal of International Relations* 20: 1–34.
- Tieku TK (2021) Punching above weight: how the African Union Commission exercises agency in politics. *Africa Spectrum* 56(3): 254–273.
- Voeten E (2019) Making sense of the design of international institutions. *Annual Review of Political Science* 22: 147–163.

Author biographies

Nicole De Silva is an Associate Professor in the Department of Political Science at Concordia University in Montreal, Canada. She holds a DPhil in International Relations from the University of Oxford. Her interdisciplinary research focuses on the role of international organizations and international law in global and regional governance.

Anne Holthoefler is an Associate Professor in the Department of Politics at St. Anselm College in Manchester, USA. She holds a PhD in International Relations from the University of Chicago. Her research focuses on legal norm development and institutional design with special interest in the historical development of international criminal law.