Dear Ostrom Workshop Participants,

Allison and I are excited to have the opportunity to present our work. We are sharing a revised version of the theory chapter from our current book project. This chapter was completely re-written in the last month. You will notice a few signs of its "in development" status: typos; a lack of citations; few clarifying footnotes. There is more to do on those fronts. That said, we think this new version of the chapter is a big improvement from our previous draft. We are eager for feedback on how to further improve it.

For context, I'm pasting below the title, abstract, and chapter structure of the book. Our empirical chapters are nearly finished. A version of the nuclear chapter is now forthcoming in *American Journal of Political Science*, much to our delight. We are hoping to complete the book manuscript by end of spring.

Thanks again for reading!
Austin Carson
University of Chicago

Secrets in Global Governance: Disclosure Dilemmas and the Challenge of International Cooperation

Allison Carnegie & Austin Carson

Abstract: How do international organizations (IOs) help states cooperate? Scholars have long argued that they address information problems by increasing transparency. This book introduces a distinct information problem that threatens cooperation and argues that IOs can mitigate it using secrecy instead. We show that states often possess information about other countries' violations of international rules but are deterred from disclosing it by concerns about revealing sensitive details, such as commercially sensitive economic information or the sources and methods used to collect intelligence. IOs equipped with effective confidentiality systems, however, can analyze and act on sensitive information while preventing its wide release. Using a multi-method approach including statistical analyses of new data, elite interviews, and original archival research, the book tests this argument in domains across international relations, including nuclear proliferation, international trade, justice for war crimes, and foreign direct investment. We show empirically that sensitive information can endanger cooperative goals, but that properly equipped IOs can elicit disclosures of such information and thereby increase international cooperation. Yet we also argue that this role for secrecy in IOs is in tension with normative goals of transparency and creates a new pathway for states to exert power in international relations. The book targets scholars of international political economy, international security, international institutions, and power in global governance.

Chapter 2

Theory

This chapter develops a theory to explain how states handle information they obtain about whether governments are upholding their cooperative commitments. States with unique insights about compliance have two basic options: disclose it or keep it private. Our theory posits that this choice depends on whether the information is *sensitive* or *non-sensitive*. While decisions about whether to reveal non-sensitive information are straightforward, states with insights about compliance often struggle with cross-cutting incentives about whether to disclose it. The sections that follow define the key concepts that we use to understand the politics of sensitive information revelation, unpack the appeals and dangers of such disclosures, and develop how international organizations with special secrecy features offer a solution that eases dilemmas surrounding disclosure decisions.

One reason these kinds of information decisions merit analysis is because states often encounter richer, more precise forms of evidence regarding compliance than do other types of actors. Despite the growth of global governance and national and transnational civil society, sovereign states have considerable advantages over non-state actors in identifying violations. As we argue, this infuses state decisions with large, systemic implications for the success of cooperative regimes in general. A second reason that states' disclosure decisions

are important to theorize is that they give rise to otherwise confusing patterns of behavior. Consider the following three episodes.

- In the mid-1990s, the American government helped to establish and fund a war crimes tribunal to punish war criminals from the former Yugoslavia. Despite lending moral, diplomatic, and financial support, Washington refused to share unique insights linking top leaders to the atrocities with the tribunal. The lead prosecutor soon complained that the lack of evidence would prevent indictments and prosecution.

- In 1998, Brazil accused Canada of violating free trade commitments by subsidizing the Canadian aircraft industry at the expense of Brazil's. The Canadian government possessed a trove of information that could help exonerate it from the claim of trade discrimination but held much of it back.

- In 1978, the United States encountered clear indications that Pakistan misled Washington about its secret nuclear weapons program. Internal American debates recognized that publicizing this information would facilitate multilateral prevention and the punishment of Pakistan, but the U.S. withheld its knowledge.

What explains these puzzling decisions? In each case, an informed state held back potentially useful information. At times this increased the perception that it was violating its cooperative commitments, as in the Canada/Brazil trade example. In others, the informed state sacrificed transnational goals that carried security and moral benefits, such as war crimes accountability and nuclear non-proliferation. While each example features variation in the substance of the information, each shows that states with unique insights about compliance often balk when intuition would expect them to disclosure it.

We theorize why states are reluctant to reveal such information, arguing that wide access to these insights can often allow harmful adaptations by others that is unrelated to compliance outcomes. We then develop how international organizations (IOs) that can protect

sensitive information frequently overcome this problem by allowing states to narrowly disclose sensitive details, thereby avoiding the problems that accompany wider dissemination while also allowing these insights to inform compliance-related disputes.

A critical implication of the theory is that it reverses important elements of the conventional wisdom about international institutions. As we note in Chapter 1, the typical scholarly view of IOs highlights their value as information conveyor belts, improving access to information about shared problems and compliance questions. While we agree that such a function enhances IOs' efficacy regarding non-sensitive information, we argue that when sensitive information is involved, a conveyor belt function exacerbates adaptation problems. Our theory instead suggests that IOs need to serve as a kind of bank vault with respect to sensitive information, protecting and securing it from broad scrutiny.

We further argue that sensitive information is often critical to solving key cooperation problems in international relations, such as securing war crimes indictments, solving trade disputes, or identifying well-hidden nuclear facilities. In short, whether IOs are endowed with a secrecy function may make the difference between the IO's success and failure. Yet there is a downside to secrecy as well: informed states retain discretion about their disclosures, empowering them to disclosure or withhold information according to their strategic interests, and thereby subtly influencing the operation of global governance.

2.1 The Challenge of Demonstrating Non-Compliance

To achieve international cooperation, timely and accurate information is essential. In particular, states must be able to determine which of their partners are cheating on their agreements in order to punish these infractions and deter future breaches. Otherwise, if noncompliant states can avoid detection, they can exploit compliant ones, which can discourage cooperation

from occurring in the first place.[1] Scholars and practitioners have embraced transparency as a means to obtain this information, to such an extent that the critical role of "compliance information to facilitate compliance with international agreements," has become "a centerpiece of neoliberal institutionalism."[2]

However, such information can be difficult to obtain. A central problem in any rules-based international political order is poor information and, more specifically, the difficulty of identifying and documenting instances of cheating. This issue can arise in domains that feature rules that prescribe or proscribe behavior and symbolic, reputational, or material penalties that come from violating those rules. This information problem thus pertains both to behavior that is regulated by norms as well as that governed by more formalized treaties and legal principles. Detecting non-compliance often requires specialized techniques or knowledge that only specific states or non-state actors can gather, especially since rule breakers typically try to hide their behavior.[3] For example, only states with large intelligence bureaucracies may gather intelligence about clandestine nuclear facilities, and only firms in affected sectors may hold detailed data about the impact of specific trade barriers. This asymmetry in knowledge is quite common, in part because institutions that states create to monitor compliance face legal and resource constraints that limit their ability to independently discover and document non-compliance.

We focus on scenarios with particularly sharp information problems about compliance in which a state or non-state actor engages in rule-violating conduct that is only *partially* observable. Our theory treats other situations – i.e., where a suspected violator's behavior is totally undetected by others or nakedly and widely observable – as subject to distinct logics, problems, and solutions. Disclosure dilemmas thus arise when intentional evasion,

[1] Keohane 1984*a*; Axelrod and Keohane 1985; Milgrom, North et al. 1990*b*; Mitchell 1998; Koremenos, Lipson and Snidal 2001; Dai 2002*a*; Lindley 2004; Carrubba 2005; Voeten 2005; Thompson 2006*a*; Lindley 2007; Guzman 2008.

[2] Dai 2002*a*, 409.

[3] Hafner-Burton 2008*a*.

legal barriers or logistical challenges make monitoring difficult but not impossible. While scholars have long established that state and non-state actors *can* play a role in monitoring compliance,[4] we focus on distinct informational dynamics that arise when compliance behavior is narrowly observed by one or a few states. In short, we ask: How do states behave when their disclosures can close informational gaps regarding compliance?

The basic scenarios our theory addresses therefore follows a straightforward structure. Depending on the kind of behavior in question, a state may observe a violation through routine bureaucratic or legal reporting, or through specific acts of solicitation or clandestine detection. While we often discuss the informed state an the violator as separate entities, they can also be the same actor; for example, a state could detect its own behavior, or a firm could do so. In such situations, an international organization (IO) – a formal supranational entity created by states to facilitate joint action to address a common problem – may be present. A given IO can perform a range of functions including serving as a forum for negotiation, gathering and disseminating information, providing services like peacekeeping or technical aid, and formally adjudicating accusations of non-compliance. We describe a feature of IOs that has not received scholarly attention: the capacity to keep secret the information that states submit. The dominant focus of research to date has instead focused on variation in how effectively IOs disseminate information.[5]

2.2 Defining the Key Concepts

This book describes a specific information problem that states often confront, links this problem to challenges of international cooperation, and theorizes an institutional solution which serves to address the problem. The resulting theory builds on a set of terms and concepts which unify a narrative that spans everything from nuclear proliferation to foreign

[4]See Dai.
[5]Mitchell; Dai

direct investment. The following concepts provide the intuitive, unifying framework for the theoretical and empirical analyses that follow.

- **Sensitive information** Private information whose wide dissemination would allow harmful changes in behavior by other state and non-state actors. This book focuses on two specific kinds: national intelligence and firm-specific data/documents.
- **Disclosure dilemmas** Situations in which sharing sensitive information would both advance multilateral cooperative goals and allow behavioral changes that create adaptation costs.
- **Incrimination benefits** Political gains that result from improved compliance outcomes – incriminating others or exonerating oneself or one's friends – when private information is revealed.
- **Adaptation costs** Strategic, financial, and other harm inflicted when state or non-state actors alter behavior in light of newly revealed private information.
- **Confidentiality systems** The organizational policy and infrastructure for managing and limiting access to specific kinds of information considered sensitive.

The remainder of this chapter builds muscle and ligaments onto this conceptual skeleton. We do so by explaining the nature of sensitive information and their adaptation costs, the incrimination-related appeal of disclosing sensitive information, and how IOs' confidentiality systems can ease disclosure dilemmas.

2.3 The Problem: Disclosure Dilemmas and Sensitive Information

Disclosure dilemmas describe a broad class of informational problems that states regularly confront anytime they acquire unique information that is relevant to their cooperative commitments. A dilemma arises when an informed state faces competing incentives – incrimination benefits and adaptation costs – to disseminate information. A dilemma does not

exist if an informed state finds little benefit from revealing what it knows or faces minimal costs from the information's wide disclosure (see Figure 2.1). The empirical analyses presented in later chapters demonstrate that disclosure dilemmas are common in international relations, arising in diverse empirical domains including peacekeeping, trade disputes, and nuclear proliferation.

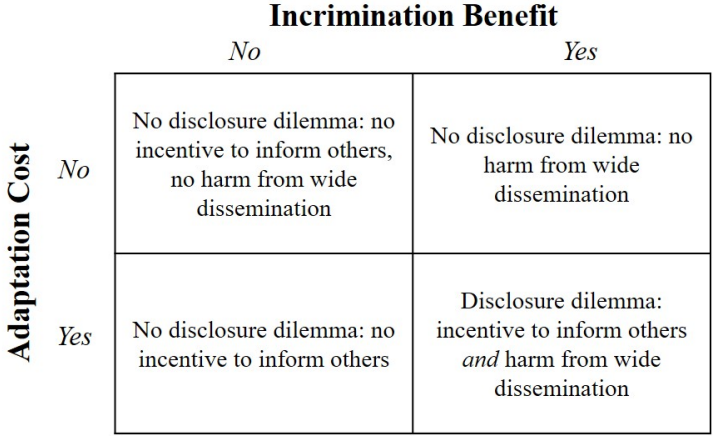

Figure 2.1: Two Necessary Conditions for Disclosure Dilemmas

The next sections describe incrimination benefits and adaptation costs in more detail, specifying their particular manifestation for the two kinds of information: national intelligence and firm-specific economic details. While other kinds of information may be sensitive and relevant to compliance issues, such as the name and private details of individuals involved in a war crime, we focus in this book on these two specific kinds of insights as a lens into the broader phenomenon. The section then addresses why removing sensitive details is a poor solution to disclosure dilemmas and why some situations fail to give rise to a dilemma in the first place.

2.3.1 Intelligence: Adaptation Costs and Incrimination Benefits

Sensitive information in the security realm tends to take the form of intelligence, or the collection of information of military or political value via clandestine means. Common types

of intelligence include human intelligence (HUMINT), which include the secret use of foreign sources to transfer classified data(**?**, 294); signals intelligence (SIGINT), which is "intelligence comprising, either individually or in combination, all communications intelligence, electronic intelligence, and foreign instrumentation signal intelligence, however transmitted" (**?**, 203); and imagery intelligence (IMINT), which collects photographic information obtained via satellite and aerial photography. While some intelligence is collected by most states, the United States is the leader in intelligence collection. Other states with strong intelligence collection capabilities include Russia, China, India, Pakistan, the UK, Germany, France, and Israel.[6] These states collect many types of intelligence.

States often benefit from making intelligence-derived insights more widely available. Sharing with allies might help improve balancing against a common foe; publicizing intelligence might build the case for war. We focus on a particular benefit specific to international cooperative commitments: using intelligence to incriminate other states and non-state actors, by demonstrating their abrogation of international norms or laws. Disclosing intelligence findings that pertain to others' misbehavior can have practical, diplomatic, or strategic effects. For example, sharing intelligence that shows violation of an arms control accord can encourage a multilateral response to coerce the violator into coming back into compliance. This can happen by altering public opinion of the violating country, persuading other heads of state, or simply by creating unwanted scrutiny which deters other potential violators.[7] A famous case of illustrating several effects of "going public" is the Cuban Missile Crisis. Publicizing American intelligence about Soviet missiles deployed to Cuba, for example,

[6]"The List: The World's Top Spy Agencies." Foreign Policy The List The Worlds Top Spy Agencies Comments. 13 July 2015.

[7]Chesterman (2006*b*, 2009). See also Allison Carnegie and Austin Carson. 2018. "The Spotlight's Harsh Glare: Rethinking Publicity and International Order," *International Organization*. Exposing breeches in international norms and laws could also influence domestic public and legislative opinions in the disclosing state, even facilitating a leader's policy agenda or re-election Hastedt (2005). Our theory only incorporates domestic politics to the extent that it influences compliance outcomes rather than parochial domestic political priorities of the informed state.

had an enormous impact on domestic political support for President Kennedy.[8] Yet it also demonstrated Soviet abrogation of a pledge not to place offensive weaponry in the Americas, casting doubt on Soviet credibility and encouraging unified pressure to reverse course.

Because of the expense and sophistication of intelligence collection, its insights can provide distinct and sometimes decisive evidence regarding compliance. Yet disclosure of intelligence, whatever the political goal, carries one ever-present risk: exposure of sources and methods. States are reluctant to share sensitive intelligence because revelations tend to expose the clandestine techniques used to collect the information, allowing targets to adapt their activities to avoid future detection.[9] For example, releasing facts derived from signals intelligence, such as intercepted phone call transcripts, can lead a target to change its communication channels. Or, sharing information provided by a human source can result in that source's expulsion, imprisonment, or death. Providing imagery intelligence can disclose that a particular site is under observation and allow the target to move or obscure it. Due to the massive expense associated with developing new sources of intelligence, countries often withhold intelligence from international and domestic audiences.[10]

The Cuban Missile Crisis again illustrates. While the revelation of American intelligence had enormous impacts on diplomacy, it also revealed sensitive sources and methods the U.S. used to obtain the photos. Internal U.S. memos from 1963 show that the U.S. was aware that it had given up important sources and methods that could be used against it. Referring to the Cuban Missile Crisis, one memo states, "US intelligence was goaded into revealing so much about its workings, from initial collection through analysis to the part it played in policy decisions, that the Soviets will know exactly what mistakes to avoid in the future."[11] Similarly, in a memo entitled "Disclosures of U.S. Intelligence Methodology," the Director

[8] See Brugioni (1991, 428-429).
[9] See Richelson (1990, 1997).
[10] See Lefebvre (2003, 523).
[11] "Possible Soviet Courses of Action in and with Respect to Cuba." Memorandum for the Director. CIA. Office of National Estimates. March 13, 1963.

of the CIA states "I wish to invite your personal attention to the attached intelligence 'damage' assessment, which describes the extent to which public disclosures concerning the Cuban missile crisis have exposed US intelligence sources and methods. This study reaches the obvious conclusion that U.S. intelligence collection ability has been impaired by these disclosures."[12]

To be clear, sharing intelligence is not always costly; countries only incur a penalty from doing so if other parties will adapt in ways that harm the informed countries. In some cases, countries are unable to adapt if they do not possess the means or technical know-how to do so. For instance, countries could learn of a satellite monitoring their nuclear facilities but lack the funds to move the program underground, or could hear of a new weapon being developed but lack the ability to develop it themselves. Alternatively, adaptation may occur whether or not the information is revealed if others are already aware of the collection method or if the exposure of a source or method is inevitable for some other reason such as an technological advance or a leak. In such a case, revealing the information costs little since the sensitive portion will be exposed regardless.

2.3.2 Firm-Specific Information: Adaptation Costs and Incrimination Benefits

The second type of sensitive information that we consider is firm-level economic data nd documentation. Due in part to increased regional and global economic integration, firms now possess troves of sensitive information that is relevant to international trade disputes, transnational crime, monetary policy, and finance. Such information tends to come in two types. First is intellectual property, which is the "legal rights that correspond to intellectual activity in the industrial, scientific, and artistic fields" and includes "patents, copyrights,

[12] "Disclosures of US Intelligence Methodology." Memorandum. The Director of Central Intelligence. March 29, 1963.

proprietary info like customer lists and information, product development data, pricing data, sales figures, marketing plans, personnel data, bid information, manufacturing costs analyses, and strategic planning information" (Nasheri 2005). In other word, these are the concepts and ideas that drive businesses such as recipes and formulas. This kind of information is most relevant to industries with high research and development costs, such as technology companies, pharmaceuticals, manufacturing, service industries, etc. Second is operation information, which includes firms' marketing, supply chain, and production details. These include items such as market share, locations of productions, customer lists, and production costs (Nasheri 2005). Banking information can also be relevant, such as a bank's loan portfolios or liquidity.

The allure of revealing sensitive firm-level information differs somewhat from the national intelligence context. For one, a state can collect or solicit sensitive firm- and bank-related details from within its territory; national intelligence is collected on foreign targets. Countries do sometimes possess detailed economic information about foreign countries and foreign firms but revealing such details does not impose harm on the disclosing state and is thus non-sensitive as we define it.[13] In the economic domain, we focus on the incentive firms and their governments have for disseminating firm-specific details that demonstrate violations by foreign governments or exonerate themselves when accused of violations.

A few examples demonstrate the kinds of incrimination benefits when we move to the world of firms. Suppose a government prefers to help regional efforts to crack down on the drug trade. Revealing firm-level information about production or sales levels in certain industries could be critical to identifying who is illegally using chemical precursors for narcotics.[14]

[13]For example, Germany revealing a Chinese firm's profit is typically not costly for the German government or its firms. An additional kind of relevant information bears on the health of the economy such as macroeconomic indicators, as well as economic policy changes. Publicly available macroeconomic indicators can be sensitive when the government alters the information, or when actors seek to obtain them before they are officially released (Hautsch, Hess and Veredas 2011).

[14]ICNB example explained.

Now suppose a government seeks to protect the integrity of a trade deal. Information about changes over time to one of its firm's profit margins and financing could be critical to a) documenting trade harm done by a foreign government's domestic subsidy program; or, b) defending the firm's home government from claims of unfair subsidization. Finally, suppose a government wishes to protect its reputation for abiding by investment agreements. Facing the need to address perceptions of foreign investment expropriation might opt to "open its books" to allow markets and the public to see that its behavior complies with an investment agreement (**?**, 14).[15]

However, just as in the intelligence context, widely disseminating firm-specific information to address cooperation-related goals has a significant downside: market adaptations. The primary concern is that revelation can trigger adaptation by economic competitors or broader market punishments. Firm-level data is sensitive because it often forms the basis of a firm's competitive advantage. Against the benefits of cracking down illegal drugs or trade discrimination, firms may worry that opening their books to reveal certain details could jeopardize their ability to compete and damage their profitability. Companies can lose market share if they are no longer able to get a product not the market first, or at the lowest cost, or lose out on obtaining a patent. While publicly traded firms are required to disclose financial information, they typically possess troves of additional detailed information which is kept private for competitive advantage.[16] Such information is often referred to as "commercially sensitive," and while there is no official legal definition of such information, it is "generally understood to be any information that has economic value or could cause economic harm if known" (**?**, 33).

Firm-level documentation on pricing or loan structures can provide critical insights to

[15]One or both parties might push for disclosure in cases where the data's availability would enable better contract management and enforcement vis-a-vis the other (**?**, 17).

[16]This can include legitimate activities as well as more dubious secrets, such as non-performing loans held off of balance sheets or involvement with shell corporations (**??**).

rival firms and risk negative investor reactions.[17] This was illustrated in a recent trade dispute between South Korea and the European Union (EU) over shipbuilding, in which South Korea refused to provide details on input costs that were essential to adjudicating the dispute, citing the risk that they would put its firms at a competitive disadvantage.[18] As a result, the EU and third parties could not ascertain whether South Korea made this claim to avoid producing evidence of its wrongdoing or because the information was, in fact, commercially sensitive. Moreover, misperceptions sharpen the problem. Revealing sensitive information about a bank's loan portfolio can prompt divestment or other forms of market punishment even if the bank is healthy, due to the complexity of this data (Hollyer, Rosendorff and Vreeland 2014, 417)[19]

To be clear, while we posit that adaptation costs often follow the public circulation of sensitive information, they do not always do so, as others may not adapt in a harmful manner as a result of the information's exposure. Other actors may be unable to change quickly, will obtain marginal strategic or economic advantages from the information, or may adapt regardless of whether sharing takes place, especially if the information is likely to leak or be disclosed for other reasons. For example, firms and countries may be comfortable revealing sensitive information if it is dated, or sensitive data may become available even if the country seeks to protect it through leaks or the expiration of a patent. In these cases, disclosure dilemmas do not exist since countries have no meaningful adaptation cost from firm and market reactions. Moreover, proprietary firm-level information is often irrelevant to other competitors and to questions of compliance with international rules. A dilemma

[17] A European Commission report that surveyed over 500 European firms concludes, "trade secrets and their protection appear to be important to all business sectors, reflecting their pervasiveness and importance to virtually all firms in EU Member States, regardless of their size." Study on Trade Secrets and Confidential Business Information in the Internal Market, Prepared for the European Commission, April 2013.

[18] Barnard, Bruce. "EU poised to file complaint with WTO over South Korean shipbuilding practices." *Journal of Commerce*. October 3, 2000.

[19] See also Colleen McCain Nelson, "U.S. Rethinks How to Release Sensitive Economic Data Potential Changes Driven From Unease Over High-Speed Trading Firms," Wall Street Journal, October 10, 2013.

only applies when countries face a genuine cost (and benefit) from disclosing information.

2.3.3 The Necessity of Widely-Sharing Sensitive Details

To summarize, we analyze two kinds of sensitive information that is relevant to questions of compliance. Each kind of information is tempting to reveal, especially as a method to document violations of laws and norms by others or to defend against such claims. Yet wide dissemination of each kind of information creates harm for the disclosing actor, which we broadly refer to as "adaptation costs." Figure 2.2 summarizes these points.

Information	Relevance to Compliance	Example	Adaptation Cost	Incrimination Benefit
National intelligence	Intelligence insights can uniquely document foreign government violations of norms/laws	Imagery and signals intelligence link a specific leader to a wartime atrocity	Exposing collection methods allows targets to adapt, avoid surveillance	Punishing rivals and improving compliance with international rules
Firm-specific economic information	Firm details can uniquely document foreign government violations of norms/laws Firm details can refute claims of home government violating norms/laws	Trends in profit margin data show damage from a foreign domestic subsidy program Detailed lending terms show lack of home government preferential treatment	Exposing operational details to market rivals allows market changes that threaten profitability	Punishing agreement defectors, defusing accusations of violations, improving compliance with international rules

Figure 2.2: Two Kinds of Sensitive Information

This discussion raises an important question. Why can't an informed state remove sensitive details and publicly disclose its conclusions? In the intelligence world, this is referred to as "scrubbing," where states remove sensitive sources and methods details before sharing information. In the economic domain, this might take the form of hiding firm-specific details but sharing aggregate statistics that speak to whether a foreign investment project was harmed.

Despite its allure, this option has two major flaws – one practical and the other political – that illustrate the necessity of disclosing sensitive details. First, withholding sensitive portions often denies other states and non-state actors the very details which make a disclosure

useful. For regimes with strong legalization, a sanitized disclosure will not provide sufficient context and precision to be useful for issuing criminal indictments or reaching a definitive conclusion about an investment dispute. For example, documenting aggregate summation of harm is often insufficient in trade disputes. Firm-specific balance sheets or contracts with suppliers can be the central axis on which a judgment of compliance or non-compliance hinges. Outside legalized contexts, removal of sensitive details can cause a disclosure to raise more questions than it answers. An informed state sharing only intelligence conclusions about a ceasefire violation, for instance, will not provide a relevant peacekeeping mission with sufficient detail to validate the claim and gather supplemental information. It will appear to be a simple assertion without evidence.

The second problem is political, as only releasing conclusions creates a credibility issue. An informed state making a scrubbed disclosure almost always has a stake in the outcome, and scrubbing removes details which can overcome known bias. Without precise details, various audiences including other states, IOs, and NGOs – struggle to evaluate the validity of the claim. As Chesterman (2006*b*, 21) notes regarding intelligence, "the lack of information about the source makes it difficult to assess reliability...a reasonable response of the recipient is to suspect that they are being manipulated."[20] Banks and firms cannot simply assert their compliance with laws or the soundness of their performance since there are obvious market and political incentives to misrepresent the information, just as states have incentives to misrepresent their private information (Fearon 1995). When others know that the informed country has a vested political or strategic interest in the outcome, they will tend to distrust claims that lack details about sources and methods or specific firms and their activities.

A possible alternative is to disclose sensitive information, but only to trusted and friendly

[20]Observers of scrubbed disclosures are aware that this practice has allowed intelligence to be manipulated historically. On manipulated intelligence to influence ally behavior in war, see JN Brown, DL Lupton, A Farrington. "Embedded Deception: Interpersonal Trust, Cooperative Expectations, and the Sharing of Fabricated Intelligence." Journal of Global Security Studies, 2018.

governments. While this may sometimes suffice if a state only seeks the cooperation of a few other states, such an approach typically provides a much smaller political benefit than the wide dissemination of information.[21] The disclosure dilemma we analyze here arises so frequently precisely because of a mismatch between a) the number of governments and publics that need to be persuaded in order to capture cooperation gains in a multilateral setting; and, b) the number and identity of governments trusted enough to receive raw, sensitive intelligence.[22] Thus, given the weakness of the alternatives, the most common reaction to the problem of sources and methods sensitivity is intelligence non-disclosure.[23] The joint effect of these two problems foreshadows the institutional solution that we develop subsequently, as international organizations with confidentiality systems can allow informed states to provide sensitive details *only to the IO*, enabling both third party validation and avoiding the adaptation costs that result from the information's broader disclosure.

2.3.4 Other Factors in Disclosure

Our theory posits that states with unique information about compliance make decisions about whether to disclose information based on the resulting adaptation problems and incrimination benefits. As with any theory, these claims necessarily ignore a range of other considerations that may bear on such disclosures. To list only a few, informed states may also weigh the domestic political risks or benefits of releasing sensitive information, wield it as leverage in private bilateral diplomacy, or prefer not to maximize compliance or otherwise support the relevant cooperative goal.

[21] On the drawbacks of intelligence sharing agreements and the limited forms of collaboration they give rise to, see (Walsh 2010).

[22] In the nuclear domain, the target audience is often other member-states of the IAEA serving on its Board of Governors. American leaders seeking greater scrutiny or punishment of a given proliferator might only trust raw intelligence with a handful of members (say, the United Kingdom and Australia) but actually need votes and support from others (say, Ghana and Estonia).

[23] NGOs also represent possible solutions, and fit into our framework as well. We discuss this in the conclusion.

While many of these considerations are clearly relevant, theoretical parsimony and clarity requires narrowing the field of relevant factors. In our empirical analyses, we control for many of these factors to try to isolate the effects of the variable of theoretical interest. Moreover, an important implication of our framework is that many of these other factors obviate the dilemma on which we focus. Consider two examples. First, an informed state knows about non-compliance but prefers to undermine the relevant institution and the goal of, say, limiting nuclear proliferation. For this state, there is no incentive to share its sensitive information since it does not obtain any incrimination benefits from doing so. The solution in these cases is to simply stay silent, both to avoid adaptation costs and avoid improving compliance for a distasteful regime. In a second scenario, an informed state's leader will win domestic political points by going public with sensitive information that exonerates it from, say, claims of ceasefire violations. Again, the state's decision about disclosure is easy to make: it will publicly release the information to reduce the chance of its own incrimination and to receive a domestic political bonus. In both, disclosure dilemmas are not present. This book is about scenarios in which incentives compete, which can result in sub-optimal outcomes without an institutional solution.

2.4 Solution: Confidentiality in International Organizations

Since Keohane (1984), scholars have highlighted the importance of reducing uncertainty about cooperation and compliance via third parties like international organizations. An important implication is that improving transparency – a traditional function of IOs – can *sharpen*, rather than alleviate, cooperation problems for states. If provided sensitive details, an IO that widely disseminates information will exacerbate adaptation costs by expanding the state and non-state actors who can adapt in harmful ways.

This section develops an alternative function for IOs: protecting and vetting sensitive details. Rather than using the metaphor of conveyor belts which funnel along information, we conceptualize IOs with confidentiality systems as informational bank vaults when information is sensitive. A properly designed IO thus adds a third option between wholesale public disclosure and keeping sensitive information private. To elicit state submissions of sensitive information, *secrecy* rather than transparency is critical because the informed state must have confidence that the IO will carefully protect its sensitive details. This targeted submission also allows the institution to act on details and data that would otherwise be unavailable, improving the chances that cooperative gains are realized. In short, IOs can change states' disclosure calculus by lowering adaptation costs while maintaining incrimination benefits.

To be clear, an institution serving this bank vault function can also widely disseminate *non-sensitive* details and conclusions. Indeed, we develop below how the dissemination of conclusions from sensitive disclosures is critical. Yet the overall process we describe has the paradoxical combination of transparency via secrecy. That is, information protection is embedded within a larger transparency function for IOs when sensitive information is key to understanding compliance. If an IO can assure countries that sensitive disclosures will be protected, then disclosures will be more common, potential violators will be less likely to violate their agreements, and cooperation is more likely to occur in the first place.

We argue that IOs must have two key characteristics to serve this function: the organizational capability to safeguard sensitive information and a reputation for technocratic, unbiased assessment. This allows an IO to serve as a trusted third party, protecting some information while credibly assessing the validity of its content. The way in which IOs integrate sensitive information can be captured in three phases, as shown in Figure 2.3: the receipt and protection of sensitive details; the validation of the accuracy of disclosures and; the wide dissemination of conclusions and supplemental detail. The sections that follow unpack each stage and explain how each unfolds for both national intelligence and firm-specific eco-

nomic details. We then detail how a successful confidentiality function improves underlying cooperative goals as well as why institutions may or may not adopt confidentiality systems.

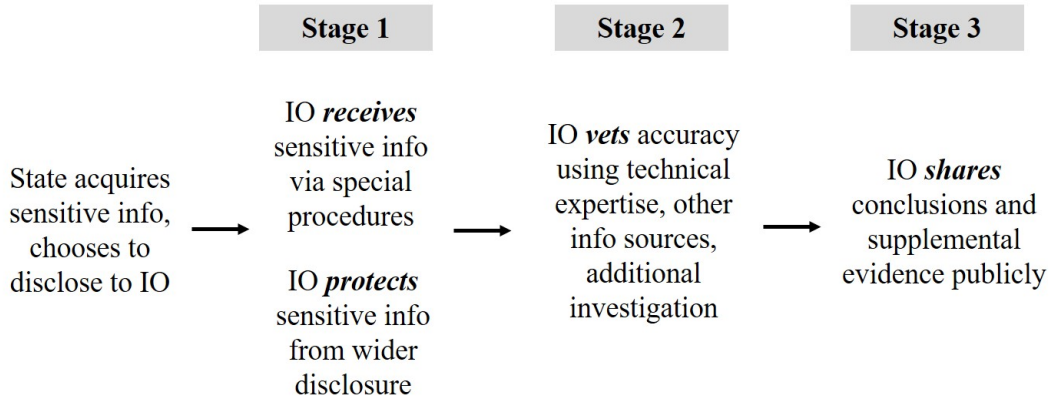

Figure 2.3: How IOs Handle Sensitive Disclosures

2.4.1 Receiving and Protecting Sensitive Disclosures

In the first phase, an informed member-state transmits sensitive material to an IO, which then must receive and protect it to avoid its wide dissemination. In practice, such transmission can take a number of forms. A briefing team from a country's national intelligence unit, for example, might meet with select members of an IO's secretariat to share details on a compliance case. Alternatively, a member-state might electronically transmit a packet of documents and data from specific firms to a team of trade dispute specialists or jurists. In the most routinized systems, an IO can establish a channel for the regular delivery of sensitive data via secure internet-based portals or an intelligence liaison.

Once received, an IO must securely store sensitive information to limit access to authorized personnel. Put differently, an IO must develop the rules and routines of organizational secrecy. Secrets on a small, individual scale are difficult enough to keep,[24] but organizational secrecy is especially challenging because protected information is distributed across

[24]Bok

more people and can be influenced by bureaucratic rivalry and mistaken disclosures.[25] Organization theorists analyzing the requirements for effective information security in private and public bureaucracies point to several features:[26] Organizations typically require a set of rules classifying the sensitivity of documents and data, the creation of physical and cyber security measures to protect designate data from broader access within the organization, clear policies identifying staff eligible to access information, the routines for accessing and handling it, and punitive policies punishing its disclosure.

How does this work in an international organization? Much depends on the specific level of confidentiality requested, the format of the information, and its manner of transmission. On one end of the spectrum, information disclosure can be ad hoc and highly confidential. An in-person intelligence briefing to a handful of senior secretariat officials requires any materials left behind to be securely stored in a physical safe. On the other end of the spectrum, a trade dispute may require a large volume of business sensitive details to be shared among technical experts, dispute panelists, and relevant lawyers. This can necessitate the development of a secure cyber transmission and storage system. Depending on the context, an IO may need to develop a system that can identify and regulate access to sensitive documents, categorizing them by their degree of sensitivity and developing policies governing different levels of access. The IO may also need measures to securely store data and documents, using physical lock-and-key systems for "hard" data and encryption and other information technology for "soft" data. These measures may also include personnel rules that establish how employees should handle sensitive information and the penalties for unauthorized disclosures.

The unifying theme across these different information management techniques is minimizing the risk of information's wide disclosure. A lingering and ever-present risk for any organizations serving a confidentiality function is leaks. Unauthorized transmissions are dif-

[25] On the risks of leaks see Pozen (2013); Sagar (2016).
[26] See Geser (1992); Gibson (2014).

ficult to avoid,[27] yet organizations may become adept dealing with the threat of leaks. For example, one secretariat staff member that we interviewed described the development of a parallel, secure computer system for the storage of sensitive technical details. Finally, it is important to note that organizational leaks do not represent an all-or-nothing problem. One interviewee candidly admitted that member-states expect a few states who are known for cyber espionage access some systems at his IO. Yet this did not render the system useless, as such a system still prevented the vast majority of states and non-state actors from doing so.

International organizations as reliable and leak-proof storehouses for information may seem implausible to some readers. One broad finding of the book, however, is that among the institutions we analyze, sensitive information is rarely leaked and confidentiality protections are perceived as broadly reliable. Our interviews with practitioners suggest several reasons for this. First, secretariats at IOs often develop organizational cultures that informally and formally reward information security. While others have noted this feature of central bank cultures,[28] our interviewees suggested that it appears in other kinds of organizations as well. Second, sensitive information is typically not disclosed broadly within an IO. Rather, states often target a small, vetted list of specific staff. Third, IOs with narrow functions and smaller bureaucracies tend to handle information more securely. Submitting information to a specific office at a functional IO – for example, the dispute settlement staff at the WTO – is qualitatively less risky than disclosing it to a sprawling United Nations bureaucracy.[29] Finally, perfect secrecy is not required to ease disclosure dilemmas. Instead, states share their information when the incrimination benefits exceed the adaptation costs. Thus, if IOs reduce – but do not eliminate – adaptation costs, states will be more likely to provide their

[27] sources

[28] E.g. Stiglitz 2003, 121.

[29] As Chesterman (2006*a*, 151) notes, "most diplomats in New York assume that their communications are routinely intercepted by the US and other intelligence services."

information. While states may continue to withhold especially sensitive details, IOs may thus improve the amount of information available.

2.4.2 Vetting Sensitive Disclosures

In the second stage of 2.3, the IO assesses the validity of states' disclosures. As we noted previously, informed states typically have well-known political and strategic interests that create credibility problems, whereby claims of innocence or guilt tend to be partially or fully discounted, especially when details are not included. In contrast, many IOs lack the narrow political and strategic interests of states, allowing them to assess claims based on confidentially shared information and to credibly communicate their conclusions.[30] Thus, the IO's technical experts must be able to scrutinize the details that are shared with them confidentially, compare them with other information sources, and integrate them into their assessment of a given compliance case.

Equally important, states must believe that they possess technical expertise and are relatively unbiased. Previous scholars have noted the importance of IOs' neutrality when the IO is tasked with providing policy advice.[31] IOs must carefully manage how their use of confidentially disclosed material is perceived to avoid allegations of unfairness or bias.[32] A record of confirming information from multiple sources and rejecting faulty information can help to reassure states that it does not take allegations at face value.

International organizations build and maintain a reputation for neutrality in several ways. One is through technical expertise and cultivating the belief that IO assessments and de-

[30]For an application to policy proposals see Thompson Iraq.

[31]On IOs' ability to validate policy proposals, see Voeten (2005); Thompson (2006*a*); Chapman (2007). In other contexts, IOs aid states by being biased or having incentives to misrepresent information. See, for example, IOs as "biased experts" (Krishna and Morgan 2001), "advocates," (Dewatripont and Tirole 1999), or "mediators" with conflicting incentives over whether to report violations of agreements (Kydd 2006; Mattes and Savun 2010).

[32]Some also argue that IOs must have access to "an exogenous source of information" in order to be believed (Fey and Ramsay 2010).

cisions are driven by technical rather than political judgment. This can include careful attention to the national identity of secretariat staff. Often, IOs have an incentive to protect this image because they have incentives to retain the trust of members in order to do their jobs effectively. As we detail in later chapters, a variety of IOs such as the IAEA and the WTO go to great pains to maintain perceptions of neutrality. Other factors that influence perceptions of neutrality are less operational. For example, the location of the tribunal for war crimes in Rwanda was chosen by member-states to avoid.[33] Thompson (2006*a*), moreover, notes that the composition and voting rules of an institution like the United Nations Security Council helps produce a perception of relative political neutrality. Economic IOs moreover have incentives to protect their reputations for political neutrality and unbiased judgments, just as they do in the security realm, as discussed previously.

The details of the vetting process vary by institution. All IOs will draw on technical experts to assess disclosures with particular expertise that can include legal, scientific, economics, engineering, and environmental. Vetting sensitive disclosures can require the IO add staff with particular expertise. For example, the World Trade Organization's integration of highly sensitive business information required allocating experts in economics and firm behavior that were not previously needed. Similarly, the International Atomic Energy Agency has added experts in overhead imagery analysis as state- and commercial-based satellite imagery has become more important. In-house experts can compare disclosures to other information sources, including sensitive and non-sensitive submissions by other states.

The other powers of an international institution can also affect the vetting process. An IO with delegated monitoring authority, for example, may have assigned experts use a sensitive disclosure to inform new monitoring missions. The new information gathered from these confidential "tips" will then be used to assess the disclosure's accuracy, and can be shared with the broader international community. Those IOs that lack independent information-

[33]See Zarek notes.

gathering powers are therefore dependent on comparisons to voluntarily submitted information, whether from states or outside non-governmental organizations. For IOs with authority to render judicial decisions, judges and their technical staff can integrate sensitive disclosures to reach a final judgment about compliance. For IOs with more modest decision powers, member-states must independently or jointly assess compliance.

2.4.3 Sharing Findings

In the third stage in Figure 2.3, the IO shares its conclusions and any supplemental information with the wider body of member-states and outside non-state groups. For example, an IO may issue a public-facing report that summarizes conclusions informed by confidential data but lacking details submitted on a confidential basis. Alternatively, the IO may generate two versions of a report. One will include the raw sensitive details and only be open to a limited group of staff. A second will be shared broadly but have sensitive details redacted. In either example, an IO with monitoring powers might also include new information the IO gathered based on confidentially shared tips, filling in gaps in the evidence and obviating the need to share the original, sensitive details.

Overall, any IO which successfully receives, vets, and shares findings from sensitive disclosures presents informed states with a third option, beyond going public or staying quiet. Figure 2.4 combines these ideas to illustrate the logic of disclosure decisions from the perspective of the informed state. A state with compliance-related information first determines whether revealing its information provides an incrimination-related benefit. If it does not, there is no reason to consider disclosing it. If so, it next assesses whether wide dissemination of these details is costly. If not, it can release the information widely without resorting to an IO. If its information is sensitive, the state next considers whether an IO exists that can evaluate its information credibly while safeguarding its sensitive details. If an IO cannot do so, most of the time the state withholds its information; if the IO can protect it, the state

discloses the information with the IO.

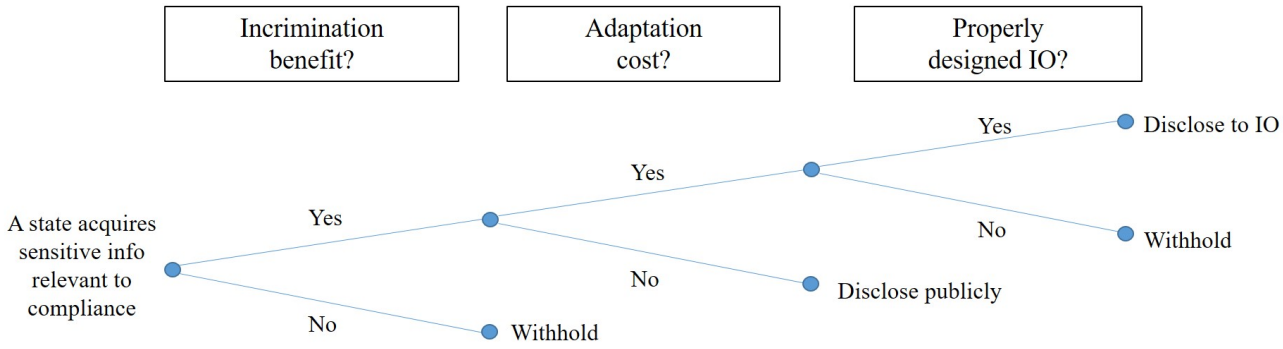

Figure 2.4: Decision Tree for States with Sensitive Information

2.4.4 Impact on Cooperation Goal

What is the significance of this third option? If an IO is equipped to securely receive and vet sensitive information, what is the practical effect? As noted above, an unresolved disclosure dilemma typically results in some unique forms of compliance-related information being kept private. Such details may be trivial, but our later empirical chapters collectively make the case that is often significant. Being denied insights based on sensitive information can leave the wider international community – both states and non-state actors – less confident that compliance with formal and informal rule is being adequately tracked. The end result can be shallower cooperation and greater risk of violations going unpunished. In practical terms, this can be quite harmful. War criminals will have greater confidence they can act with impunity. Hosts of foreign direct investment will have less fear of market punishment of expropriation. Peacekeeping missions will be less confident they can detect, address, and deter ceasefire violations.

A properly equipped IO can reverse these dynamics. If an IO can receive sensitive disclosures and vet them, the broader international community benefits from better insights into compliance issues. While secrecy reduces the adaptation costs that a disclosing state

faces, an IO can infuse such claims with greater credibility, screen out inaccurate claims, and improve wider trust that violators are being identified and punished. This, in turn, opens the door to deeper cooperation. As our later empirical chapter show, integration of sensitive information has concrete implications for important policy and normative goals. A tribunal that can draw on intelligence submissions, we argue, is more successful in securing indictments and arrests of war criminals. Similarly, integration of sensitive business information into the trade dispute system makes it harder to pull off non-tariff barriers via subsidies. Our theory highlights two institutional features: the capacity to keep sensitive details secret and the ability to credibly validate disclosures. Paradoxically, the institutional solution we develop uses *secrecy* – that is, intentionally limited information access – to solve a problem of information *underprovision*.

2.4.5 When Do IOs Have Confidentiality Systems?

Given its potential benefits, readers may naturally wonder which IOs have confidentiality systems and when they obtain them. When does an IO adopt a confidentiality system if it originally was not equipped with one? Put differently, if states can better identify noncompliance by giving IOs this function, then why would IOs ever lack such powers?

To be clear, our goal is not to explain the initial design or reform of IOs. As we describe below, an institution's features are a central *independent* variable in our framework. We use it and political incentives to disclose to understand how states use information and its impact on cooperative goals. Yet other scholars have given the issue of when institutions are endowed with cooperation-enhancing functions considerable attention. Two basic viewpoints exist. At one end of the spectrum is a smooth functionalist view, typified by the "rational design" project: institutions acquire authorities like a confidentiality system when the cooperation problem states seek to address demands it. Others take a far less sanguine view, noting that institutional design is sticky, cooperation problems evolve, and reinforcing feedback

loops make change difficult.[34] Change is not impossible but often requires exogenous shocks that generate critical historical junctures in which stakeholders re-evaluate and reconfigure institutions to better suit their needs.

In the empirical chapters that follow, the bulk of our descriptive findings supports the latter view. States and IO secretariats tend to prefer the status quo to politically controversial reforms; moreover, such reforms may be expensive and some countries or firms may see the use of sensitive information as threatening or generally oppose the IO. We use a before/after design in our chapters, which dramatically demonstrates that disclosure dilemmas exist long before institutions for trade, nuclear proliferation, and war crimes adopt confidentiality systems that address them. While not the focus of our empirical analysis, our chapters include evidence that unexpected, politically signifciant "shocks" can create junctures in which reforms are possible. For example, we argue that the end of the Cold War and the polarization of the international system under bipolarity created the political space for reforms at the International Atomic Energy Agency and for tribunals for war crimes. Moreover, stickiness and reforms can both directions. In Chapter X, we show that leaks and a legitimacy crisis led to a weakening of the confidentiality system for an IO used to arbitrate foreign direct investment disputes.

2.5 Discretion and Selective Disclosure

Thus far, our theorization of the politics of sensitive disclosures has focused on an institutional solution. We argue that a properly equipped institution can ease the tension between adaptation costs and incrimination benefits, giving informed states a third option of narrow disclosure to an IO. However, even if an IO provides a third *option*, will an informed state always take advantage of that option?

[34] E.g. Pierson 2011.

Our theory explicitly incorporates this issue, answering the question with "no." In practice, states retain significant discretion over whether to disclose and exercise it even with an institutional third option. This is partly due to the anarchic nature of the international system. States retain sovereignty over their information even when an IO offers to protect what they disclose. Discretion is also unavoidable for practical reasons. The private insights of an informed states are, by definition, unobservable unless disclosed. Thus, IOs and other states usually will not know about the insights that informed states choose to withhold. For example, outsiders may observe that France provides intelligence which implicates a top Serbian leader in a war crimes trial, but will they know whether France withheld other intelligence about other leaders? Thus, an institutional solution does not eliminate the reality that sensitive information will be disclosed *selectively* even with an institutional alternative.

We focus on who information incriminates to capture this second layer of the theoretical story. We assume that informed states seek to protect their friends and themselves from information that harms them in questions of compliance. Informed states, in contrast, have especially strong incentives to make use of compliance insights that bear on the non-compliance of rival states. Sharing here helps build a political or legal case against a rival even as it risks adaptation costs. In short, we argue that both institutional design and strategic interests matter.

The concept of incrimination benefits, discussed above, captures this intuition. The informed state's selfish strategic interests shape whether there is an incentive to disclose in the first place. Informed states with information that implicates themselves or their political allies do not face a meaningful disclosure dilemma. This situation lacks a "pull" for making unique insights about compliance more widely known; the obvious course of action is to withhold sensitive information. Countries seldom seek to get themselves or their allies into trouble by revealing such misdeeds both because strengthening allies tends to boost a country's own security and because countries often possess greater leverage over allies.

Indeed, scholars have long recognized that security externalities exist whereby countries gain from strong allies and lose from breakdowns in their allies' security.[35] Moreover, countries are often able to persuade allies to alter their behavior through bilateral actions, since they tend to have more influence over such countries and stronger economic and security ties with which to bargain.[36] Similarly, in the economic domain, firms and their governments would have no reason to disseminate sensitive details that demonstrate the home government's own non-compliance.

We argue that an incrimination benefit must be present in order for the informed country to be said to face a dilemma. This is most obvious when the informed country has sensitive information about an adversary's failure to comply. If a country has incriminating evidence about a friend, sharing it could endanger is relationship and the security of a friend now at risk of reputational and substantive punishment. In economic contexts, an informed state can possess information which helps show its own trade or investment violations, or which implicate a trade partner in trade or investment violations. The former scenario lacks an incrimination benefit; the latter features one.

2.6 General Hypotheses

Two core empirical expectations flow from these claims. First, our theory outlines conditions under which states with sensitive information should be more or less likely to disclose it. When adaptation costs and incrimination-related benefits are relevant, governments should disclose sensitive details only if an IO has adopted a reliable confidentiality system and the information does not harm the informed state's strategic interests. Second, if sensitive information is successfully elicited from governments, our theory suggests that IOs should gain new insights into compliance-related questions and, as a result, international cooper-

[35] See Gowa (1989); Gowa and Mansfield (1993).
[36] See (Miller 2014*b*).

ative outcomes should improve. These broad expectations are the foundation for a set of issue-specific hypotheses we articulate and test in each subsequent empirical chapter. For example, Chapter X presents evidence that new confidentiality reforms at the World Trade Organization made commercially sensitive disclosures more common and deepened trade in economic sectors where sensitive information is especially common.

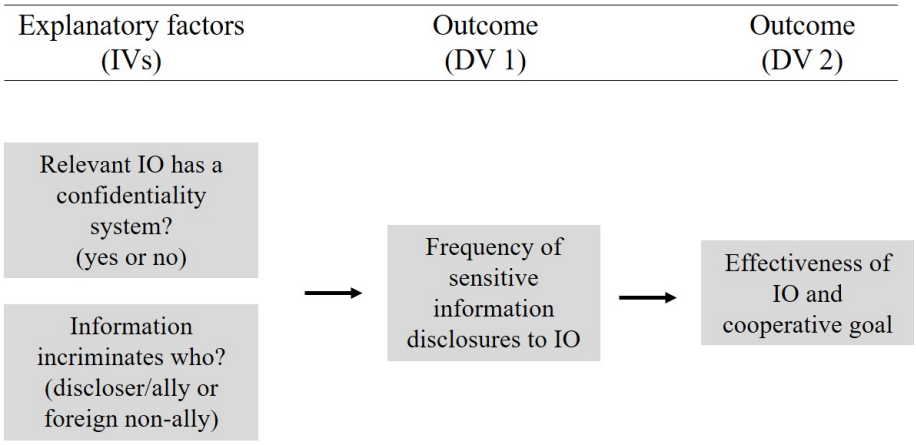

Figure 2.5: Research Questions

To formalize the intuition, we expect that in both the economic and security domains, IOs can reduce the harm of disclosing sensitive information by providing a third option between open revelation and staying silent. Lowering adaptation cost while still allowing disclosures to affect compliance questions should lead more states to supply it when they would get a benefit from doing so. The better an IO can perform this function, the more it can address disclosure dilemmas.

Hypothesis 1. *Suppose that countries will adapt to compliance information in ways that will harm the informed state. Then the better an IO can protect sensitive information, the more sensitive information informed states share with it, particularly about states or firms which constitute a security or economic rival.*

Moreover, IOs that can handle information confidentially should obtain more of it. Since

sensitive details are often the most precise material regarding compliance questions, member-states operating in the shadow of a properly equipped IO should know violations are more likely to be caught. This should dissuade violations of international rules and make makes members more likely to initiate and sustain cooperation.

Hypothesis 2. *Suppose that countries will adapt to compliance information in ways that will harm the informed state. Then the better an IO can protect sensitive information, the more success the IO will have in facilitating cooperation and the deeper cooperation will be overall, particularly regarding actors and activities that endanger the interests of the informed state.*

We adapt these general hypotheses to four areas of international relations: war crimes, nuclear proliferation, trade, and investment. The first two focus on intelligence, where the adaptation cost from wide dissemination by the informed state is the loss of sources and methods. In the third and fourth areas, the focus is firm-level data and documents, where the wide revelation of market sensitive information endangers a firm's market position. The table below summarizes these differences among the empirical chapters. In each , we collect new data on information disclosures to the relevant IO as well as new qualitative and quantitative material on improved cooperative outcomes. We elaborate on these empirical strategies in the following chapters.

Issue Area	Adaptation Cost	Incrimination Benefit	Relevant IO	Outcomes
War Crimes	Sources & Methods	Support Justice; Punish Regional Threat	ICTY/ICTR	Intel Disclosed; Arrests; Indictments
Nuclear	Sources & Methods	Punish Adversary	IAEA	Intel Disclosed; Facility Closures
Trade	Market Rivals	Win Dispute	WTO	Redactions; Trade
Investment	Market Rivals	Win Dispute	ICSID	Redactions; FDI

2.7 Three Remaining Questions

To this point, this chapter has laid out our theoretical story for the logic driving decisions by states that possess unique, sensitive information about compliance. Building on the concepts of adaptation costs and incrimination benefits, we argue that state decisions are based on the availability of an institution with confidentiality features and the informed state's strategic interests. Absent an option for limited disclosure to an IO, states tend to withhold sensitive information despite its utility for clarifying compliance questions. If an IO equipped to handle and integrate sensitive information exists, informed states will disclose more frequently, especially when sensitive information implicates rivals.

This section addresses three lingering questions raised by our story about IOs and confidentiality systems. Addressing them helps clarify why other functions for IOs are unable to ease disclosure dilemmas, why confidentiality systems are broadly useful for a range of IOs, and how our theoretical story both supports and is in tension with a functionalist view of institutions.

2.7.1 Why Not Delegate Sensitive Information Collection?

The first question raises the possibility that IOs might help states address disclosure dilemmas in other ways. In particular, past research into other informational and commitment problems suggests delegation to IOs is a common approach.[37]. Why not delegate the collection of sensitive information to a third party institution like an IO?

While global governance institutions are sometimes empowered to monitor and collect information, rarely does that encompass the kinds of sensitive details we theorize, such as firm-specific pricing data or clandestinely derived intelligence insights. This is rare for several reasons.

[37] delegation volume

First, governments – who must jointly agree to delegate such powers for an IO to acquire them – have understandable concerns about a significant loss of sovereignty. The mechanisms by which states obtain sensitive information are some of the core elements of sovereignty. For intelligence, countries are especially loath to sacrifice their informational sovereignty in the intelligence domain.[38]. Retaining technical and human source advantages over other states and non-state actors is usually seen as central to ensuring self-preservation in an anarchic setting.[39] On the economic side, states are able to demand that firms share under-the-hood details relevant to compliance because those firms are located in sovereign territory and subject to network of national laws. Because IOs lack territorial sovereignty and a network of property and economic regulations, their ability to force a firm like Boeing to provide sensitive business details is highly circumscribed.

Endowing an IO with the authority to replicate such activities would be a far larger concession than allowing IOs to collect pollution data or monitor ceasefire terms. General sovereignty concerns are compounded by the risk that an IO could turn those powers against those who originally delegated them. Many states legitimately suspect an intelligence capability could be put to use against themselves. Similarly, authority to collect sensitive firm data could be subverted for corrupt or personally enriching ends. States are reluctant to give such authority to an independent IO lest a fishing expedition against one of their own firms takes place in the future.

Finally, providing IOs with fact-finding capacities can be costly; member-states may balk at assuming the expense. Thus, most of the time, much or all of the sensitive information that IOs obtain must come from governments themselves.[40] Put differently, a central theme of the book is that access to sensitive information constitutes a form of power. Relinquishing

[38]See Walsh 2010.

[39]On information advantages and security seeking, see [].

[40]Some sensitive information also comes from non-state actors such as NGOs as well as open source intelligence. We discuss these types of information further in Chapter 2.

this to an IO would weaken states' control.

We hasten to add that there are some rare exceptions of IOs with authority to collect what can be taken as sensitive information. One notable example is the use of drone flights flown for UN peacekeeping. We discuss these examples in more depth in Chapter 8. For now, it is sufficient to note that these exceptions prove the rule and remain extremely controversial with "intelligence" remaining a dirty word in global governance.

2.7.2 Are Only Powerful IOs Capable of Addressing Disclosure Dilemmas?

A second question is about scope conditions. A reader might suspect sensitive information integration is only feasible for the most powerful IOs. An organization weak in other areas, the thinking might go, could not affect high stakes decisions to disclose. A related concern might be that confidentiality systems only work if an IO has specific features like monitoring.

We argue there are relatively modest scope conditions on our claims about how IOs can address disclosure dilemmas. Regarding overall power of the IO, the logic of our theory suggests IOs need a relatively modest resource base to perform the basic confidentiality functions. Like any organization exercising information security, an IO needs sufficient resources, policy, and infrastructure to protect sensitive disclosures. While important, this is misleading. If anything, smaller IOs will feature greater accountability and fewer risks of leakage. A second driver of costs is the vetting process. However, most IOs originate as bodies of technical experts. While some sensitive disclosures might require hiring specially trained staff, most of the reputational and practical resources should be already present. Similarly, sharing the results of this vetting can often draw on existing capacities developed for the traditional information dissemination function of IOs.

What about specific features like monitoring? In the sections above, we have been careful

to describe how the confidentiality process can work in institutions with different roles. In the empirical chapters that follow, we provide additional details for IOs with judicial functions compared to IOs with monitoring powers. In Chapter 8, we review other examples of IOs, like United Nations peacekeeping missions, which engage in on-the-ground operations. Even an IO with extremely limited powers – as a decision forum and home for technical experts – can still serve the vetting function that helps convert potentially biased claims into credible statements regarding compliance.

2.7.3 Is This a Functionalist Story?

Finally, readers might still wonder whether our theory embraces a purely functionalist view of institutions, in the spirit of earlier rational institutional design literature.[41] To be clear, our opening theoretical move uses problem/solution terminology and argues that an institutional solution exists for a unique challenge from sensitive information. Yet our theory includes an important twist: selective disclosure. We argue that states do not automatically submit sensitive information once the optimal institutional design is reached. Figure X and the core hypotheses in Section 2.6 underscore that our expectations about the relevant outcomes (e.g. disclosure patterns; cooperation success) depends on institutions and states' strategic interests. Moreover, our empirical chapters focus analyze periods *before* confidentiality reforms were adopted when disclosure dilemmas remained unaddressed. In short, we embrace a hybrid view. Institutions can solve informational problems for states, but *how* they acquire this capacity and *whether* states take advantage of it depends on non-functional variables.

A related point is worth clarifying. Does easing disclosure dilemmas have an egalitarian effects (e.g. "Pareto optimal") or does it help some states or other actors more than others? We do not argue that all states benefit equally from confidentiality systems. Sensitive infor-

[41] Koremenos et al.

mation and the capacity to use it are not evenly distributed. Intelligence, in particular, is predominantly in the hands of powerful states or regional powers that invest in intelligence bureaucracies and technologies. As we discuss in the next section, this endows states with a subtle form of power.

2.8 The Downstream Consequences for IOs

Our central claim is that confidentiality systems enable states to share sensitive details relevant to compliance with a lower risk of facing harm from adaptation costs. While states retain discretion about whether to disclose, an institutional alternative to the publicize/withhold dichotomy will tend to increase the occasions in which trade disputes and war crimes are informed by firm-specific details or national intelligence. Yet adding new institutional functions, such as a confidentiality system, inevitably causes other – often unintended – effects. We briefly review three of these downstream consequences here and return to these issues again in Chapter 8.

One downstream effect that can follow if IOs serve a confidentiality function is creation of a new tactic of power. Informed states can use sensitive information disclosure to influence the information landscape. When they choose to disclose sensitive details to an IO about a particular episode, this can improve scrutiny of individuals and states. In contrast, withholding sensitive details from a properly equipped IO can serve to obscure a particular rule violation and generally obstruct scrutiny of the government or firm in question. In short, integrating sensitive information improves cooperation but also opens up a new channel for a subtle form of influence. This influence-via-information effect is akin to what past scholars have deemed the "second face of power," in which an actor influences a target's choice by altering the agenda and choice set rather than by directly coercing or bribing it.[42]

[42]See Carson and Thompson (2014); the original second face of power conceptualization is in Bachrach and Baratz (1962).

Second, confidentiality systems may impact the accountability and transparency of an international institution. Our theory posits that a necessary condition for integrating sensitive information is organizational secrecy. That is, IOs must develop routines, procedures, and policies to limit access to sensitive details. This takes the form of a proto-secrecy bureaucracy: documents are labeled according to a classification system, penalties for unauthorized disclosures by staff are created, etc. Moreover, the vetting process requires IO staff and leadership to validate claims without having all of the details available to member-states. In both appearance and practical operation, some aspects of an IO's operations will be necessarily opaque. This, in turn, increases the risk of corruption and the appearance of unaccountable judgments. This can sharpen a general normative tension with democratic, participatory global governance, a theme we develop at length in Chapter 8.

Lastly, confidentiality systems and sensitive information integration may also have downstream effects on the autonomy of the IO. There are good reasons to believe that IO autonomy is both helped and hurt by this function, corresponding to the power and accountability effects just described. On the one hand, autonomy is reduced if a select state or group of states only shares sensitive details about some subset of compliance issues. An IO secretariat may find itself amply supplied about some episodes and starved in others. On the other hand, the confidentiality process imbues IOs with a qualitatively new role in analyzing compliance issues. Recall that vetting requires states make a leap of faith, trusting that an IO with access to some details they lack will exercise fair and sound judgment. An IO only drawing on non-sensitive information, in contrast, can have its books checked more thoroughly. We therefore suspect IO autonomy may be both enhanced and degraded.

Given the importance of institutional unbiasedness in our theoretical story, it is worth noting whether downstream effects like these endanger that very reputation for neutrality. This risk is ameliorated in a few ways. First, our theory focuses on an IO's reputation for fairly analyzing information it receives. This is distinct from the perception that it

receives a balanced, diverse supply of information. Even an IO with selective submissions can use clear procedures and technical expertise to show its assessments are fair. Second, observability is an important when thinking about perceptions and reputation. In each of our empirical chapters that follow, there is significant uncertainty about the details of any sensitive disclosures, i.e. by whom and about what, and what sensitive details are *not* disclosed. This means state and non-state views of the IO's credibility are may be insulated from operational changes. Finally, IOs that value their reputation for non-biased, technical expertise can correct course if sensitive information poses a threat. Later chapters include episodes where IOs reject sensitive information disclosures as non-credible and clarify procedures for evaluating disclosures as a way to safeguard their reputation. We discuss these points further in Chapter 8.

2.9 Conclusion

TBD!