International Cooperation, Information Transmission, and Delegation*

Emiel Awad† Nicolás Riquelme‡

July 10, 2023

Abstract

Do international organizations (IOs) help states to solve coordination problems over policy choices? We analyze a formal model of coordinated adaptation in which states use costly signals to transmit information about their preferences. We show that states only delegate to IOs if states are sufficiently aligned and face little uncertainty about each other's preferences. Although states gain from delegation by achieving more policy coordination, they also incur more costs because of inefficient signaling. States misrepresent their preferences to ensure policies are coordinated on their own preferred outcome, and delegation to IOs makes states want to misrepresent their preferences more strongly. This effect can be so strong that the gains from international coordination are insufficient to warrant delegation to IOs. We discuss the robustness of our results to different types of IOs and provide implications for the design of institutions.

*We thank Jason Davis, Zuheir Desai, Torun Dewan, Stefan Elbl, Mark Fey, Jacque Gao, Michael Gibilisco, Gleason Judd, Rafael Hortala-Vallve, Brenton Kenkel, Vanand Meliksetian, Jack Paine, Kris Ramsay, Patricio Romero, Bradley Smith, Duncan Snidal, Randy Stone, Stephane Wolton, several anonymous reviewers, the editors at the Quarterly Journal of Political Science, and audiences at the University of Rochester, London School of Economics and Political Science, Quebec Political Economy Webinar Series, and the conference of Political Economy of International Organizations 2022 for helpful comments. Riquelme gratefully acknowledges financial support from ANID/CONICYT (FONDECYT Iniciación #11220584). A previous version of this paper had the title 'International Coordination and the Informational Rationale for Delegation.'

†Associate Research Scholar, Department of Politics, Princeton University, emielawad@gmail.com.

‡Assistant Professor, School of Business and Economics, Universidad de los Andes, Chile, nriquelme@uandes.cl.

States face various international coordination problems that they could either solve *ad hoc* or via formal international organizations (IOs). For example, they may want to coordinate the regulation of technology companies or of transport technologies. Why do some states delegate authority to IOs to solve coordination problems while others do not? A sizable international relations literature identifies several reasons for delegation to IOs (Voeten, 2019). IOs may help solve international collective action problems, allow for credible commitments, and reduce transaction costs in policy-making (Abbott and Snidal, 1998; Hawkins et al., 2006; Bradley and Kelley, 2008; Koremenos, 2008; Hooghe and Marks, 2015). Also, IOs can have specific expertise that states do not have, or help states reduce uncertainty by pooling their information (Koremenos, Lipson and Snidal, 2001; Koremenos, 2008).

If states value international policy coordination and IOs can help with this, why would states *not* delegate authority to IOs? Our answer is that states may face too much uncertainty about other states' preferences. Even if states can create an IO that optimally coordinates policies, states waste too many resources to influence the IO's decisions under too much uncertainty. The reason is that states want to misrepresent their preferences. By misrepresenting preferences, state A induces state B's policy to shift toward state A's preferred outcome. State A's misrepresentation of preferences lets B believe that A plans to pick a policy that is farther removed from B's ideal policy. This incentivizes B to 'acquiesce' and move its policy closer to A's policy to maintain sufficient coordination. Although state A has incentives to misrepresent both with and without IOs, this incentive is stronger when IOs actively lead to more international coordination. Thus, even though states gain from international coordination, their incentives to misrepresent may be so strong that IOs are not created. The gains from coordination may be insufficient to compensate states for the losses they incur due to their incentives to misrepresent.

We develop a formal model in which two states transmit information and make policies in the presence and absence of an IO. States have private information about which policies are domestically optimal. Their payoffs are determined by domestic policy adaptation and international coordination. On one hand, states want to choose policies that are optimal for their own economy, while on the other, they want to coordinate with other states. Our informational setting differs from most related models of IOs, as the private information of states is not *commonly* valued. Instead, states are privately informed about their *preferences*, but one state still cares about another state's private information indirectly to have correct expectations about what policy the other state plans to implement. Especially when coordination is relatively important, states want to make sure that information about their preferences and policy-making intentions are transmitted. Besides transmitting this information in written or spoken form (cheap talk), states may engage in costly activities (money burning) as a signal about their policy preferences.

Equilibrium policies depend on whether states make policies absent from IOs or delegate authority to IOs. Without IOs, two states make policies absent external influence. If there is an IO, then this organization is a separate actor which makes policies on the states' behalf. We downplay the exogenous costs and benefits of information transmission and coordination in IOs. That is, our results do not rely on the fact that IOs are costly to create or that states cannot transmit information without IOs. Instead, we are interested in the *strategic* effects of institutions on international cooperation and information transmission.

Our model produces two sets of results, relating to the effect of IOs on coordination and signaling. First, on the plus side, international delegation always helps with international cooperation. In isolation, states make policies without internalizing the effects on other states. They coordinate too little compared to what they could have achieved if they could keep their promises. The power of IOs allows countries to lock in more coordination than they could achieve individually and they benefit from this. The more capable IOs are in imposing costs on states for deviating from agreements, the easier it is to achieve more beneficial cooperative agreements. The fact that IOs are created to choose policies that take both states' welfare into account implies that both states benefit from delegation via improved international coordination.

The second result is this article's main novelty and contribution. International delegation and the involvement of IOs in policy-making lead to more costly signaling. We show that international institutions do not help states share more information. States fully transmit their information regardless of the existence of an IO. However, the cost of information transmission is always greater when policy-making occurs in the presence of international institutions. One reason is that following delegation, states have to incur costs to influence a policy that they would have made themselves absent delegation. Even if the IO is less biased than the other state, the negative effect of the IO's increase in authority always outweighs the positive effect of transmitting information to a less biased receiver. The other reason is that the IO cares more about coordination than domestic adaptation than each state in isolation. Hence, IOs more strongly adapt policies to states' claims about their domestic circumstances as long as these policies are coordinated. This leads to stronger incentives to misrepresent and increases the costliness of information transmission. State A wants to misrepresent its preferences to ensure that state B's policy is more coordinated to state A's policy. Although this incentive exists without and with IOs, it is stronger with IOs.

Taken together, these two results imply that states do not automatically benefit from IOs if they want to cooperate under uncertainty. Even if international coordination is a crucial objective for states, too strong incentives to send costly signals in IOs make states worse off delegating. The problem of increased costly signaling is more pronounced with greater uncertainty, combined with disagreement among states. Thus, delegation is only beneficial for states if there is not too much uncertainty. This ensures that the benefits from increased coordination outweigh the increased costs of signaling.

We extend our model to better understand when and why states delegate to IOs.[1] The first

[1] The online appendix contains additional extensions, including formal analyses of the implications of preference alignment and asymmetric uncertainty. Two extensions tackle the issue of institutional design in

extension considers partially delegated authority to the IO. The two states can deviate from the IO's recommendations at a cost. When the IO has insufficient enforcement capabilities to punish states' deviations, it cannot achieve the optimal level of coordination. This is because the IO must make it sufficiently unattractive to prevent deviations. When the cost of violating agreements is low, IOs have no effect, which means that decisions and signals are equivalent to a situation without an IO. The level of necessary enforcement capabilities is especially binding when two states have highly divergent preferences.

The second main extension conceptualizes the IO as an organization in which two states bargain over policies. As in standard bargaining models, the underlying assumption is that states cannot deviate from agreements if a proposal is accepted, i.e., contracts are binding. The main result is that if states have equal bargaining power, decision-making and signaling strategies are equivalent to a world in which the IO was conceptualized as a separate actor. This suggests that in our context of coordinated adaptation, it is irrelevant whether IOs are separate agents or member-led institutions. The bargaining process ensures that states choose decisions like the IO would have done as an agent.

The applicability of our model is subject to several scope conditions. First, IOs must influence policy-making through their enforcement capabilities. Our model does not apply when IOs only serve as a forum in which states communicate without any sort of enforcement mechanism. Our results are more applicable in IOs such as the European Union, which is relatively capable to impose restrictions and costs on its member states and punish non-compliance. Second, we study situations in which states have an informational advantage over the IO (Stone, 2009, 35–36). This setting is especially relevant for, e.g., newly formed IOs, IOs that operate with limited budgets, or if states have access to private information of domestic firms and markets that IOs cannot obtain or easily verify. Third, we study policy domains in which states at least partially want to coordinate policies. This is important be-

more detail, including the selection of the IO's preferences and the effects of limitations on its discretion. We postpone a brief discussion about these extensions to the end of the results section.

cause it generates incentives to misrepresent information; there clearly are no informational costs or benefits of IOs if states always prefer to tell the truth regardless.

An example that partially fits our model comes from the European Union. EU member states recently agreed to climate neutrality by 2050.[2] One aspect of tackling climate change is a collective action problem in which multiple countries need to cooperate and exert effort in regulating their economies. Another aspect, however, is an international coordination problem.[3] At the time of the negotiations, it was a contentious issue whether nuclear power could be part of a member state's strategy to reduce greenhouse gas emissions. The European Commission offered two possible policy options to its member states, where nuclear power would either get a *green label* or not. Although each state may individually benefit from using nuclear power, it would benefit more if other states used the same energy sources. This is because it could spur technological advancement and because scale economies could yield cheaper energy production. That is, France has incentives to push for a green label to help its domestic nuclear energy producers, and there are potentially further benefits if other states coordinate on the same technology. States such as Germany are opposed to nuclear energy, referring to it as a dangerous energy source, and plan to use hydrogen energy sources, among others. This case highlights how states may have different preferences based on their domestic economy, while still valuing coordination on specific policies.

Another component of this example is incomplete information about states' preferences.[4] Broadly speaking, there are two ways to transmit information to influence another party. One is direct communication (i.e., what do states *say*?), while another is indirect signaling (what do states *do*?). A way to model direct communication is cheap talk, where representatives

[2] `https://ec.europa.eu/clima/policies/strategies/2050_en` (accessed on April 20, 2021).

[3] For example, United States Treasury Secretary Janet Yallen urged the European Union to help improve international coordination on carbon-cutting policies. See `https://www.reuters.com/business/sustainable-business/g20-countries-can-take-several-paths-cut-carbon-emissions-yellen-says-2021-07-09/` (accessed on May 23, 2023).

[4] Toulemonde (2013) describes that states may have private information about their adaptation costs in setting standards.

of states may send letters to each other or to international organizations.[5] Another way to model indirect signaling is money burning. This is a reduced form of various activities that are costly and informative about a state's preferences. For example, French investments in nuclear power plants and research and development are not just productive in their own right. Other countries learn about French preferences in favor of nuclear power this way too. An interpretation of money burning stems from an example of TV standards in which European states attempted to influence their neighbors by assistance in marketing surveys, subsidies, exhibitions, fairs, and demonstrations (Crane, 1978).[6] Another interpretation is the intentional delay of international agreements. States may delay acceptance of a deal, which is costly and informative about their preferences (Rubinstein, 1985; Abreu and Gul, 2000), and thus delay can be a form of money burning. Finally, policies and investments that are made domestically (or the lack thereof) can also be a signal. Some claim that in tackling climate change, the United States' lack of leadership domestically hurts its ability to persuade other countries to cooperate.[7] Whatever these activities are, part of their use is a *signal*. The signal's *costliness* is the key component that we model below.

Related Literature

Our paper builds on the literature in international relations and organizational economics. It specifically builds on the rational design literature (Koremenos, Lipson and Snidal, 2001). A substantial literature studies IOs that have an informational advantage over states (Johns, 2007; Chapman, 2007; Fang, 2008; Fang and Stone, 2012; Crombez, Huysmans and Van Ges-

[5] Several states mentioned their support for nuclear energy in a letter about EU climate and energy policy addressed to the European Commission. This was a joint letter from the Czech Republic, French Republic, Hungary, Republic of Poland, Romania, the Slovak Republic, and the Republic of Slovenia (March 19, 2021). Available at `https://www.euractiv.com/wp-content/uploads/sites/2/2021/03/Nuclear-letter-march-2021.pdf` (accessed on April 20, 2021).

[6] Angulo, Calzada and Estruch (2011) discuss similar signaling behavior in Latin American countries that wanted to influence their neighbors to adopt similar digital television standards. Although firms are the main affected parties, governments also care because of concerns for their domestic economies and tax revenue.

[7] As the director of the International Centre for Climate Change and Development in Bangladesh mentioned: "John Kerry goes around the world saying all the right things, but he can't make the U.S. deliver them (...) He loses credibility when he comes and preaches to everyone else." See `https://www.nytimes.com/2022/07/01/climate/biden-climate-agenda-global.html` (accessed on May 23, 2023).

tel, 2017). Such articles are based on a vast principal-agent literature in which a principal may give an agent decision-making authority if the agent has more expertise and preferences are sufficiently aligned (Aghion and Tirole, 1997; Bendor, Glazer and Hammond, 2001). Another strand of the international relations literature studies the ability of IOs to help with informational issues if states are better informed. Keohane (1984) notes that "By reducing asymmetries of information through a process of upgrading the general level of available information, international regimes reduce uncertainty (p. 94)." Similarly, Thompson (2015) highlights the informational benefits of IOs in monitoring and exchanging information.[8] This literature is generally optimistic about how IOs can solve informational problems.[9]

Our article is also related to the literature on mediation in international conflict because of its focus on the role of institutions in promoting cooperation under uncertainty. The idea is that states' strength is private information which they can reveal to international institutions to help achieve peaceful outcomes. Fey and Ramsay (2010) show that if mediators do not have independent abilities to acquire information, they are unable to help solve states' incentives to misrepresent in favor of generating peaceful outcomes. Hörner, Morelli and Squintani (2015) and Meirowitz et al. (2019) study mediators as mechanisms that can acquire states' information and reveal it to the other state to reduce the probability of conflict.

Our results also build on the organizational economics literature (Gibbons, Roberts et al., 2013). Our setup is similar to the ones in Alonso, Dessein and Matouschek (2008) and Rantakari (2008). They study organizations with an uninformed headquarters (the principal) that values coordination adaptation and multiple divisions (the agents) that are better informed about local markets. Our main innovations are that we also consider *costly* signals and that we switch who are the principals (states) and the agent (IO).

[8] "Moreover, the exchange of information and discourse that takes place within an IO tends to reveal information about countries' preferences and intended actions [...], leading to more effective monitoring and higher quality signaling at the international level" (Thompson, 2015, p. 30).

[9] However, IOs do not necessarily reduce uncertainty. See Carson and Thompson (2014) and Carnegie and Carson (2019) for research on situations in which states are better informed. Carnegie and Carson (2019) find that IOs may help states share information on nuclear matters, but then withhold it from the public.

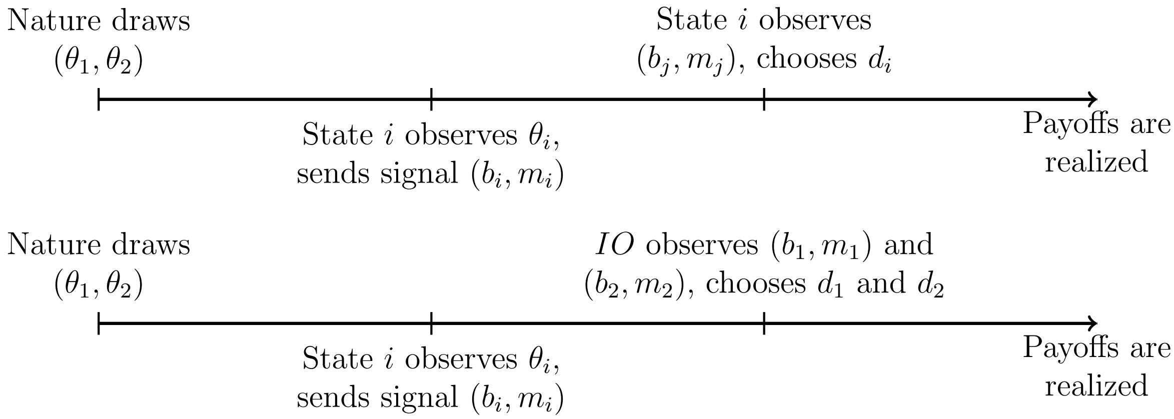
The Model

We consider two states that decide whether to create an IO or make policy *ad hoc*. Both states care about how their policy adapts to their domestic economy and how coordinated it is with the other state's policy. The two states are incompletely informed about the other state's preferences over policies. We compare and contrast two institutions that symbolize the absence and presence of an IO. In the first, states act without a formalized structure and make their own decisions. In the second, an IO makes decisions on the states' behalf. Our goal is to analyze the strategic implications of the creation of IOs and their impact on states' policies and information transmission.

The timing is as follows. First, Nature draws two random variables, each one being private information for each state, i.e., each state's type. These types determine a state's preferences over policy. The type of 1 (θ_1) is drawn from a uniform distribution with support $\Theta_1 = [\underline{\theta}_1, \overline{\theta}_1] = [-1-s, -1+s]$. Similarly, 2's type θ_2 is independently drawn from a uniform distribution with support $\Theta_2 = [\underline{\theta}_2, \overline{\theta}_2] = [1-s, 1+s]$. Note that state 1's type distribution has a mean of -1, while state 2's distribution has a mean of 1. The value $s \in [0, 2)$ indicates the amount of uncertainty about each state's private information.[10] The larger is s, the larger the interval from which a type is drawn, and the greater the amount of uncertainty about a state's type. In the second stage, each state $i \in \{1, 2\}$ observes its private information θ_i and sends a signal (b_i, m_i), where $b_i \geq 0$ is the amount of burned money and m_i is a cheap talk message. We allow states to transmit information using costly signals *and* communication to grant states flexibility in information transmission. This setup follows from an economic theory literature with cheap talk and burned money (Austen-Smith and Banks, 2000; Kartik, 2007; Karamychev and Visser, 2016). That is, related models of delegation which only allow for cheap talk are unable to speak about other types of states' signaling behavior in the international arena. Thus, money burning can be a way to model other behavior in a

[10]The results are similar when $s \geq 2$, but it complicates exposition; we discuss this in the online appendix.

Figure 1: Timing without and with Delegation



Note: The top (bottom) panel illustrates the no-delegation game (delegation game).

reduced form that can be informative to other states and IOs because of its cost.

The two institutions differ in the third stage. In the first model without an IO (the *no-delegation game*), each state makes a decision individually. In the second model with an IO (the *delegation game*), both states' decisions are delegated to the IO. In the third stage of the no-delegation game, each state i observes how much the other state has burned (b_j) and what cheap talk message was sent (m_j), and makes decision $d_i \in \mathbb{R}$. In the delegation game, authority is delegated to an agent, conceptualized as an IO, who makes both decisions on behalf of the states (the two principals). The IO observes amounts of burned money (b_1, b_2) and messages (m_1, m_2), and makes decisions $(d_1, d_2) \in \mathbb{R} \times \mathbb{R}$. Figure 1 illustrates the timing of both games. Each state's utility consists of a *policy payoff* and the cost of burned money, which can be written as $u_i(d_i, d_j, \theta_i, b_i) = \pi_i(d_i, d_j, \theta_i) - b_i$, where its policy payoff is

$$\pi_i(d_i, d_j, \theta_i) = -(1 - \beta)(d_i - \theta_i)^2 - \beta(d_i - d_j)^2.$$

This policy payoff consists of two parts. The first term is a state's adaptation motive, where $-(d_i - \theta_i)^2$ is the cost of decisions that are not adapted to domestic conditions. The further d_i is from a state's bliss point θ_i, the higher the cost that state i incurs. The second term measures coordination, where $-(d_i - d_j)^2$ is the cost of uncoordinated decisions. The parameter $\beta \in (0,1)$ measures the importance of coordination relative to tailoring decisions to domestic conditions. State 1 is best off if $d_1 = d_2 = \theta_1$ and both states coordinate on 1's bliss point θ_1. Symmetrically, state 2 is best off if both coordinate on 2's bliss point with $d_1 = d_2 = \theta_2$. Given delegation, we study an IO that treats states symmetrically. The IO weighs both states' policy payoff $\pi_1(\cdot)$ and $\pi_2(\cdot)$ equally, i.e.,

$$u_{IO}(d_1, d_2, \theta_1, \theta_2) = \frac{1}{2}\left[\pi_1(d_1, d_2, \theta_1) + \pi_2(d_1, d_2, \theta_2)\right].$$

Equilibrium

Equilibrium definitions depend on every actor's set of strategies. In the no-delegation game, state i's strategy is (i) a mapping from types θ_i to signals (b_i, m_i) and (ii) from types θ_i, state i's and j's signals to decisions d_i. In the delegation game, a strategy of state i consists of a mapping from types to signals; and for the IO it is a mapping from both states' signals to decisions d_1 and d_2.

We study perfect Bayesian equilibria (PBE), where players use sequentially rational strategies and update beliefs using Bayes' rule wherever possible. A strategy is sequentially rational if it is a best response to other players' strategies at any stage of the game. The requirement of Bayes' rule means that players' beliefs are consistent with other players' strategies. We fully specify our equilibrium concept in the appendix.

A general issue in signaling models with money burning and cheap talk is the existence of multiple equilibria.[11] Indeed, Lemma 0 in the Appendix uses existing results to characterize the infinite set of equilibria that differ in the degree of transmitted information.

We focus on the most informative equilibria in each game and justify this selection by applying a commonly used refinement in models with costly signaling.[12] From now on, we

[11] Karamychev and Visser (2016) study equilibria that are ex-ante sender-optimal. Proposition 1 shows that equilibrium existence requires a partitional structure, and that any arbitrary partitioning of the type-space can be generated through cheap talk and burned money. See also Kolotilin and Li (2021).

[12] Refinements may delete a large set of these equilibria, potentially guaranteeing a unique one (Cho and

call the most informative PBE an *equilibrium*.[13]

Lemma 1. *Every equilibrium is fully separating. It is without loss of generality to ignore cheap talk messages.*

Lemma 1 states that after observing how much money a state burned, every player perfectly infers the state's type on the equilibrium path. Thus, every equilibrium is outcome-equivalent regardless of the use of cheap talk messages. Therefore, we restrict attention to money-burning strategies in describing equilibria. In the appendix, we show that cheap talk is necessary to characterize equilibrium if the amount of uncertainty, s, is too large.

Results

The structure of our results section is as follows. First, we derive what decisions are made without and with an IO, and compare the amount of coordination and adaptation under each institution. This first step helps us to understand the benefits of delegation without taking into account the presence of uncertainty. Second, given decision-making strategies, we study how states burn money to transmit information in equilibrium and show how this differs across institutions. The fact that decision-making strategies and payoffs are characterized in the first step allows us to understand how incentives to misrepresent are shaped by the institution in which states operate. Third, we use the results from the first two steps to characterize our main results about when states delegate to IOs.

Improved Coordination

The presence and absence of an IO result in different decisions for each state. States make decisions optimizing their own policy payoffs. These decisions are driven by a state i's own type θ_i, its own signal b_i, and state j's signal b_j. The IO, however, makes decisions optimizing an equally weighted average of the policy payoffs of both states. Unlike both states, the IO's

Kreps, 1987; Chen, Kartik and Sobel, 2008). We adopt the monotonic D1 (mD1) refinement (Bernheim and Severinov, 2003), discussed further in the appendix.

[13] This equilibrium does not necessarily maximize efficiency. See Karamychev and Visser (2016) for a discussion on *ex ante* efficient equilibria. Although the receiver's preferred equilibrium is fully separating, the sender prefers to reveal information through a finite number of intervals.

decisions are only driven by the signals of both states. They are not directly determined by their true types as those are unobserved by the IO.

Crucially, because the IO considers the effects of decisions on both states' payoffs, these decisions are more coordinated. The IO internalizes the externalities from more coordinated decisions from state i for the payoff of state j. That is, the distance between d_1 and d_2 ($|d_2 - d_1|$) is smaller under delegation. States gain from more coordination but lose because of less adaptation. In sum, in terms of policy payoffs, they gain from delegation to IOs. Delegation to IOs allows states to achieve payoffs that they could have achieved if they had commitment power and could keep their promises. That is, IOs can grant states commitment power in making policy.

Proposition 1. *In equilibrium, the decision for state $i = 1, 2$ is as follows:*

$$\text{\textit{In the no-delegation game:}} \qquad d_i^{ND} = (1-\beta)\theta_i + \frac{\beta}{1+\beta}\mathbb{E}_i[\theta_j|b_j] + \frac{\beta^2}{1+\beta}\mathbb{E}_j[\theta_i|b_i],$$
$$\text{\textit{In the delegation game:}} \qquad d_i^{D} = \frac{1+\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_i|b_i] + \frac{2\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_j|b_j].$$

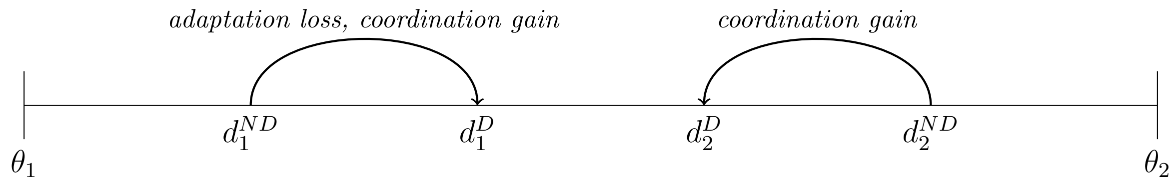
Delegation increases coordination and reduces adaptation. Without uncertainty, states always delegate to IOs.

Figure 2 displays the logic where the states' types are θ_1 and θ_2. Decision d_1^{ND} increases to d_1^D and becomes more moderate, while decision d_2^{ND} decreases to d_2^D and becomes more moderate too. In sum d_1^D and d_2^D are closer to each other than d_1^{ND} and d_2^{ND} are. Without considering the role of endogenous information transmission, IOs increase coordination and states benefit from the involvement of IOs. This implies that if there is no uncertainty, states always gain from delegating to IOs.

Deteriorated Signaling

How do IOs shape states' equilibrium signaling incentives? Proposition 1 characterizes equilibrium decision-making strategies and Lemma 1 shows that information is fully revealed in

Figure 2: Delegation and its Effects on Coordination and Adaptation



Note: The interval displays the policy-space in between the two states' types θ_1 and θ_2. Holding everything fixed except for the institutional arrangement, the IO leads to more coordinated policies. Each state loses in adaptation but gains in coordination, and achieves, in sum, a gain through delegation.

equilibrium via money burning. States' money-burning incentives are proportional to their incentives to misrepresent. Prior to presenting the results, we define two functions of θ_1 and θ_2 that determine money burning. For each function, the value of state i's type θ_i and its most moderate type matter ($\overline{\theta}_1$ for state 1 and $\underline{\theta}_2$ for state 2).

$$f_1(\theta_1) := \left(\theta_1 - \overline{\theta}_1\right)\left(\frac{\theta_1 + \overline{\theta}_1}{2} - 1\right),$$
$$f_2(\theta_2) := \left(\theta_2 - \underline{\theta}_2\right)\left(\frac{\theta_2 + \underline{\theta}_2}{2} + 1\right).$$

The function $f_1(\theta_1)$ increases when θ_1 moves *downward*, which means that state 1 burns more when its type is *lower*. Analogously, $f_2(\theta_2)$ increases when θ_2 moves *upward*, implying that state 2 burns more when its type is *higher*. Proposition 2 shows how money-burning strategies are determined in equilibrium and that IOs increase the amount of burned money.

Proposition 2. *In equilibrium, the money burning function for state $i = 1, 2$ is*

In the no-delegation game: $$b_i^{ND} = \frac{2(1-\beta)\beta^2}{(1+\beta)^2} f_i(\theta_i),$$

In the delegation game: $$b_i^{D} = \frac{2(1-\beta)\beta}{1+3\beta} f_i(\theta_i).$$

Delegation increases the amount of burned money for each type of each state.

To understand this result, we expand on the sources of a state's incentives to misrepresent.

A state's money burning incentives are proportional to incentives to misrepresent their type, claiming that it is θ_i' instead of $\theta_i \neq \theta_i'$. If misrepresenting is beneficial, state i must prevent itself from misrepresenting by burning money. It needs to exactly offset the profit of misrepresenting by an additional amount of burned money. For each pair of two types θ_i and θ_i' for which θ_i wants to mimic θ_i', the difference in the amount of burned money can be characterized as

$$\underbrace{\pi_i(d_i, d_j, \theta_i | \theta_i') - \pi_i(d_i, d_j, \theta_i | \theta_i)}_{\text{policy benefit}} = \underbrace{b_i(\theta_i') - b_i(\theta_i)}_{\text{misrepresenting cost}} . \tag{1}$$

That is, the left-hand side in Equation 1 measures the increase in a state's payoff if it tells a lie (θ_i' instead of θ_i). This must be equal to the increase in burned money, displayed on the right-hand side.

Given that states care about coordination, they have incentives to pull the other state's decision closer to its own. Specifically, state 1 wants to pull d_2 downward, while state 2 wants to pull d_1 upward. Thus, state 1 has incentives to misrepresent its information and claim its type is lower than θ_1. The reverse is true for state 2, as it wants to say θ_2 is higher than it actually is. Incentives to misrepresent information prevent states from fully transmitting their information through cheap talk. That is, if misrepresenting information is free, there is nothing that prevents them from doing so. Hence, influential information transmission requires money burning.

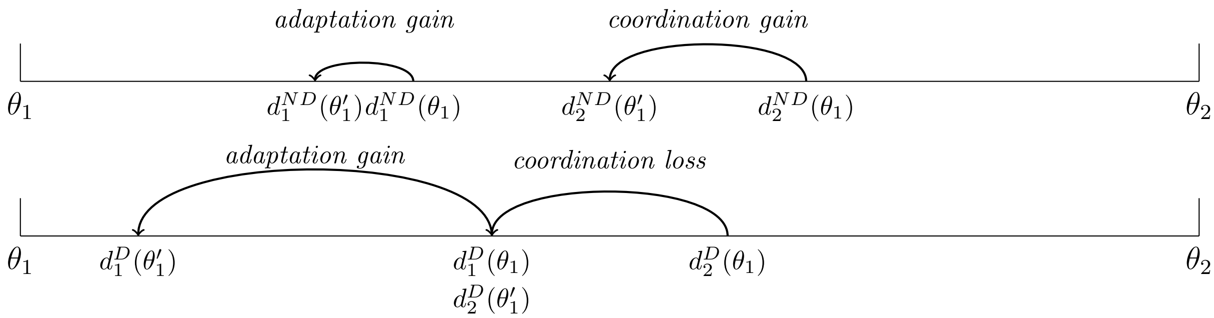
To understand how much money is burned in equilibrium, we characterize the magnitude of the benefits of misrepresenting information without and with delegation. Consider first the situation for state 2 without delegation. State 2 wants to say its type is higher than it actually is. By saying that state 2's type is θ_2' instead of $\theta_2 < \theta_2'$, state 1's decision changes, which ultimately also results in a change of state 2's decision. If state 2 says its type is higher (i.e., $\theta_2' > \theta_2$), state 1 expects that d_2' will be higher than state 2's decision d_2 if it would have been truthful about θ_2. However, state 2 will not move all the way to d_2' and instead only deviates a relatively small amount towards its ideal point. This means that, by misrepresenting preferences, both states' decisions will be relatively closer to each other. This generates a coordination gain from state 2's perspective. Misrepresenting information also changes the marginal benefit of adaptation relative to coordination. Given that state 1's decision will be more in line with state 2's decision, there is more coordination, making adaptation more valuable at the margin. As a result, state 2 can shift its decision more toward its ideal point θ_2. In sum, by misrepresenting preferences, state 2 gains both in terms of adaptation and coordination.

Now consider state 2's incentives to misrepresent if the IO makes decisions. These incentives may now differ because (i) the IO is a separate player, and (ii) the IO makes decisions differently. If state 2 misrepresents its type, then the IO believes this.[14] The IO shifts d_2 to the right if state 2 says θ_2' instead of θ_2. As a result, state 2 benefits from misrepresenting in terms of the adaptation motive. On the other hand, the IO also chooses d_1 differently because d_2 is now further apart. Hence, at the margin, coordination becomes more valuable, and the IO chooses d_1 which is then closer to d_2. In total, however, because the IO makes decisions under the impression that θ_1 and θ_2 are farther apart, there is less coordination. Therefore, if state 2 misrepresents its preferences, it loses with respect to coordination. In sum, however, state 2 gains from misrepresenting.

Figure 3 illustrates how misrepresenting information affects the magnitude of changes in equilibrium policies across institutions. A key takeaway from the figure is simply the magnitude of policy shifts. With delegation, the same misrepresentation of information has more significant effects on equilibrium decisions. This is not the full story, however, as the main driver of equilibrium money burning is not just the sensitivity of *decisions*, but the sensitivity of *payoffs*. Therefore, to formally understand the difference in signaling incentives without and with an IO, we define the two following functions for state 2. The expected payoffs given

[14]This is different than if state 2 would misrepresent its type because it would know the true θ_2.

Figure 3: Delegation and Stronger Incentives to Misrepresent



Note: We assume that $\beta = \frac{1}{2}$, $\theta_1 = -1$ and $\theta_2 = 1$. We look at the effects of type $\theta_1 = -1$ misrepresenting to $\theta_1' = -2$. The upper panel shows how misrepresenting information affects decisions and payoffs absent delegation to an IO, where state 1 gains in adaptation and coordination. The lower panel shows the same effect but when authority is delegated to the IO. It shows that a state gains from adaptation to a greater extent, and loses in terms of coordination as the distance between decisions increases.

no delegation (ND) and delegation (D) are

$$U_2^{ND} = -(1-\beta)(d_2^{ND} - \theta_2)^2 - \beta \left(\Delta d^{ND}\right)^2,$$
$$U_2^{D} = -(1-\beta)(d_2^{D} - \theta_2)^2 - \beta \left(\Delta d^{D}\right)^2,$$

where Δd^{ND} and Δd^D are the distances between decisions d_1 and d_2 under no delegation and delegation respectively. These terms measure the amount of coordination.

The above functions are relevant because they allow us to determine the benefit of misrepresenting at the margin. By taking the derivative of 2's expected utility function with respect to state 2's reported type, we characterize the marginal benefit of misrepresenting. To emphasize, the formal equations help to measure the benefits of misrepresenting preferences with respect to coordination and adaptation separately. First, without delegation, a misrepresentation of preferences has the following effects:

$$\frac{\partial U_2^{ND}}{\partial \theta_2'} = \underbrace{2\beta\Delta d^{ND}\left(\frac{\beta-\beta^2}{1+\beta}\right)}_{\textit{increasing} \text{ in coordination}} + \underbrace{2(1-\beta)(\theta_2 - d_2^{ND})\left(\frac{\beta^2}{1+\beta}\right)}_{\textit{increasing} \text{ in adaptation}}. \tag{2}$$

With delegation, however, coordination and adaptation are affected differently. The total marginal benefit of misrepresenting preferences equals

$$\frac{\partial U_2^D}{\partial \theta_2'} = \underbrace{-2\beta\Delta d^D\left(\frac{1-\beta}{1+3\beta}\right)}_{\textit{decreasing} \text{ in coordination}} + \underbrace{2(1-\beta)(\theta_2 - d_2^D)\left(\frac{1+\beta}{1+3\beta}\right)}_{\textit{increasing} \text{ in adaptation}}. \tag{3}$$

To compare the effect of delegation on the benefits of misrepresenting information, we need to compare equations 2 and 3. The latter is greater, which stems from the fact that in terms of the adaptation motive, a marginal misrepresentation of information affects decisions more strongly. This can be seen by comparing the second term in equations 2 and 3:

$$(\theta_2 - d_2^D)\left(\frac{1+\beta}{1+3\beta}\right) > (\theta_2 - d_2^{ND})\left(\frac{\beta^2}{1+\beta}\right). \tag{4}$$

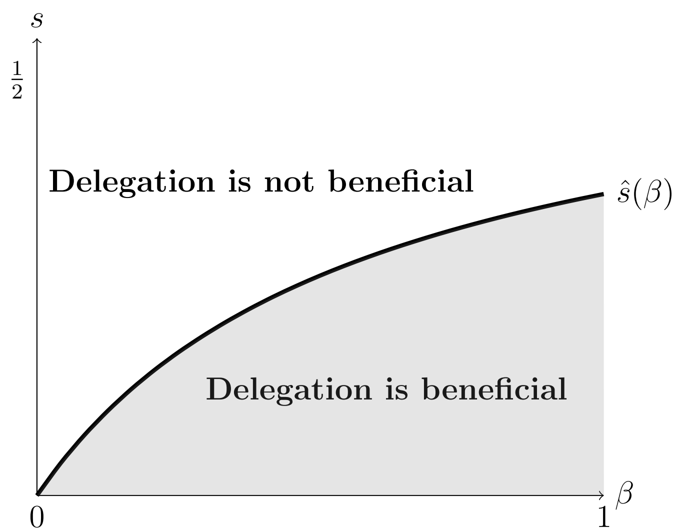
This effect dominates the fact that misrepresenting *hurts* a state under delegation in terms of coordination while it would be *beneficial* without delegation. In total, the benefits of misrepresenting under delegation are higher. Importantly, as well, the function makes clear that the most moderate types of each state ($\bar{\theta}_1$ and $\underline{\theta}_2$) are relevant in determining the amount of burned money. Thus, uncertainty works in a particular way: holding fixed a type θ_1, the more uncertainty there is in terms of s, the more type θ_1 has to burn to separate from the most moderate type $\bar{\theta}_1 = -1 + s$ and every other more moderate type.[15]

Endogenous Delegation

Given that states anticipate how costly their signals will be and how decisions are made, under what conditions do they prefer to delegate? For our first result, we assume that if both states benefit from delegation, an IO is created, while if neither state benefits, it is not created. Two key parameters are relevant in this calculus. On one hand, the value of coordination, β, influences decisions and money burning. On the other, the level of

[15]This is the only way in which uncertainty matters, as the money-burning functions do not depend on s or $Var(\theta_i)$ other than through the function's initial condition. Hence, it is not necessary to assume uniform distributions for these functions. The functions only depend on the distribution's expected value.

Figure 4: Coordination, Uncertainty, and the Value of Delegation



Note: The x-axis illustrates values of coordination, β, and the y-axis indicates the amount of uncertainty, s. The curve illustrates the function $\hat{s}(\beta)$ for which delegation and no delegation are equally good from the states' perspectives ex-ante.

uncertainty, s, affects money burning. Proposition 3 demonstrates the conditions under which states benefit from delegation in expectation.

Proposition 3. *After delegation, each state benefits from the increase in coordination but loses from the increase in burned money. States delegate if and only if the level of uncertainty is sufficiently low relative to the value of coordination.*

Proposition 3 formally establishes that states delegate if and only if the level of uncertainty s is sufficiently low relative to the importance of coordination. This means that, for a fixed value of β, there exists a value $\hat{s}(\beta)$ such that if $s < \hat{s}(\beta)$, states delegate, while if $s > \hat{s}(\beta)$, states do not create an IO. Figure 4 illustrates this graphically. There are two main takeaways. First, higher β makes states gain more from delegation for a fixed level of uncertainty s, and for a fixed β, more uncertainty makes states gain less from IOs. Thus, if states value coordination to a large extent, they benefit from delegation for a wider range of values s. For states to gain from delegation, it is necessary that they sufficiently value coordination but that there is not too much uncertainty. Second, if there is too much uncertainty, then states never find it beneficial to delegate, even if coordination is highly important ($\beta \approx 1$). This highlights the detrimental negative effect of delegation on signaling, especially when states face a very uncertain environment.

The following illustrates how each type's payoffs are determined when the delegation decision was made before a state learns its type. We say that a state's type θ_i is more *moderate* if it is closer to the expected value of θ_j given the prior ($\mathbb{E}_0[\theta_j]$) and more *extreme* otherwise.

Corollary 1. *A state's relative value of delegation decreases in how extreme its type is.*

A state's preference for delegation is therefore also determined by its type θ_i. We now consider the game where states individually decide whether to delegate authority to an international organization (IO) after they learn this type. An IO is created if and only if both states delegate. The game unfolds as follows after Nature draws each state's type:

1. Each state observes its type θ_i and independently decides whether to delegate authority.

2. If both states delegate authority, the IO is created. Otherwise, the IO is not created.

3. Each state observes the delegation choice made by the other state and burns b_i.

4. Finally, decisions are made as in the main model. If an IO is created, it makes decisions. If an IO is not created, then both states make decisions.

This game captures the decision-making process in which the creation of an IO is contingent on the joint delegation decision of both states. In solving this game, we focus on symmetric equilibria in cut-point strategies. That is, the symmetry assumption implies that type θ_1 chooses the same equilibrium strategy as type $\theta_2 = -\theta_1$. The cut-point structure of equilibria implies that (i) every type delegates, (ii) every type does not delegate, or (iii) there are two distinct intervals where types in one interval delegate, and types in the other do not delegate. The following proposition demonstrates the structure of equilibria.

Proposition 4. *Consider symmetric equilibria in cut-point strategies. There always exists* *an equilibrium in which every type does not delegate. Further,*

1. *If uncertainty is sufficiently small $s \in [0, \underline{s}(\beta)]$, then there exists an equilibrium in which every type delegates.*

2. *If uncertainty is intermediate $s \in (\underline{s}(\beta), \bar{s}(\beta))$, then there exists an equilibrium in which a set of moderate types delegate, while a set of extreme types does not delegate.*

With intermediate uncertainty, the ex-ante probability that a state delegates (i) decreases in the level of uncertainty s, and (ii) increases in the relative value of coordination β.

First, we establish that there always exists a trivial equilibrium where both states do not delegate, no matter their type. This equilibrium arises because neither state can unilaterally create an IO, implying that any deviation by any type of player has no impact on payoffs. As neither state is pivotal, there always exists an equilibrium in which states burn money and make decisions as in the no-delegation game above.

Other equilibria exist as well, but this depends on the amount of uncertainty about preferences. The value of s determines how uncertain each state is about the other state's preferences. If the value of uncertainty is sufficiently low, it is possible to construct an equilibrium in which every type of both states delegates authority to an IO. The gains from IO's and higher coordination always exceed the cost of additional money burning. The main constraint is driven by the most extreme type, as it burns the most money. The deviation to non-delegation is most attractive because even though it would lead to less preferred policies, that type would benefit most from saving on the costs of burned money. On the other extreme, when the value of uncertainty is sufficiently high, then the only equilibrium that exists is one in which every type does not delegate. Money burning incentives are too strong to offset the value of increased coordination.

In between, when the amount of uncertainty is intermediate, there exists an equilibrium in which some type delegate and some do not. The structure of that equilibrium is precisely that moderate types create an IO while extreme types prefer to remain sovereign. Even though they would benefit from delegation under complete information, the necessity of significant money burning prevents the creation of IOs. As moderate types want to mimic extreme types, the latter group of types need to spend more on money burning.

In essence, the key takeaway of the model in which states endogenously create IOs after learning their type shows the importance of several factors. The presence of uncertainty s decreases the scope for potential delegation, and it makes the creation of IOs less likely. The value of coordination β increases the scope for delegation, implying that states want to delegate more often when international coordination is a salient dimension of policymaking. Finally, international disagreement–in the sense that states' types are farther apart–also inhibits the chances with which IOs are created. The reason is that, in the presence of uncertainty, higher disagreement requires state i to burn more money to convince state j to coordinate on i's preferred policy.

Extensions

In this section, we study several extensions in which we relax some of the model's assumptions. Our goal is to better understand the key components of international organizations that generate the results. That is, what do IOs do that states cannot do by themselves? Why and how do IOs generate more coordination and more costly signaling? To this end, we focus on the importance of enforcement and international bargaining. These extensions allow us to understand (i) whether and how an IO's *power* to enforce decisions is important, and (ii) whether and how it matters that the IO is a separate actor or whether states can still be actively bargaining inside IOs. We briefly discuss other extensions after, with more details and formal statements in the appendix.

The Importance of Enforcement

States generally remain autonomous over policy-making even if IOs are involved. Even in a powerful IO such as the European Union, states can still decide to exit or may opt out of certain policies, at a cost. Other IOs, such as the G20, may only serve as a forum in which states exchange information without any enforcement capabilities. How do such enforcement capabilities of IOs affect the strategic behavior of states? We extend the delegation game by allowing states to deviate from the IO's recommended decisions. That is, at some cost $c \geq 0$–which measures the IO's capabilities to enforce policies–each state may deviate from the IO's proposal. The *adapted* delegation game is now as follows after Nature draws types and states burn money:

1. The IO sets policies (d_1, d_2).
2. States 1 and 2 individually make decisions. If state $i = 1, 2$ deviates from d_i, it pays cost $c \geq 0$.

In observing a proposal, states contemplate whether they want to deviate from it. Unless the IO chooses decisions as in the no-delegation game–which is the unique equilibrium without an IO–a profitable deviation always exists if there is no cost of deviating. Hence, if the IO anticipates potential deviations with low c, it wants to prevent this. Alternatively, if c is high enough, it can simply take decisions as in the delegation game in which no deviations were possible. The lower is c, the more the IO must accommodate states to prevent them from deviating, especially if states disagree more.

In turn, the value of c also affects money burning strategies as states anticipate less coordinated decisions from the IO if c is low. Put differently, if c is low, decisions and money burning strategies are closer to the no-delegation game. If c is high, they are closer to strategies as in the delegation game with higher levels of money burning. The following proposition summarizes the results of this extension.

Proposition 5. *In every equilibrium in which the IO makes a proposal that is accepted and from which neither state will deviate, the IO makes proposals as follows:*

- *If* $c \leq \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}\mathbb{E}_{IO}[\theta_1 - \theta_2]^2$, *then* $d = \left(d_1^{ND} + \frac{\sqrt{c}}{1+\beta}, d_2^{ND} - \frac{\sqrt{c}}{1+\beta}\right)$.

- *Otherwise, decisions are as in the delegation game.*

States' money burning strategies are weakly increasing in the level of enforcement c.

Our main results are thus two extreme cases in terms of enforcement capabilities. The key driver of the results is that IOs help solve commitment problems. Another salient aspect is that the level of necessary enforcement to generate decisions as in the delegation model is greater for higher levels of disagreement. That is, when θ_1 and θ_2 are very different, states have very different bliss points, and the IO must have relatively large enforcement capabilities—measured by c—to sustain coordinated decisions.

Bargaining in International Organizations

In the baseline delegation game, we assume that a separate player—the IO—weighed both states' utilities equally and took decisions on their behalf. We now analyze a variation of the model, the *international bargaining* game, in which states bargain over policy. This is to investigate whether the results are purely driven by a focus on centralized IOs rather than member-led IOs. In the latter, it is more natural to think of states being the relevant decision-makers. Still, IOs perform a particular task in helping with cooperation. The idea is that, once states are members of an IO, it becomes easier to commit to agreed-upon policies. That is, once a proposal is accepted, both states are bound to stick to the agreement. The timing of the international bargaining game is as follows. After Nature draws states' types θ_1 and θ_2 and states burn money,

1. Nature selects state 1 as the proposer with probability $p \in [0, 1]$ and state 2 with probability $1 - p$.
2. The proposer observes signals and makes offer (d_1, d_2, T), where $T \in \mathbb{R}$ is a transfer.
3. The non-proposer observes signals and (d_1, d_2, T), and accepts or rejects. If the offer is accepted, the outcome is (d_1, d_2, T), otherwise, states' decisions are as if there were no delegation.

Payoffs are equivalent to the main model and the transfer T is valued linearly. Our equilibrium concept is the same as in the main model. If the offer is rejected, states make decisions as in the no-delegation game. Hence, in contemplating whether to accept or reject an offer, a state compares the offer (d_1, d_2, T) with its outside option of the no-delegation game. Knowing this, the proposer maximizes its payoff subject to the other state's acceptance.

The first result is that the proposer always offers (d_i^D, d_j^D) as the IO does in the delegation game. The proposer then sets the transfer T to make the other state indifferent between the offered (d_i^D, d_j^D) and its outside option from the no-delegation game. Thus, proposal power is beneficial. Equilibrium decisions (d_i^D, d_j^D) are the same regardless of proposal power but the proposer extracts a transfer from the other state. An important determinant of strategies is a state's *bargaining power*, which is the probability that a state makes a proposal (p for state 1 and $1-p$ for state 2). The following proposition establishes the results.

Proposition 6. *In the international bargaining game, the following statements are true in equilibrium:*

1. *If states have equal bargaining power, strategies are equivalent to the delegation game.*
2. *For every bargaining power distribution, states burn more in the delegation game than in the no delegation game.*
3. *If a state obtains more bargaining power, it burns more money.*

With equal bargaining power ($p = \frac{1}{2}$), both states are equally likely to be the proposer, and decision-making strategies and money burning strategies are equivalent to those in the delegation game. Also, for every distribution of bargaining power $p \in [0, 1]$, international bargaining leads to more aggregate money burning than in the no-delegation game. This highlights how even with asymmetric bargaining power, states have stronger incentives to burn money. Finally, states burn more money if their bargaining power increases. For example, if state 1 is the proposer, then if it signals that it is an extreme type, it reveals negative information about the other state's outside option. This allows state 1 to extract a greater transfer from state 2 than otherwise, which determines money burning strategies.

Alternatively, if state 2 is the proposer, then state 1 wants to mimic extreme types to induce the proposer to offer (d_1, d_2) that are closer to 1's bliss point. These different incentives to burn money generate the result that misrepresenting preferences is more attractive if a state is the proposer than otherwise.

Discussion of Other Extensions

In Appendix D, we analyze several other extensions with two main goals. First, we aim to further our understanding of how design choices within IOs affect how and whether states gain from delegation. By designing institutions more carefully, there may be a greater scope for international cooperation and delegation. Second, we aim to understand how asymmetries between the two states affect their value of delegation, and whether such asymmetries have implications for designing institutions.

The main result of this paper highlights the interplay between international coordination and incentives to misrepresent information. In fact, Proposition 7 in the appendix shows how IOs that make more coordinated decisions ensure that states have even greater incentives to misrepresent information. Hence, one way for states to benefit more from international delegation is by limiting the IO's discretion. That is, it may be optimal to let the IO make decisions from a restricted interval. Proposition 8 provides results on optimal delegation intervals from the perspectives of states.

In other extensions, we focus on asymmetries in the environment in which states operate. First, we consider one large state that cares relatively little about coordination compared to a small state. We study the optimal allocation of authority to maximize the gains from co-operation. Proposition 9 shows that with more uncertainty about preferences, it is generally better to give more authority to the larger state. The reason is that the larger state must be compensated for its loss of sovereignty as it does not value the gains from cooperation as much as the smaller state. Another extension allows for asymmetric uncertainty s_i and s_j. That is, it may be the case that one state's domestic conditions (θ_i) are relatively well known compared to the other state's (θ_j). Proposition 10 shows that states with less uncertainty are more likely to delegate authority to IOs. This again emphasizes how uncertainty negatively impacts a state's potential gains from international cooperation within IOs.

Discussion and Conclusion

A key objective of many international organizations is international coordination. They often successfully do so. For example, Pelkmans (2001) shows how in the early 1980s, Western Europe was highly uncoordinated with respect to mobile communication absent the involvement of IOs.[16] Starting in 1985, the European Commission took an active role in promoting the GSM standard in Western Europe and was successful over the following years. Similarly, without formal institutions in the 1960s, European countries failed to develop regional color TV standards, while they were much more successful in the European Community to set HDTV standards in the 1980s (Austin and Milner, 2001).

While states understand that coordination brings about benefits, they also disagree on how to coordinate. This disagreement causes states to attempt to influence their neighbors or IOs in setting policies. Our results indicate that states incur greater signaling costs when IOs are involved in decision-making. It is, however, more difficult to empirically establish what effect IOs have on information transmission. One of the reasons is that costly signaling could take many different forms, which makes it difficult to measure. As in our example in the introduction, the French may give more subsidies and support for nuclear energy than it would otherwise have done if the EU did not exist, partially to influence other EU member states to allow nuclear energy to get a *green label*. Our model predicts that, although decisions would be less coordinated without IOs, states would have fewer incentives to influence other states and misrepresent their preferences. Thus, although the costs states choose to incur would still exist, they would be lower compared to a world with IOs.

[16] Each country had its own monopolistic market with different standards. This was costly for monopolists—and thus also for the governments due to decreased tax revenue—because they could not benefit from the wider European market. For the production of all types of equipment, scale economies were necessary to do so profitably as business demand was simply too small in each individual country.

The value of coordination, uncertainty about preferences, and interstate preference alignment matter for the delegation calculus. The more valuable coordination is to states, the greater is the gain from IOs. Uncertainty, however, makes states less willing to delegate because their incentives to misrepresent information are stronger in the presence of IOs. Therefore, states delegate to IOs as long as they sufficiently value coordination relative to the amount of preference uncertainty and disagreement. The key driver of our results is that IOs brings about commitment power to make more coordinated policies. That is, as long as the IO has a capability of punish states' deviations from agreements, IOs can achieve policy coordination. IOs in our framework are not merely actors or mechanisms that allow states to share information. Instead, these organizations must also actively help states keep their commitments by punishing non-compliance.

Our model has several implications for the design of institutions.[17] The first implication delineates the conditions under which IOs are created in the first place. If coordination is sufficiently important relative to the degree of preference uncertainty, the commitment power of IOs is a valuable asset for states to improve cooperation. Still, however, our results demonstrate that incentives to misrepresent may generate welfare losses that outweigh the gains from cooperation within IOs. As a result, states may prefer to delegate to weaker IOs with fewer enforcement capabilities or prefer to limit their discretion so as to manage their incentives to misrepresent information.

Future work is required to understand the design of IOs when states have different policy-making objectives. For example, if states face a free-rider problem (Kenkel, 2019), it is an open question if IOs are still beneficial when states must transmit information prior to policy-making. Also, states preferences may be determined by a competition motive rather than a coordination motive (Lazer, 2001). Our results indicate that informational issues in

[17]Naturally, IOs are not necessarily designed to maximize the welfare of states. Instead, states bargain over institutional design, and those that have better outside options often have a greater say in the design process (Johns, 2007; Lipscy, 2017). Additionally, states are not in complete control over design and international bureaucrats play an important role (Johnson, 2013, 2014; Johnson and Urpelainen, 2014).

international cooperation can be severe and that the effect of IOs on international cooperation and information transmission is not straightforward.

References

Abbott, Kenneth W and Duncan Snidal. 1998. "Why States Act Through Formal International Organizations." *Journal of Conflict Resolution* 42(1):3–32.

Abreu, Dilip and Faruk Gul. 2000. "Bargaining and Reputation." *Econometrica* 68(1):85–117.

Aghion, Philippe and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105(1):1–29.

Alonso, Ricardo, Wouter Dessein and Niko Matouschek. 2008. "When Does Coordination Require Centralization?" *American Economic Review* 98(1):145–79.

Angulo, Jorge, Joan Calzada and Alejandro Estruch. 2011. "Selection of Standards for Digital Television: The battle for Latin America." *Telecommunications Policy* 35(8):773–787.

Austen-Smith, David and Jeffrey S. Banks. 2000. "Cheap Talk and Burned Money." *Journal of Economic Theory* 91(1):1–16.

Austin, Marc T and Helen V Milner. 2001. "Strategies of European Standardization." *Journal of European Public Policy* 8(3):411–431.

Bendor, Jonathan, Amihai Glazer and Thomas Hammond. 2001. "Theories of Delegation." *Annual Review of Political Science* 4(1):235–269.

Bernheim, B Douglas and Sergei Severinov. 2003. "Bequests as Signals: An explanation for the equal division puzzle." *Journal of Political Economy* 111(4):733–764.

Bradley, Curtis A and Judith G Kelley. 2008. "The Concept of International Delegation." *Law and Contemporary Problems* 71(1):1–36.

Carnegie, Allison and Austin Carson. 2019. "The Disclosure Dilemma: Nuclear intelligence and international organizations." *American Journal of Political Science* 63(2):269–285.

Carson, Austin and Alexander Thompson. 2014. “The Power in Opacity: Rethinking Information in International Organizations.”.

Chapman, Terrence L. 2007. “International Security Institutions, Domestic Politics, and Institutional Legitimacy.” *Journal of Conflict Resolution* 51(1):134–166.

Chen, Ying, Navin Kartik and Joel Sobel. 2008. “Selecting Cheap-Talk Equilibria.” *Econometrica* 76(1):117–136.

Cho, In-Koo and David M Kreps. 1987. “Signaling Games and Stable Equilibria.” *The Quarterly Journal of Economics* 102(2):179–221.

Crane, Rhonda J. 1978. “Communication Standards and the Politics of Protectionism: The case of colour television systems.” *Telecommunications Policy* 2(4):267–281.

Crawford, Vincent P. and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica* 50(6):1431–1451.

Crombez, Christophe, Martijn Huysmans and Wim Van Gestel. 2017. “Choosing an Informative Agenda Setter: The appointment of the Commission in the European Union.” *European Union Politics* 18(2):145–167.

Fang, Songying. 2008. “The Informational Role of International Institutions and Domestic Politics.” *American Journal of Political Science* 52(2):304–321.

Fang, Songying and Randall W Stone. 2012. “International Organizations as Policy Advisors.” *International Organization* 66(4):537–569.

Fey, Mark and Kristopher W Ramsay. 2010. “When is Shuttle Diplomacy Worth the Commute? Information sharing through mediation.” *World Politics* 62(4):529–560.

Gibbons, Robert, John Roberts et al. 2013. “The Handbook of Organizational Economics.”.

Hawkins, Darren G, David A Lake, Daniel L Nielson and Michael J Tierney. 2006. *Delegation and Agency in International Organizations*. Cambridge University Press.

Hooghe, Liesbet and Gary Marks. 2015. "Delegation and Pooling in International Organizations." *The Review of International Organizations* 10(3):305–328.

Hörner, Johannes, Massimo Morelli and Francesco Squintani. 2015. "Mediation and Peace." *The Review of Economic Studies* 82(4):1483–1501.

Johns, Leslie. 2007. "A Servant of Two Masters: Communication and the selection of international bureaucrats." *International Organization* 61(02):245–275.

Johnson, Tana. 2013. "Looking beyond States: Openings for international bureaucrats to enter the institutional design process." *The Review of International Organizations* 8(4):499–519.

Johnson, Tana. 2014. *Organizational Progeny: Why governments are losing control over the proliferating structures of global governance*. Transformations in Governance.

Johnson, Tana and Johannes Urpelainen. 2014. "International Bureaucrats and the Formation of Intergovernmental Organizations: Institutional design discretion sweetens the pot." *International Organization* 68(1):177–209.

Karamychev, Vladimir and Bauke Visser. 2016. "Optimal Signaling with Cheap Talk and Money Burning." *International Journal of Game Theory* 46(3):1–38.

Kartik, Navin. 2007. "A Note on Cheap Talk and Burned Money." *Journal of Economic Theory* 136(1):749–758.

Kartik, Navin. 2009. "Strategic Communication with Lying Costs." *The Review of Economic Studies* 76(4):1359–1395.

Kenkel, Brenton. 2019. "The Efficacy of Cheap Talk in Collective Action Problems." *Journal of Theoretical Politics* 31(3):370–402.

Keohane, Robert O. 1984. *After Hegemony: Cooperation and discord in the world political economy*. Princeton University Press.

Kolotilin, Anton and Hongyi Li. 2021. "Relational Communication." *Theoretical Economics* 16(4):1391–1430.

Koremenos, Barbara. 2008. "When, What, and Why do States Choose to Delegate?" *Law and Contemporary Problems* 71(1):151–192.

Koremenos, Barbara, Charles Lipson and Duncan Snidal. 2001. "The Rational Design of International Institutions." *International Organization* 55(04):761–799.

Lazer, David. 2001. "Regulatory Interdependence and International Governance." *Journal of European Public Policy* 8(3):474–492.

Lipscy, Phillip Y. 2017. *Renegotiating the World Order: Institutional Change in International Relations*. Cambridge University Press.

Meirowitz, Adam, Massimo Morelli, Kristopher W Ramsay and Francesco Squintani. 2019. "Dispute Resolution Institutions and Strategic Militarization." *Journal of Political Economy* 127(1):378–418.

Pelkmans, Jacques. 2001. "The GSM Standard: Explaining a success story." *Journal of European Public Policy* 8(3):432–453.

Rantakari, Heikki. 2008. "Governing Adaptation." *The Review of Economic Studies* 75(4):1257–1285.

Rubinstein, Ariel. 1985. "A Bargaining Model with Incomplete Information about Time Preferences." *Econometrica* 53(5):1151–1172.

Stone, Randall W. 2009. Institutions, Power, and Interdependence. In *Power, interdependence, and nonstate actors in world politics*. Princeton University Press pp. 31–49.

Thompson, Alexander. 2015. *Channels of Power: The UN Security Council and US statecraft in Iraq*. Cornell University Press.

Toulemonde, Eric. 2013. "A Welfare Analysis of the Principle of Mutual Recognition." *European Economic Review* 60:1–16.

Voeten, Erik. 2019. "Making Sense of the Design of International Institutions." *Annual Review of Political Science* 22:147–163.

Supplementary Information for "International Cooperation, Information Transmission, and Delegation"

Online Appendix

Table of Contents

A Model Setup and Equilibrium **1**

A.1 Equilibrium Concept 1

A.2 mD1-Refinement 1

B Proofs for Baseline Model **3**

B.1 Lemma 0 3

B.2 Proof of Lemma 1 4

B.3 Proof of Proposition 1 4

B.4 Proof of Proposition 2 6

B.5 Proof of Proposition 3 (Ex-ante delegation) 8

B.6 Proof of Corollary 1 (Interim Comparison) 9

B.7 Proof of Proposition 4 (Endogenous ex-interim Delegation) 10

C Proofs of Extensions **13**

C.1 Proof of Proposition 5 (Costly Deviations) 13

C.2 Proof of Proposition 6 (International Bargaining) 16

D Other Extensions: Discussions and Proofs **18**

D.1 Coordination Sensitivity and Proposition 7 18

D.2 Limited Discretion and Proposition 8 19

D.3 Heterogeneous Value of Coordination and Proposition 9 21

D.4 One-sided Incomplete Information and Proposition 10 24

A Model Setup and Equilibrium

A.1 Equilibrium Concept

Strategies. In the no-delegation game, state i's strategy is (i) a mapping from types to signals $\left(b_i^{ND}, m_i^{ND}\right) : \Theta_i \to \mathbb{R}_+ \times \mathbb{R}_+$ and (ii) a mapping from state i's type and state j's signals to decisions $d_i^{ND} : \Theta_i \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$. In the delegation game, state i's strategy is a mapping from types to signals $\left(b_i^{D}, m_i^{D}\right) : \Theta_i \to \mathbb{R}_+ \times \mathbb{R}_+$; and for the IO it is a mapping from the signals of both states to decisions $\left(d_1^D, d_2^D\right)$ with $d_i^D : \mathbb{R}_+^2 \times \mathbb{R}_+^2 \to \mathbb{R}$. Let $\mu_i^{ND}\left(b_j, m_j\right) \in \Delta\left(\Theta_j\right)$ be state i's posterior belief about state j's type after observing $\left(b_j, m_j\right)$ and $\mu_{IO}^{D}\left(b_i, m_i, b_j, m_j\right) \in \Delta\left(\Theta_i\right) \times \Delta\left(\Theta_j\right)$ the IO's posterior beliefs about states i and j's types after observing $\left(b_i, m_i, b_j, m_j\right)$. Let $b^{\mathcal{I}} = \left(b_1^{\mathcal{I}}, b_2^{\mathcal{I}}\right)$, $m^{\mathcal{I}} = \left(m_1^{\mathcal{I}}, m_2^{\mathcal{I}}\right)$, $d^{\mathcal{I}} = \left(d_1^{\mathcal{I}}, d_2^{\mathcal{I}}\right)$, $\mu^{ND} = \left(\mu_1^{ND}, \mu_2^{ND}\right)$ and $\mu^D = \mu_{IO}^D$.

Equilibrium. Formally, a *perfect Bayesian equilibrium* (PBE), and from now on an *equilibrium*, is a tuple $\left(b^{\mathcal{I}}, m^{\mathcal{I}}, d^{\mathcal{I}}, \mu^{\mathcal{I}}\right)$ where $\left(b^{\mathcal{I}}, m^{\mathcal{I}}, d^{\mathcal{I}}\right)$ is sequentially rational given $\mu^{\mathcal{I}}$ and $\mu^{\mathcal{I}}$ is Bayesian consistent with $\left(b^{\mathcal{I}}, m^{\mathcal{I}}\right)$.

In the no-delegation game, $\left(b^{ND}, m^{ND}, d^{ND}\right)$ is *sequentially rational* given μ^{ND} if

For each θ_i,

$$\left(b_i^{ND}(\theta_i), m_i^{ND}(\theta_i)\right) \in \text{argmax}_{(b_i,m_i)}\mathbb{E}_i^0\Big[u_i\Big(d_i^{ND}\big(\theta_i, b_j^{ND}(\theta_j), m_j^{ND}(\theta_j)\big), d_j^{ND}\big(\theta_j, b_i, m_i\big), \theta_i, b_i\Big)\Big].$$

For each θ_i, b_j and m_j,

$$d_i^{ND}(\theta_i, b_j, m_j) \in \text{argmax}_{d_i}\mathbb{E}_i\Big[\pi_i\Big(d_i, d_j^{ND}\big(\theta_j, b_i^{ND}(\theta_i), m_i^{ND}(\theta_i)\big), \theta_i\Big)|b_j, m_j\Big].$$

In the delegation game, (b^D, m^D, d^D) is *sequentially rational* given μ^D if

For each θ_i, $(b_i^D(\theta_i), m_i^D(\theta_i)) \in$

$$\text{argmax}_{(b_i,m_i)}\mathbb{E}_i^0\Big[u_i\Big(d_i^D\big(b_i, m_i, b_j^D(\theta_j), m_j^D(\theta_j)\big), d_j^D\big(b_j^D(\theta_j), m_j^D(\theta_j), b_i, m_i\big), \theta_i, b_i\Big)\Big].$$

For each b_i, m_i, b_j and m_j,

$$\left(d_i^D(b_i, m_i, b_j, m_j), d_j^D(b_j, m_j, b_i, m_i)\right) \in \text{argmax}_{(d_i,d_j)}\mathbb{E}_{IO}\Big[u_{IO}\Big(d_i, d_j, \theta_i, \theta_j\Big)|b_i, m_i, b_j, m_j\Big].$$

$\mu^{\mathcal{I}}$ is *Bayesian consistent* with $(b^{\mathcal{I}}, m^{\mathcal{I}})$ if $\mu_i^{\mathcal{I}}$ is the conditional probability distribution of θ_j given (b_j, m_j) derived from the joint distribution over $\Theta_j \times \mathbb{R}_+ \times \mathbb{R}_+$ that the prior distribution and $(b_j^{\mathcal{I}}, m_j^{\mathcal{I}}) : \Theta_j \to \mathbb{R}_+ \times \mathbb{R}_+$ induce.

A.2 mD1-Refinement

We adopt the *monotonic D1* refinement introduced by Bernheim and Severinov (2003). According to the D1 criterion proposed by Cho and Kreps (1987), players should not believe a deviation is made by type θ_i if there is some other type θ_i' who would strictly prefer to deviate for any response from the players that type θ_i would weakly prefer to deviate for. The monotonicity requirement implies that higher types use signals that are weakly more costly (or weakly less costly), and the posterior beliefs should exhibit a monotonic relationship with respect to these signals, including out-of-equilibrium signals.

The D1 refinement rules out the possibility of pooling intervals. The monotonicity requirement gives some order to posterior beliefs regarding signals. In summary, both conditions allow for a unique equilibrium outcome in which information is fully transmitted. This focus on fully informative equilibria is important for two reasons. First, we aim to understand the negative consequences of states' private information. A fully informative equilibrium represents the worst-case scenario in terms of the amount of resources spent on transmitting information. On the other extreme, when states do not incur any costs for transmitting information, information can only be conveyed through cheap talk, which may involve interval equilibria as discussed by Crawford and Sobel (1982). Any other semi-separating equilibrium lies between these two extremes. Second, the uniqueness of the equilibrium allows for a fair comparison between the two games.

We may also consider imposing a cap on the amount of resources that a state can spend, as suggested by Kartik (2009). In such a case, states would have fewer options to differentiate themselves, leading to a pooling interval for extreme types in equilibrium. However, we do not impose this restriction and instead interpret our equilibrium as one where the spending cap is sufficiently high.

Let us consider an equilibrium $\left(b^{\mathcal{I}}, m^{\mathcal{I}}, d^{\mathcal{I}}, \mu^{\mathcal{I}}\right)$. To illustrate the refinement, we focus on the no-delegation game, with similar implications extending to the delegation game. We introduce the following definitions for clarity:

$$\underline{\nu}_1\left(\tilde{b}_1\right) \equiv \max\left\{\frac{\underline{\theta}_2}{1+\beta}+\frac{\underline{\theta}_1}{1+\beta}\beta, \sup_{\theta_1 : b_1^{ND}(\theta_1)\leq \tilde{b}_1} d_2^{ND}\left(\underline{\theta}_2, b_1^{ND}(\theta_1), m_1^{ND}(\theta_1)\right)\right\},$$

$$\overline{\nu}_1\left(\tilde{b}_1\right) \equiv \min\left\{\frac{\overline{\theta}_2}{1+\beta}+\frac{\overline{\theta}_1}{1+\beta}\beta, \inf_{\theta_1 : b_1^{ND}(\theta_1)\geq \tilde{b}_1} d_2^{ND}\left(\overline{\theta}_2, b_1^{ND}(\theta_1), m_1^{ND}(\theta_1)\right)\right\},$$

$$\underline{\nu}_2\left(\tilde{b}_2\right) \equiv \max\left\{\frac{\underline{\theta}_1}{1+\beta}+\frac{\underline{\theta}_2}{1+\beta}\beta, \sup_{\theta_2 : b_2^{ND}(\theta_2)\leq \tilde{b}_2} d_1^{ND}\left(\underline{\theta}_1, b_2^{ND}(\theta_2), m_2^{ND}(\theta_2)\right)\right\},$$

$$\overline{\nu}_2\left(\tilde{b}_2\right) \equiv \min\left\{\frac{\overline{\theta}_1}{1+\beta}+\frac{\overline{\theta}_2}{1+\beta}\beta, \inf_{\theta_2 : b_2^{ND}(\theta_2)\geq \tilde{b}_2} d_1^{ND}\left(\overline{\theta}_1, b_2^{ND}(\theta_2), m_2^{ND}(\theta_2)\right)\right\}.$$

The function $\underline{\nu}_i(\tilde{b}_i)$ represents the lowest policy action that is chosen as an equilibrium response to $\tilde{b}_i$. If no type chooses $\tilde{b}_i$, then it corresponds to the highest policy action chosen as an equilibrium response to $b_i \leq \tilde{b}_i$. In the event that no type chooses $b_i \leq \tilde{b}_i$, $\underline{\nu}_i(\tilde{b}_i)$ represents the highest rationalizable action. On the other hand, the function $\overline{\nu}_i(\tilde{b}_i)$ denotes the highest policy action chosen as an equilibrium response to $\tilde{b}_i$.

Let us denote

$$\hat{u}_i\left(d_j, \theta_i, b_i\right) \equiv \mathbb{E}_i^0\Big[u_i\Big(d_i^{ND}\big(\theta_i, b_j^{ND}(\theta_j), m_j^{ND}(\theta_j)\big), d_j, \theta_i, b_i\Big)\Big].$$

Now, we define

$$A_i\left(\tilde{b}_i, \theta_i\right) \equiv \left[\underline{\nu}_i\left(\tilde{b}_i\right), \overline{\nu}_i\left(\tilde{b}_i\right)\right] \cap \left\{d_j : \hat{u}_i\Big(d_j, \theta_i, \tilde{b}_i\Big) \geq \hat{u}_i\Big(d_j^{ND}\big(\theta_j, b_i^{ND}(\theta_i), m_i^{ND}(\theta_i)\big), \theta_i, b_i^{ND}(\theta_i)\Big)\right\},$$

$$\overline{A}_i\left(\tilde{b}_i,\theta_i\right)\equiv\left[\underline{\nu}_i\left(\tilde{b}_i\right),\overline{\nu}_i\left(\tilde{b}_i\right)\right]\cap\left\{d_j:\hat{u}_i\Big(d_j,\theta_i,\tilde{b}_i\Big)>\hat{u}_i\Big(d_j^{ND}(\theta_j,b_i^{ND}(\theta_i),m_i^{ND}(\theta_i)),\theta_i,b_i^{ND}(\theta_i)\Big)\right\}.$$

Let us consider a fixed amount of burned money $\tilde{b}_i$. We define two sets of responses within the interval $\left[\underline{\nu}_i\left(\tilde{b}_i\right),\overline{\nu}_i\left(\tilde{b}_i\right)\right]$ that induce different incentives for type θ_i to deviate towards $\tilde{b}_i$. The first set $A_i\left(\tilde{b}_i,\theta_i\right)$ comprises responses that provide type θ_i with a weak incentive to deviate towards $\tilde{b}_i$. The second set $\overline{A}_i\left(\tilde{b}_i,\theta_i\right)$ is a stricter version of the first set, including only responses that strictly incentivize type θ_i to deviate towards $\tilde{b}_i$. Let $G_i^{ND}\left(\cdot|b_j,m_j\right)$ denote the cumulative distribution function of $\mu_i^{ND}\left(b_j,m_j\right)$.

An equilibrium $\left(b^{ND},m^{ND},d^{ND},\mu^{ND}\right)$ satisfies the *mD1 criterion* if it fulfills the following conditions:

i) b_i^{ND} is a monotonic function.

ii)
  1. For all m_1,m_1',θ_1, and $b_1>b_1'$, $G_2^{ND}\left(\theta_1|b_1,m_1\right)\geq G_2^{ND}\left(\theta_1|b_1',m_1'\right)$.
  2. For all m_2,m_2',θ_2, and $b_2>b_2'$, $G_1^{ND}\left(\theta_2|b_2,m_2\right)\leq G_1^{ND}\left(\theta_2|b_2',m_2'\right)$.

iii) *Support* $\left[\mu_i^{ND}\left(\tilde{b}_j,\tilde{m}_j\right)\right]=\left\{\theta_j'\right\}$ for any θ_j' and any out-of-equilibrium $\tilde{b}_j$ such that $A_i\left(\tilde{b}_j,\theta_j\right)\subseteq\overline{A}_i\left(\tilde{b}_j,\theta_j'\right)$ for all $\theta_j\neq\theta_j'$ and $A_i\left(\tilde{b}_j,\theta_j'\right)\neq\emptyset$.

B Proofs for Baseline Model

B.1 Lemma 0

Lemma 0. *There are infinite equilibria. Each equilibrium is characterized by a collection of disjoint pooling intervals.*

Proof. Based on the result from Austen-Smith and Banks (2000), all equilibria exhibit the following structure: for each state i, there exists a partition

$$\left(B_0\equiv\underline{\theta}_i,A_1,B_1,\ldots,A_N,B_N,A_{N+1}\equiv\overline{\theta}_i\right),$$

where $B_{j-1}\leq A_j<B_j\leq A_{j+1}$ for all $j\in I=1,\ldots,N$. Within each interval (A_j,B_j), a state pools all types θ_i together by employing a constant amount of burned money and sending the same message $\left(b_i^{\mathcal{I}}(\theta_i),m_i^{\mathcal{I}}(\theta_i)\right)=\left(b_i^{\mathcal{I}}(j),m_i^{\mathcal{I}}(j)\right)$. On the other hand, for types $\theta_i\in(B_j,A_{j+1})$, a state differentiates between them by employing distinct amounts of burned money $b_i^{\mathcal{I}}(\theta_i)$. Furthermore, for any equilibrium where $\theta_i,\theta_i'\in(B_j,A_{j+1})$ and $m_i^{\mathcal{I}}(\theta_i)\neq m_i^{\mathcal{I}}(\theta_i)$, there exists another equilibrium that is output-equivalent, except for the fact that $m_i^{\mathcal{I}}(\theta_i)=m_i^{\mathcal{I}}(\theta_i)$. Hence, for this set of types, the specific message they send in equilibrium becomes irrelevant as they convey information through money burning. Exploiting this property, there is no need to specify the equilibrium messages for any such set of types. The partition is uniquely determined by its collection of pooling intervals $P=\{(A_j,B_j)\,|j\in I\}$. As highlighted by Austen-Smith and Banks (2000), the set of equilibria encompasses a continuum of semi-separating equilibria, ranging from the separating equilibrium $P=\emptyset$ to the pooling equilibrium $P=\Theta_i$. $\square$

B.2 Proof of Lemma 1

We provide a proof sketch that applies to both the no-delegation and delegation games. Without loss of generality, we focus on state 1. Suppose there are two types, θ_1 and θ_1', with $\theta_1 < \theta_1'$, who burn the same amount of money but send different messages m_1 and m_1', respectively, in equilibrium. In this case, it is profitable for type θ_1' to deviate and send message m_1, pretending to be a lower type, in order to induce a lower policy. Therefore, in any equilibrium where different types burn the same amount of money, the posterior beliefs after burning that amount must be the same regardless of the chosen message. We demonstrate in Proposition 2 that b_1^{ND} (the amount of money burned by type θ_1 in the no-delegation game) is a non-increasing function. Intuitively, lower types are willing to burn more money because the benefits of misrepresenting and inducing a lower policy are greater for them. Our refinement eliminates cases where types pool on the same amount of burned money, ensuring that there is differentiation among the types in terms of the money burning strategy.

Suppose, by contradiction, that b_1^{ND} is not a one-to-one function. This implies the existence of an interval $[\theta', \theta''] \subseteq \Theta_1$ such that $b_1^{ND}(\theta) = b^* \geq 0$ and $m^{ND}(\theta) = m^*$ for every $\theta \in [\theta', \theta'']$, while $b_1^{ND}(\theta) \neq b^*$ for every $\theta \notin [\theta', \theta'']$. Considering the uniform prior belief, we have $\mathbb{E}_2[\theta_1 | b^*, m^*] = \frac{(\theta'' - \theta')}{2}$. Furthermore, for any $b > b^*$, it holds that $\mathbb{E}_2[\theta_1 | b, m] \leq \theta'$. Hence, we obtain the inequality:

$$\mathbb{E}_2\left[\theta_1 | b, m\right] \leq \theta' < \frac{(\theta'' - \theta')}{2} = \mathbb{E}_2\left[\theta_1 | b^*, m^*\right].$$

By invoking the fact that state 1's strategy is sequentially rational, the previous inequality implies that

$$\lim_{\theta \to \theta'^-} b_1^{ND}(\theta) > b^*.$$

Now, consider type $(\theta' - \epsilon)$ and a deviation to an off-path action b such that $b^* < b < \lim_{\theta \to \theta'^-} b_1^{ND}(\theta)$. For sufficiently small ϵ, this type has a profitable deviation by choosing action b. Intuitively, this deviation is profitable because type $(\theta' - \epsilon)$ burns a strictly smaller amount b and signals that it is type θ' (since the mD1 criterion restricts the posterior belief after b to assign probability one to type θ'), which is ϵ close to their actual type. This creates a contradiction with the sequential rationality condition for type $(\theta' - \epsilon)$. Hence, we conclude that b_1^{ND} is a one-to-one mapping. Thus, the amount of money burned is fully informative of a state's type and is characterized by an ordinary differential equation (ODE) equation that has a unique solution, which is a strictly monotone function (refer to the proof of Proposition 2 for further details). As a result, the equilibrium is fully informative and unique. □

B.3 Proof of Proposition 1

First, we examine the decisions of the states in the no-delegation game. Then, we analyze the decisions of the international organization (IO) in the delegation game. Finally, we compare the decisions made in both games. Our aim is to demonstrate that in the no-delegation game, state i selects

$$d_i^{ND} = (1-\beta)\theta_i + \beta\left[\frac{1}{1+\beta}\mathbb{E}_i[\theta_j|b_j] + \frac{\beta}{1+\beta}\mathbb{E}_j[\theta_i|b_i]\right].$$

In the final stage, state i has observed its own type θ_i and the amount of money burned b_j by the other state. To simplify notation, let us denote $\mathbb{E}_i[\cdot] \equiv \mathbb{E}_i[\cdot|b_j]$ as the expected value with respect to θ_j using state i's beliefs induced by b_j. State i solves the following optimization problem:

$$\max_{d_i} \mathbb{E}_i\left[-(1-\beta)(d_i-\theta_i)^2 - \beta\left(d_i - d_j(\theta_j)\right)^2\right].$$

By calculating the first-order condition, we obtain the following expression:

$$0 = -2(1-\beta)\left(d_i - \theta_i\right) - 2\beta\mathbb{E}_i\left(d_i - d_j(\theta_j)\right).$$

This expression leads to the following result:

$$d_i = (1-\beta)\theta_i + \beta\mathbb{E}_i[d_j(\theta_j)].$$

Similarly, for state j, we obtain the following expression:

$$d_j = (1-\beta)\theta_j + \beta\mathbb{E}_j[d_i(\theta_i)].$$

Taking the expected values of both expressions, we have:

$$\mathbb{E}_i[d_j(\theta_j)] = \mathbb{E}_i\left[(1-\beta)\theta_j + \beta\mathbb{E}_j[d_i(\theta_i)]\right] = (1-\beta)\mathbb{E}_i[\theta_j] + \beta\mathbb{E}_j[d_i(\theta_i)],$$

$$\mathbb{E}_j[d_i(\theta_i)] = \mathbb{E}_j\left[(1-\beta)\theta_i + \beta\mathbb{E}_i[d_j(\theta_j)]\right] = (1-\beta)\mathbb{E}_j[\theta_i] + \beta\mathbb{E}_i[d_j(\theta_j)].$$

By solving the previous system of equations, we obtain the following result:

$$\mathbb{E}_j[d_i(\theta_i)] = \frac{1}{1+\beta}\mathbb{E}_j[\theta_i] + \frac{\beta}{1+\beta}E_i[\theta_j],$$

$$\mathbb{E}_i[d_j(\theta_j)] = \frac{1}{1+\beta}\mathbb{E}_i[\theta_j] + \frac{\beta}{1+\beta}\mathbb{E}_j[\theta_i].$$

By substituting these expected values into the expressions, we obtain that:

$$d_i^{ND} = (1-\beta)\theta_i + \beta\left[\frac{1}{1+\beta}\mathbb{E}_i[\theta_j] + \frac{\beta}{1+\beta}\mathbb{E}_j[\theta_i]\right],$$

$$d_j^{ND} = (1-\beta)\theta_j + \beta\left[\frac{1}{1+\beta}\mathbb{E}_j[\theta_i] + \frac{\beta}{1+\beta}\mathbb{E}_i[\theta_j]\right].$$

Next, we examine the delegation game. Let $\mathbb{E}_{IO}[\cdot] \equiv \mathbb{E}_{IO}[\cdot|b_i, b_j]$ denote the expected value with respect to θ_i and θ_j given the IO's beliefs induced by b_i and b_j, respectively. Our objective is to demonstrate that in the delegation game, the IO selects

$$d_i^D = \frac{1+\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_i] + \frac{2\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_j].$$

The IO solves the following optimization problem:

$$\max_{d_i,d_j} \frac{1}{2} \cdot \mathbb{E}_{IO}\left[-(1-\beta)\left((d_i-\theta_i)^2+(d_j-\theta_j)^2\right) - \beta\left((d_i-d_j)^2+(d_j-d_i)^2\right)\right].$$

By calculating the first-order condition for d_i, we obtain the following expression:

$$0 = -(1-\beta)\left(d_i - \mathbb{E}_{IO}[\theta_i]\right) - 2\beta\left(d_i - d_j\right).$$

Similarly, we obtain the following expression for d_j by calculating the first-order condition:

$$0 = -(1-\beta)\left(d_j - \mathbb{E}_{IO}[\theta_j]\right) - 2\beta\left(d_j - d_i\right).$$

By solving the previous system of equations, we obtain the following solutions:

$$\begin{aligned} d_i^D &= \frac{1+\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_i] + \frac{2\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_j], \\ d_j^D &= \frac{1+\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_j] + \frac{2\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_i]. \end{aligned}$$

Using the previous expressions, we can calculate the coordination term in equilibrium for each game.

$$\begin{aligned} \Delta d^{ND} &= \left(d_j^{ND} - d_i^{ND}\right) = (1-\beta)\left(\theta_j - \theta_i\right) + \frac{\beta-\beta^2}{1+\beta} \cdot \left(\mathbb{E}_j[\theta_i] - \mathbb{E}_i[\theta_j]\right), \\ \Delta d^D &= \left(d_j^D - d_i^D\right) = \frac{1-\beta}{1+3\beta} \cdot \left(\mathbb{E}_{IO}[\theta_j] - \mathbb{E}_{IO}[\theta_i]\right). \end{aligned}$$

In a fully informative equilibrium, when $\mathbb{E}_{IO}[\theta_i] = \mathbb{E}_j[\theta_i] = \theta_i$ and $\mathbb{E}_{IO}[\theta_j] = \mathbb{E}_i[\theta_j] = \theta_j$, we can simplify the expressions and analyze the comparative results. After some algebraic manipulations, we find that $\Delta d^{ND} > \Delta d^D$, $d_i^D > d_i^{ND}$, and $d_j^D < d_j^{ND}$ when $\theta_i < \theta_j$. This implies that delegation improves coordination but worsens adaptation compared to no delegation. Moreover, each state benefits from delegation. □

B.4 Proof of Proposition 2

In this section, we analyze the money-burning strategies employed by each state in the different game scenarios. Additionally, we demonstrate that a state tends to burn more money in the delegation game compared to the non-delegation game. Let $d_i\left(\theta_i, \theta_i', \theta_j\right)$ represent the decision made by state i when the following conditions are met: (i) state i has type θ_i, (ii) state i signals that its type is θ_i', and (iii) state i believes with probability one that state j is of type θ_j. Similarly, we denote state i's political payoff under the assumption that state

j truthfully signals its type.

$$\tilde{\pi}_i(\theta_i, \theta_i', \theta_j) \equiv \pi_i(d_i(\theta_i, \theta_i', \theta_j), d_j(\theta_j, \theta_j, \theta_i'), \theta_i).$$

Let's consider that state i expends an amount of $b_i(\theta_i')$ in order to signal its type as θ_i'. As a result of this signaling strategy, state i achieves the following payoff:

$$\tilde{\pi}_i(\theta_i, \theta_i', \theta_j) - b_i(\theta_i').$$

A function $b_i(\theta_i)$ is considered incentive-compatible and fully reveals state i's type if the following condition holds:

$$\begin{aligned} &\theta_i \in \arg\max_{\theta_i'} \mathbb{E}_i^0 [\tilde{\pi}_i(\theta_i, \theta_i', \theta_j)] - b_i(\theta_i'), \text{ and} \\ &b_i(\theta_i) \text{ is a strictly monotone function.} \end{aligned}$$

The first requirement implies that when $\theta_i' = \theta_i$, the following first-order condition must be satisfied:

$$\frac{\partial b_i(\theta_i')}{\partial \theta_i'} = \mathbb{E}_i^0 \left[\frac{\partial \tilde{\pi}_i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'} \right].$$

The right-hand side of the equation depends on the prior distribution only through the expected value of θ_j, denoted as $\mathbb{E}_i^0[\theta_j]$. This can be observed by noting that $\frac{\partial \tilde{\pi}_i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}$ is a linear function of θ_j. Hence,

$$\mathbb{E}_i^0 \left[\frac{\partial \tilde{\pi}_i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'} \right] = \frac{\partial \tilde{\pi}_i(\theta_i, \theta_i', \mathbb{E}_i^0[\theta_j])}{\partial \theta_i'}.$$

Hence, the slope of the money burning functions is solely determined by the expected value of the prior. The specific expressions on the right-hand side vary depending on the institution and state being considered. By integrating these expressions with respect to θ_i, we obtain the following:

- In the case of no delegation, we have the following expressions for the money burning functions:
  - i) State 1 burns $b_1^{ND}(\theta_1) = \frac{2(1-\beta)\beta^2}{(1+\beta)^2} \left(\frac{\theta_1^2}{2} - \theta_1 \right) + C$,
  - ii) State 2 burns $b_2^{ND}(\theta_2) = \frac{2(1-\beta)\beta^2}{(1+\beta)^2} \left(\frac{\theta_2^2}{2} + \theta_2 \right) + C$.
- In the case of delegation, we have the following expressions for the money burning functions:
  - i) State 1 burns $b_1^{D}(\theta_1) = \frac{2(1-\beta)\beta}{(1+3\beta)} \left(\frac{\theta_1^2}{2} - \theta_1 \right) + C$,
  - ii) State 2 burns $b_2^{D}(\theta_2) = \frac{2(1-\beta)\beta}{(1+3\beta)} \left(\frac{\theta_2^2}{2} + \theta_2 \right) + C$,

where C is a constant term. These functions exhibit strict convexity and are centered around

1 for state 1 and around -1 for state 2. In order for these functions to be equilibrium strategies, the type that burns the lowest amount for each state and institution must burn zero. Hence, we have the conditions: $b_1^{\mathcal{I}}\left(\min\{\overline{\theta}_1, 1\}\right) = 0$ and $b_2^{\mathcal{I}}\left(\max\{\underline{\theta}_2, -1\}\right) = 0$ for any institution $\mathcal{I} \in \{D, ND\}$. Taking these restrictions into account, we obtain the following expressions:

- In the case of no delegation, we have the following expressions for the money burning functions:
  i) State 1 burns $b_1^{ND}(\theta_1) = \frac{2(1-\beta)\beta^2}{(1+\beta)^2}\left(\theta_1 - \min\{\overline{\theta}_1, 1\}\right)\left(\frac{\theta_1+\min\{\overline{\theta}_1,1\}}{2} - 1\right)$,
  ii) State 2 burns $b_2^{ND}(\theta_2) = \frac{2(1-\beta)\beta^2}{(1+\beta)^2}\left(\theta_2 - \max\{\underline{\theta}_2, -1\}\right)\left(\frac{\theta_2+\max\{\underline{\theta}_2,-1\}}{2} + 1\right)$.
- In the case of delegation, we have the following expressions for the money burning functions:
  i) State 1 burns $b_1^{D}(\theta_1) = \frac{2(1-\beta)\beta}{(1+3\beta)}\left(\theta_1 - \min\{\overline{\theta}_1, 1\}\right)\left(\frac{\theta_1+\min\{\overline{\theta}_1,1\}}{2} - 1\right)$,
  ii) State 2 burns $b_2^{D}(\theta_2) = \frac{2(1-\beta)\beta}{(1+3\beta)}\left(\theta_2 - \max\{\underline{\theta}_2, -1\}\right)\left(\frac{\theta_2+\max\{\underline{\theta}_2,-1\}}{2} + 1\right)$.

If $s \leq 2$, these functions are strictly monotonic, implying the existence of a fully informative equilibrium. However, if $s > 2$, these functions become non-monotonic but still strictly convex, allowing for the possibility that multiple types from the same state may burn the same amount of money. To ensure full information revelation, we can assume that each of these types sends a distinct message. For instance, considering state 1, we can assume that types $\theta_1 < 1$ send the signal m_l, while types $\theta_1 > 1$ send the signal m_r, where $m_l \neq m_r$. Consequently, a fully informative equilibrium still exists. Moreover, it is direct to see check that $b_i^D(\theta_i) > b_i^{ND}(\theta_i)$, $i \in \{1, 2\}$. Consequently, for a fixed state and type, the amount of money burned under delegation is higher compared to the no delegation scenario. □

B.5 Proof of Proposition 3 (Ex-ante delegation)

For any institution $\mathcal{I} \in \{D, ND\}$, denote $\Pi_i^{\mathcal{I}} \equiv \mathbb{E}^0[\pi_i^{\mathcal{I}}(\theta)]$ and $B_i^{\mathcal{I}} \equiv \mathbb{E}^0[b_i^{\mathcal{I}}(\theta_i)]$. States' payoffs are symmetric so

$$\Pi^{\mathcal{I}} \equiv \Pi_1^{\mathcal{I}} = \Pi_2^{\mathcal{I}},$$
$$B^{\mathcal{I}} \equiv B_1^{\mathcal{I}} = B_2^{\mathcal{I}}.$$

Using the previous results, after some algebra we obtain

- In case of no delegation:

$$\Pi^{ND} = -\frac{2}{3}\frac{(1-\beta)\beta(6+s^2)}{(1+\beta)^2},$$

$$B^{ND} = \frac{1}{3}\frac{(1-\beta)\beta^2(9+s^2-3\max\{-1,1-s\}^2-6\max\{-1,1-s\})}{(1+\beta)^2}.$$

- In case of delegation:

$$\Pi^D = -\frac{2}{3}\frac{(1-\beta)\beta(6+s^2)}{(1+3\beta)},$$

$$B^D = \frac{1}{3}\frac{(1-\beta)\beta(9+s^2-3\max\{-1,1-s\}^2-6\max\{-1,1-s\})}{(1+3\beta)}.$$

It is direct that $\Pi^D > \Pi^{ND}$ and $B^D > B^{ND}$.

We need to compare $(\Pi^{ND} - B^{ND})$ and $(\Pi^D - B^D)$. After some algebra we obtain:

$$(\Pi^{ND} - B^{ND}) - (\Pi^D - B^D) = \begin{cases} \frac{2}{3}\frac{(1-\beta)^2\beta\left(s^2+12\beta+12\right)}{(\beta+1)^2(1+3\beta)} & \text{if } s \geq 2 \\ \frac{4}{3}\frac{(1-\beta)^2\beta\left[(6-s)s-3\beta\left(s^2-4s+2\right)\right]}{(1+\beta)^2(1+3\beta)} & \text{if } s < 2. \end{cases}$$

The term $[(6-s)s - 3\beta\,(s^2-4s+2)]$ is increasing in s and has one zero whenever $s < 2$. Let $\hat{s}$ be the value in s that makes zero the last term. Then we have the following:

If $s \leq \hat{s}$, then $(\Pi^{ND} - B^{ND}) \leq (\Pi^D - B^D)$,

If $s > \hat{s}$, then $(\Pi^{ND} - B^{ND}) > (\Pi^D - B^D)$.

Thus, delegation is beneficial only if the level of uncertainty is sufficiently low

$$s \leq \hat{s}(\beta) \equiv \frac{3+6\beta-(9+30\beta+18\beta^2)^{\frac{1}{2}}}{1+3\beta}.$$

B.6 Proof of Corollary 1 (Interim Comparison)

Define $\tilde{s} \equiv \frac{2(2+3\beta)}{\beta} - 4\left(\frac{1+3\beta+2\beta^2}{\beta^2}\right)^{\frac{1}{2}}$. We will show the following: (i) There exists a cutoff $\hat{\theta}_2$ such that if $\theta_2 \leq \hat{\theta}_2$, state 2 prefers delegation, and if $\theta_2 > \hat{\theta}_2$, the state prefers no delegation. (ii) If $s > \tilde{s}$, then $\underline{\theta}_2 < \hat{\theta}_2 < \overline{\theta}_2$. In this case, $\frac{\partial\hat{\theta}_2}{\partial s} < 0$ and $\frac{\partial\hat{\theta}_2}{\partial\beta} > 0$. (iii) If $s \leq \tilde{s}$, then $\hat{\theta}_2 = \overline{\theta}_2$.

Consider the difference in ex-post utilities for state 2 between the no-delegation and delegation game in equilibrium:

$$\begin{aligned}\Delta u \equiv &\left[-(1-\beta)\left(d_2^{ND}-\theta_2\right)^2 - \beta\left(\Delta d^{ND}\right)^2 - b_2^{ND}(\theta_2)\right] \\ &- \left[-(1-\beta)\left(d_2^{D}-\theta_2\right)^2 - \beta\left(\Delta d^{D}\right)^2 - b_2^{D}(\theta_2)\right].\end{aligned}$$

Here, $\Delta d^{ND} = \left(d_2^{ND} - d_1^{ND}\right)$ and $\Delta d^D = \left(d_2^D - d_1^D\right)$. After simplification, we obtain the following:

$$\frac{\partial\Delta u}{\partial\theta_2} = \frac{2(1-\beta)^2\beta}{(1+\beta)^2(1+3\beta)}\left(1+\theta_2\beta\left(2+\theta_1+\theta_2\right)\right).$$

The term above is linear in θ_1. Taking the expected value with respect to θ_1, we get:

$$\frac{2(1-\beta)^2\beta}{(1+\beta)^2(1+3\beta)}\left(1+\theta_2\beta\left(1+\theta_2\right)\right),$$

which is strictly positive. Hence, if a type θ_2 prefers no delegation, any type θ_2' with $\theta_2' > \theta_2$ strictly prefers no delegation. Since $\underline{\theta}_2$ strictly prefers delegation, there exists a cut-off value $\hat{\theta}_2 > \underline{\theta}_2$ such that if $\theta_2 \le \hat{\theta}_2$, state 2 prefers delegation, and if $\theta_2 > \hat{\theta}_2$, the state prefers no delegation.

The cut-off is defined by the unique solution of $\mathbb{E}_2^0[\Delta u] = 0$ when the solution is strictly lower than $\overline{\theta}_2$, and it is equal to $\overline{\theta}_2$ when $\mathbb{E}_2^0[\Delta u] \le 0$ for every $\theta_2 \in [\underline{\theta}_2, \overline{\theta}_2]$. After some algebra, we obtain that $\hat{\theta}_2 < \overline{\theta}_2$ is equivalent to $s > \tilde{s}$.

If $s > \tilde{s}$, we find that $\frac{\partial \hat{\theta}_2}{\partial s} < 0$ and $\frac{\partial \hat{\theta}_2}{\partial \beta} > 0$. These results imply that $\Pr\left(\theta_2 \le \hat{\theta}_2\right) = F_s\left(\hat{\theta}_2\right)$ decreases with s and increases with β when F_s is a symmetric distribution. Note that as we change s, we must consider a different distribution function to represent a probability measure. Hence, we explicitly indicate the dependence of the distribution on the value of s. □

B.7 Proof of Proposition 4 (Endogenous ex-interim Delegation)

In describing cut-points on s ($\underline{s}$ and $\overline{s}$), we omit the description of its dependence on β.

Proof. We prove the following statement: When uncertainty is sufficiently low ($0 < s < \underline{s}$), there is an equilibrium where every state's type prefers to delegate. In the case of intermediate values of uncertainty ($\underline{s} < s < \overline{s}$), there is an equilibrium where a state prefers to delegate if and only if that state's type is sufficiently moderate. Furthermore, in this intermediate case, the ex-ante probability that a state delegates is (i) decreasing as the level of uncertainty s increases, and (ii) increasing as the relative value of coordination β increases.

Define $\underline{s} \equiv \frac{3(3\beta^2+3\beta+2)}{2(1-\beta)\beta} - \frac{\sqrt{3(23\beta^4+62\beta^3+59\beta^2+36\beta+12)}}{2(1-\beta)\beta}$ and $\overline{s} \equiv 1 - \sqrt{\frac{1+3\beta}{2(1+\beta)}}$. We begin by proving the following: if the value of s lies between $\underline{s}$ and $\overline{s}$, there exists an equilibrium where the following conditions hold: There exist cutoff types $\hat{\theta}_1$ and $\hat{\theta}_2$ satisfying $\hat{\theta}_1 = -\hat{\theta}_2$ and $1-s < \hat{\theta}_2 < 1+s$. Under this equilibrium, if a state's type is less than or equal to $\hat{\theta}_1$ (or its type is greater than or equal to $\hat{\theta}_2$), the state prefers delegation. On the other hand, if the state's type is greater than $\hat{\theta}_1$ (or its type is less than $\hat{\theta}_2$), the state prefers not to delegate. Let us define the following intervals: $\Theta_1^D = [\hat{\theta}_1, -1+s]$, $\Theta_1^{ND} = [-1-s, \hat{\theta}_1]$, $\Theta_2^D = [1-s, \hat{\theta}_2]$, and $\Theta_2^{ND} = [\hat{\theta}_2, 1+s]$.

We now consider the scenario where states adhere to the aforementioned strategy. We can utilize our previous findings to comprehend the outcomes following each history once the decision to establish an IO has been made. Assuming an IO is formed, the subsequent game resembles our baseline model, with the exception that the types of state 1 are believed to be drawn from the set Θ_1^D, while the types of state 2 are believed to be drawn from the set Θ_2^D. This modification impacts our construction in a single aspect: the expected value of state 1's type is $\frac{\hat{\theta}_1+(-1+s)}{2}$, while the expected value of state 2's type is $\frac{\hat{\theta}_2+(1-s)}{2}$.

In the scenario where an IO is not created, there are three possible histories leading to this outcome:

(i) Both states choose not to delegate, in which case the types are known to be drawn from Θ_1^{ND} and Θ_2^{ND}.

(ii) State 1 delegates while state 2 does not delegate, in which case the types are known to be drawn from Θ_1^{D} and Θ_2^{ND}.

(iii) State 1 does not delegate while state 2 delegates, in which case the types are known to be drawn from Θ_1^{ND} and Θ_2^{D}.

In each of these cases, the subsequent game follows a similar structure to our baseline model, as discussed previously. However, our construction introduces an additional consideration: the determination of the most moderate type for each state is influenced by the specific history that led to the non-creation of the IO.

In all of these continuation games, it is important to note that the equilibrium behavior is unique, and the history only influences the money-burning behavior. Along the equilibrium path, information is perfectly transmitted, and decisions are made without any uncertainty regarding the types of states.

To demonstrate that the delegating behavior is indeed an equilibrium strategy, let's concentrate our analysis on state 1. We will calculate the expected payoff for a type θ_1 when choosing not to delegate.

$$\begin{aligned}\mathbb{E}u_1^{ND}(\theta_1) = &\int_{1-s}^{1+s}\left(-\frac{(1-\beta)\beta(\theta_1-\theta_2)^2}{(1+\beta)^2}\right)\times\frac{d\theta_2}{2s}\\&-\int_{1-s}^{\hat{\theta}_2} b_1^{ND}\left(\theta_1|\mathbb{E}\left[\theta_2\right]=\frac{\hat{\theta}_2+(1-s)}{2}\right)\times\frac{d\theta_2}{2s}\\&-\int_{\hat{\theta}_2}^{1+s} b_1^{ND}\left(\theta_1|\mathbb{E}\left[\theta_2\right]=\frac{\hat{\theta}_2+(1+s)}{2}\right)\times\frac{d\theta_2}{2s}.\end{aligned}$$

Similarly, let's consider the expected payoff for a type θ_1 when choosing delegation.

$$\begin{aligned}\mathbb{E}u_1^{D}(\theta_1) = &\int_{1-s}^{\hat{\theta}_2}\left(-\frac{(1-\beta)\beta(\theta_1-\theta_2)^2}{1+3\beta} - b_1^{D}\left(\theta_1|\mathbb{E}\left[\theta_2\right]=\frac{\hat{\theta}_2+(1-s)}{2}\right)\right)\times\frac{d\theta_2}{2s}\\&+\int_{\hat{\theta}_2}^{1+s}\left(-\frac{(1-\beta)\beta(\theta_1-\theta_2)^2}{(1+\beta)^2} - b_1^{ND}\left(\theta_1|\mathbb{E}\left[\theta_2\right]=\frac{\hat{\theta}_2+(1+s)}{2}\right)\right)\times\frac{d\theta_2}{2s}.\end{aligned}$$

After performing the algebraic calculations, we obtain the expression:

$$\frac{\partial\left(\mathbb{E}u_1^{D}(\theta_1)-\mathbb{E}u_1^{ND}(\theta_1)\right)}{\partial\theta_1} = \frac{(1-\beta)^2\beta}{2(1+\beta)(1+3\beta)}\left(3\hat{\theta}_2+1-s\right)\left(\hat{\theta}_2-1+s\right).$$

This expression is positive because $1 - s < \hat{\theta}_2 < 1 + s$. Hence, there exists a unique value θ_1, denoted as $\hat{\theta}_1$, such that $\mathbb{E}u_1^{ND}(\theta_1) = \mathbb{E}u_1^{D}(\theta_1)$. We can directly calculate the value of $\hat{\theta}_1$ from the condition $\mathbb{E}u_1^{D}(\hat{\theta}_1) = u_1^{ND}(\hat{\theta}_1)$ when $\hat{\theta}_1 = -\hat{\theta}_2$. The expression for $\hat{\theta}_1$ is as follows:

$$
\begin{aligned}
\hat{\theta}_1 = &-\frac{\beta\left(\beta(7s+2)+5s-2\right)}{5\beta^2+\beta-6} \\
&-\sqrt{\frac{\beta^4\left(8s^2+56s+23\right)+10\beta^3(6s-1)+\beta^2\left(4s^2-32s-37\right)+12\beta\left(2s^2-5s+1\right)+12(s-1)^2}{\left(5\beta^2+\beta-6\right)^2/3}}.
\end{aligned}
$$

We observe that $\hat{\theta}_1 < -1 + s$ is equivalent to $s < \bar{s}$. On the other hand, $\hat{\theta}_1 > -1 - s$ is equivalent to $s > \underline{s}$. Therefore, when $\underline{s} < s < \bar{s}$, types $\theta_1 \geq \hat{\theta}_1$ ($\theta_2 \leq \hat{\theta}_2$) prefer to delegate, while types $\theta_1 < \hat{\theta}_1$ ($\theta_2 > \hat{\theta}_2$) prefer not to delegate.

To prove that if $0 < s \leq \underline{s}$, there exists an equilibrium where every state's type prefers to delegate, let's consider the case where states adhere to the previous strategy. For a state 1 type θ_1 who proposes delegation, the expected payoff obtained is given by:

$$\mathbb{E}u_1^{D}(\theta_1) = \int_{1-s}^{1+s}\left(-\frac{(1-\beta)\beta(\theta_1-\theta_2)^2}{1+3\beta} - b_1^D\left(\theta_1 | \mathbb{E}\left[\theta_2\right]=1\right)\right) \times \frac{d\theta_2}{2s}.$$

Now, suppose that after choosing not to delegate, the other state believes that state 1 is of type θ_1' with probability 1. The expected payoff for a type θ_1 in the case of choosing not to delegate is given by:

$$\mathbb{E}u_1^{ND}(\theta_1) = \int_{1-s}^{1+s}\left(-\frac{(1-\beta)\beta(\theta_1+\beta(\theta_1-\theta_1')-\theta_2)^2}{(1+\beta)^2}\right) \times \frac{d\theta_2}{2s}.$$

Let's consider the case where $\theta_1' = -1 - s$. In this scenario, we find that

$$\frac{\partial(\mathbb{E}u_1^{D}(\theta_1) - \mathbb{E}u_1^{ND}(\theta_1))}{\partial\theta_1} = \frac{2(1-\beta)\beta s^2\left(3\beta^2(s+\theta_1+1)+\beta(s+2\theta_1)-\theta_1+1\right)}{(1+\beta)(1+3\beta)} > 0.$$

Hence, it is sufficient to examine whether the type $-1 - s$ prefers delegation or not. After performing some algebraic calculations, we obtain that:

$$\mathbb{E}u_1^{D}(-1-s) - \mathbb{E}u_1^{ND}(-1-s) = -\frac{4(1-\beta)\beta s^2\left(\beta^2\left(s^2+9s+3\right)-\beta\left(s^2-9s+3\right)+6s\right)}{3(1+\beta)^2(1+3\beta)}.$$

Thus, $\mathbb{E}u_1^{D}(-1-s) > \mathbb{E}u_1^{ND}(-1-s)$ if and only if $s < \underline{s}$. Therefore, when $s < \underline{s}$, every type θ_1 prefers to delegate. Next, let's justify why $\theta_1' = -1 - s$. Note that when $0 < s < \underline{s}$, we observe that $\frac{\partial(\mathbb{E}u_1^{D}(\theta_1) - \mathbb{E}u_1^{ND}(\theta_1))}{\partial\theta_1} > 0$ for any θ_1'. Thus, based on the D1 refinement, we can conclude that $\theta_1' = -1 - s$. Consequently, when $0 < s < \underline{s}$, there is an equilibrium where every type prefers to delegate. $\square$

C Proofs of Extensions

C.1 Proof of Proposition 5 (Costly Deviations)

We restate Proposition 5 formally:

Proposition 5. *In every equilibrium, the IO proposes*

$$d = \begin{cases} \left(d_1^{ND} + \frac{\sqrt{c}}{1+\beta}, d_2^{ND} - \frac{\sqrt{c}}{1+\beta}\right) & \text{if } c \leq \frac{(1-\beta)^2\beta^2\mathbb{E}_{IO}[\theta_1-\theta_2]^2}{(1+3\beta)^2}, \\ \left(d_1^D, d_2^D\right) & \text{if } c > \frac{(1-\beta)^2\beta^2\mathbb{E}_{IO}[\theta_1-\theta_2]^2}{(1+3\beta)^2}. \end{cases}$$

Additionally, the money burning functions b_i^c exhibit the following properties for a fixed θ_i:

i) $b_i^{ND}(\theta_i) \leq b_i^c(\theta_i) \leq b_i^D(\theta_i)$,

ii) For $c \leq \underline{c} \equiv \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}\left(\max\{\underline{\theta}_2 - \overline{\theta}_1, 0\}\right)^2$, $b_i^c(\theta_i) = b_i^{ND}(\theta_i)$,

iii) For $c \geq \overline{c} \equiv \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}(2+2s)^2$, $b_i^c(\theta_i) = b_i^D(\theta_i)$,

iv) $b_i^c(\theta_i)$ *is weakly increasing in c.*

Proof. Given the IO's proposal (d_1, d_2), let $d_i^{br}(d_j)$ denote state i's best response policy to d_j, defined as the solution to:

$$d_i^{br}(d_j) \in \text{argmax}_{d_i} - (1-\beta)(d_i - \theta_i)^2 - \beta(d_i - d_j)^2.$$

Note that by construction, $d_i^{br}(d_j^{ND}) = d_i^{ND}$. Furthermore, we have $d_i^{br}(d_j) = (1-\beta)\theta_i + \beta d_j$. State i will deviate to $d_i^{br}(d_j)$ if the following condition is satisfied:

$$-(1-\beta)\left(d_i^{br}(d_j) - \theta_i\right)^2 - \beta\left(d_i^{br}(d_j) - d_j\right)^2 - c > -(1-\beta)(d_i - \theta_i)^2 - \beta(d_i - d_j)^2.$$

This condition defines the policies for each state given the IO's recommendation (d_1, d_2).

We define the updated policies as:

$$d_i'(d_1, d_2) = \begin{cases} d_i^{br}(d_j) & \text{if state } i \text{ deviates,} \\ d_i & \text{if state } i \text{ does not deviate.} \end{cases}$$

In the analysis of the IO's proposed policies, we consider two extreme cases and then draw conclusions for the intermediate case. If we consider the IO's most preferred policies without potential deviations, which correspond to the solution of the delegation model, we find that state i will not deviate if and only if:

$$c \geq \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}(\theta_i - \theta_j)^2.$$

The maximum value that the right-hand side (RHS) can take is

$$\overline{c} \equiv \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}\left(\overline{\theta}_2 - \underline{\theta}_1\right)^2 = \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}(2+2s)^2 > 0.$$

The minimum value that the RHS can take is

$$\underline{c} \equiv \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}\left(\max\{\underline{\theta}_2 - \overline{\theta}_1, 0\}\right)^2 = \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}\left(\max\{2-2s, 0\}\right)^2 \geq 0.$$

If $c \geq \overline{c}$, both states will not deviate for any possible types. The IO will propose its most preferred policies, which are derived in the delegation game, and these policies will be accepted:

$$d_1 = \frac{1+\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_1] + \frac{2\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_2],$$
$$d_2 = \frac{1+\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_2] + \frac{2\beta}{1+3\beta}\mathbb{E}_{IO}[\theta_1].$$

Suppose $c \leq \underline{c}$. If the IO were to propose its most preferred policies, both states would have a profitable deviation for any type. In this case, the IO will optimally propose policies to make both states indifferent between the proposed policies and their most profitable deviations.

$$\mathbb{E}_{IO}\left[-(1-\beta)(d_1-\theta_1)^2 - \beta(d_1-d_2)^2\right] = \mathbb{E}_{IO}\left[-(1-\beta)(d_1'(d_2)-\theta_1)^2 - \beta(d_1'(d_2)-d_2)^2\right] - c,$$
$$\mathbb{E}_{IO}\left[-(1-\beta)(d_2-\theta_2)^2 - \beta(d_1-d_2)^2\right] = \mathbb{E}_{IO}\left[-(1-\beta)(d_2'(d_1)-\theta_2)^2 - \beta(d_2'(d_1)-d_2)^2\right] - c.$$

These conditions yield the following policies:

$$d_1 = \frac{1}{1+\beta}\mathbb{E}_{IO}[\theta_1] + \frac{\beta}{1+\beta}\mathbb{E}_{IO}[\theta_2] + \frac{\sqrt{c}}{1+\beta},$$
$$d_2 = \frac{1}{1+\beta}\mathbb{E}_{IO}[\theta_2] + \frac{\beta}{1+\beta}\mathbb{E}_{IO}[\theta_1] - \frac{\sqrt{c}}{1+\beta}.$$

Note that $d_1 = d_1^{ND} + \frac{\sqrt{c}}{1+\beta}$ and $d_2 = d_2^{ND} - \frac{\sqrt{c}}{1+\beta}$.

Consider now the incentives for money burning. If $c \geq \overline{c}$, each state always anticipates that the IO will propose its most preferred policies. Therefore, the money burning functions in this case remain the same as in the delegation game.

$$b_1(\theta_1) = \frac{2(1-\beta)\beta}{1+3\beta}f_1(\theta_1), \qquad b_2(\theta_2) = \frac{2(1-\beta)\beta}{1+3\beta}f_2(\theta_2).$$

If $c \leq \underline{c}$, each state anticipates that the IO will always strive to make both states indifferent between the proposal and their optimal deviations. Therefore, the money burning functions are given by:

$$b_1(\theta_1) = \frac{2(1-\beta)\beta^2}{(1+\beta)^2}f_1(\theta_1) - \frac{2(1-\beta)(1+2\beta)}{(1+\beta)^2}\theta_1\sqrt{c} = b_1^{ND}(\theta_1) - \frac{2(1-\beta)(1+2\beta)}{(1+\beta)^2}\theta_1\sqrt{c},$$
$$b_2(\theta_2) = \frac{2(1-\beta)\beta^2}{(1+\beta)^2}f_2(\theta_2) + \frac{2(1-\beta)(1+2\beta)}{(1+\beta)^2}\theta_2\sqrt{c} = b_2^{ND}(\theta_2) + \frac{2(1-\beta)(1+2\beta)}{(1+\beta)^2}\theta_2\sqrt{c}.$$

Note that for a fixed θ_i, b_i is increasing in c.

In general, state 1 anticipates that the IO proposes its most preferred policies if and only if the following condition holds:

$$c \geq \frac{(1-\beta)^2\beta^2}{(1+3\beta)^2}\left(\theta_2 - \theta_1'\right)^2,$$

which is equivalent to the condition:

$$\theta_2 < \hat{\theta}_2(\theta_1') \equiv \theta_1' + \frac{(1+3\beta)}{(1-\beta)\beta}\sqrt{c}.$$

Let $\hat{U}^i(\theta_i, \theta_i', \theta_j)$ denote the ex-ante payoff U^i when the IO proposes its most preferred policies, and let $U_c^i(\theta_i, \theta_i', \theta_j)$ denote the ex-ante payoff U^i when the IO proposes its restricted policies. Define

$$P\left(\theta_j < \hat{\theta}_j(\theta_i')\right) \equiv \frac{\min\{\max\{1+s-\hat{\theta}_j(\theta_i'), 0\}, 1\}}{2s}.$$

Thus

$$\begin{aligned}
\mathbb{E}_i^0\left[\frac{\partial U^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\right] &= \mathbb{E}_i^0\left[\frac{\partial \hat{U}^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\middle|\theta_j < \hat{\theta}_j(\theta_i')\right] P\left(\theta_j < \hat{\theta}_j(\theta_i')\right) \\
&\quad + \mathbb{E}_i^0\left[\frac{\partial U_c^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\middle|\theta_j > \hat{\theta}_j(\theta_i')\right]\left(1 - P\left(\theta_j < \hat{\theta}_j(\theta_i')\right)\right) \\
&= \int_{1-s}^{\hat{\theta}_j(\theta_i')} \frac{\partial \hat{U}^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\frac{1}{2s}d\theta_j \\
&\quad + \int_{\hat{\theta}_j(\theta_i')}^{1+s} \frac{\partial U_c^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\frac{1}{2s}d\theta_j.
\end{aligned}$$

Also,

$$\frac{\partial}{\partial c}\mathbb{E}_1^0\left[\frac{\partial U^1(\theta_1, \theta_1', \theta_2)}{\partial \theta_1'}\right] = \int_{\hat{\theta}_2(\theta_1')}^{1+s} \frac{\partial}{\partial c}\frac{\partial U_c^1(\theta_1, \theta_1', \theta_2)}{\partial \theta_1'}\frac{1}{2s}d\theta_2 < 0,$$

and

$$\frac{\partial}{\partial c}\mathbb{E}_2^0\left[\frac{\partial U^2(\theta_2, \theta_2', \theta_1)}{\partial \theta_2'}\right] = \int_{\hat{\theta}_1(\theta_2')}^{1+s} \frac{\partial}{\partial c}\frac{\partial U_c^2(\theta_2, \theta_2', \theta_1)}{\partial \theta_2'}\frac{1}{2s}d\theta_1 > 0.$$

The money burning function satisfies the following expression when $\theta_i' = \theta_i$:

$$\frac{\partial b_i(\theta_i')}{\partial \theta_i'} = \mathbb{E}_i^0\left[\frac{\partial U^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\right].$$

Then, the slope of the function $b_i(\theta_i)$ at θ_i becomes more pronounced as c increases. Additionally, since $b_1(\overline{\theta}_i) = 0$ and $b_2(\underline{\theta}_i) = 0$, for a fixed θ_i, the value $b_i(\theta_i)$ increases as c increases. Now, suppose that

$$c \leq \frac{(1-\beta)^2\beta^2\mathbb{E}_{IO}\left[\theta_1 - \theta_2\right]^2}{(1+3\beta)^2}.$$

Consider the difference in ex-post utilities in equilibrium for state j between the no-delegation and delegation game:

$$\Delta u = \underbrace{\left[-(1-\beta)(d_j^{ND}-\theta_j)^2 - \beta\left(\Delta d^{ND}\right)^2 - b_j^{ND}(\theta_j)\right]}_{\text{No delegation}} - \underbrace{\left[-(1-\beta)(d_j^{c,D}-\theta_j)^2 - \beta\left(\Delta d^{c,D}\right)^2 - b_j^{D}(\theta_j)\right]}_{\text{Delegation}},$$

where $\Delta d^{ND} = \left(d_j^{ND} - d_i^{ND}\right)$ and $\Delta d^{c,D} = \left(d_j^{c,D} - d_i^{c,D}\right)$. After some algebraic manipulation, we obtain:

$$\frac{\partial^2 \Delta u}{\partial c \partial \theta_j} = \frac{c^{1/2} + \beta\left(3c^{1/2} - 1 + \theta_i - \theta_j\right) + \beta^2\left(1 - \theta_j - \theta_i\right)}{(1+\beta)^2 c^{1/2}} + \frac{\partial^2 b_j^{c,D}}{\partial c \partial \theta_j}.$$

□

C.2 Proof of Proposition 6 (International Bargaining)

Let $U_i^{ND} \equiv -(1-\beta)(d_i^{ND}-\theta_i)^2 - \beta(d_i^{ND}-d_j^{ND})^2$ denote the outside option of state i. First, we will prove the following lemma.

Lemma 2. *In the equilibrium of the international bargaining game, state $i = 1, 2$ proposes the following policy:*

$$\begin{aligned} d_i^{IB} &= \frac{1+\beta}{1+3\beta}\theta_i + \frac{2\beta}{1+3\beta}\mathbb{E}_i[\theta_j], \\ d_j^{IB} &= \frac{1+\beta}{1+3\beta}\mathbb{E}_i[\theta_j] + \frac{2\beta}{1+3\beta}\theta_i, \\ T^{IB} &= \mathbb{E}_i\mathbb{E}_j\left[-(1-\beta)(d_j^{IB}-\theta_j)^2 - \beta(d_i^{IB}-d_j^{IB})^2\right] - \mathbb{E}_i\mathbb{E}_j\left[U_j^{ND}\right]. \end{aligned}$$

State i accepts a proposal (d_i, d_j, T) if and only if

$$\mathbb{E}_i\left[-(1-\beta)(d_i-\theta_i)^2 - \beta(d_i-d_j)^2\right] - T \geq \mathbb{E}_i\left[U_i^{ND}\right].$$

Proof. We limit our analysis to strategies where states accept an offer when they are indifferent. If this is not the case, the maximization problem may not have a solution.

When state i is the proposer, it solves the following problem:

$$\begin{aligned} \max_{d_i, d_j, T} \quad & \mathbb{E}_i\left[-(1-\beta)(d_i-\theta_i)^2 - \beta(d_i-d_j)^2 + T\right] \\ \text{s.t.} \quad & \mathbb{E}_i\mathbb{E}_j\left[-(1-\beta)(d_j-\theta_j)^2 - \beta(d_i-d_j)^2 - T\right] \geq \mathbb{E}_i\mathbb{E}_j\left[U_j^{ND}\right]. \end{aligned}$$

In the optimum, the restriction is binding. The problem can be formulated as follows:

$$\begin{aligned}\max_{d_i,d_j} \quad & \mathbb{E}_i \left[-(1-\beta)(d_i-\theta_i)^2 - \beta(d_i-d_j)^2 + \mathbb{E}_j \left[-(1-\beta)(d_j-\theta_j)^2 - \beta(d_i-d_j)^2\right]\right] \\ & - \mathbb{E}_i\mathbb{E}_j \left[U_j^{ND}\right]\end{aligned}$$

with

$$T = \mathbb{E}_i\mathbb{E}_j \left[-(1-\beta)(d_j-\theta_j)^2 - \beta(d_i-d_j)^2\right] - \mathbb{E}_i\mathbb{E}_j \left[U_j^{ND}\right].$$

The optimum is given by the following solution:

$$\begin{aligned}d_i^{IB} &= \frac{1+\beta}{1+3\beta}\theta_i + \frac{2\beta}{1+3\beta}\mathbb{E}_i\left[\theta_j\right], \\ d_j^{IB} &= \frac{1+\beta}{1+3\beta}\mathbb{E}_i\left[\theta_j\right] + \frac{2\beta}{1+3\beta}\theta_i, \\ T^{IB} &= \mathbb{E}_i\mathbb{E}_j \left[-(1-\beta)(d_j^{IB}-\theta_j)^2 - \beta(d_i^{IB}-d_j^{IB})^2\right] - \mathbb{E}_i\mathbb{E}_j \left[U_j^{ND}\right].\end{aligned}$$

The proposal is accepted because state j is indifferent between the offer and his outside option. □

We can now formally restate Proposition 6, incorporating explicit money burning functions:

Proposition 6. *In the equilibrium of the international bargaining game, states* 1 *and* 2 *burn the following amounts of money, respectively:*

$$\begin{aligned}b_1^{IB}(\theta_1) &= \frac{2(1-\beta)\beta\left[(1+\beta)+2((1-p)\beta^2+p\beta)\right]}{(1+\beta)^2(1+3\beta)}f_1(\theta_1), \\ b_2^{IB}(\theta_2) &= \frac{2(1-\beta)\beta\left[(1+\beta)+2(p\beta^2+(1-p)\beta)\right]}{(1+\beta)^2(1+3\beta)}f_2(\theta_2).\end{aligned}$$

Moreover, if $p = 1/2$, $b_i^{IB}(\theta_i) = b_i^D(\theta_i)$. For any value of p, $b_i^{IB}(\theta_i) > b_i^{ND}(\theta_i)$. Finally, since $\beta > \beta^2$, $b_i^{IB}(\theta_i)$ is increasing in state i's proposing probability.

Proof. Let us denote the expected utility that state i receives when it proposes as follows:

$$\begin{aligned}U_i^i \equiv & \mathbb{E}_i \left[-(1-\beta)(d_i^{IB}-\theta_i)^2 - \beta(d_i^{IB}-d_j^{IB})^2 + \mathbb{E}_j \left[-(1-\beta)(d_j^{IB}-\theta_j)^2 - \beta(d_i^{IB}-d_j^{IB})^2\right]\right] \\ & - \mathbb{E}_i\mathbb{E}_j \left[U_j^{ND}\right].\end{aligned}$$

Let us denote the expected utility that state i receives when state j proposes as follows:

$$\begin{aligned}U_i^j \equiv & \mathbb{E}_i \left[-(1-\beta)(d_i^{IB}-\theta_i)^2 - \beta(d_i^{IB}-d_j^{IB})^2\right] - \mathbb{E}_j\mathbb{E}_i \left[-(1-\beta)(d_i^{IB}-\theta_i)^2 - \beta(d_i^{IB}-d_j^{IB})^2\right] \\ & + \mathbb{E}_j\mathbb{E}_i \left[U_i^{ND}\right].\end{aligned}$$

From an ex-ante perspective, before knowing who is going to be the proposer, state i's payoff can be denoted as follows:

$$U^i \equiv pU_i^i + (1-p)U_i^j.$$

Suppose state i is of type θ_i, signals his type as θ_i', and believes that the other state is of type θ_j with probability one. Let $U^i(\theta_i, \theta_i', \theta_j)$ denote the ex-ante payoff U^i when these conditions hold.

Suppose that state i burns $b_i(\theta_i')$ in order to signal his type as θ_i'. Then, he obtains the following payoff:

$$U^i\left(\theta_i, \theta_i', \theta_j\right) - b_i(\theta_i').$$

A function $b_i(\theta_i)$ is incentive-compatible and fully reveals state i's type if the following condition holds:

$$\theta_i \in \arg\max_{\theta_i'} \mathbb{E}_i^0\left[U^i(\theta_i, \theta_i', \theta_j)\right] - b_i(\theta_i').$$

$b_i(\theta_i)$ is a strictly monotone function.

The first requirement implies that $\theta_i' = \theta_i$ satisfies the following first-order condition:

$$\frac{\partial b_i(\theta_i')}{\partial \theta_i'} = \mathbb{E}_i^0\left[\frac{\partial U^i(\theta_i, \theta_i', \theta_j)}{\partial \theta_i'}\right].$$

Integrating with respect to θ_i and considering the initial condition yields the following expression for each state:

$$b_1^{IB}(\theta_1) = \frac{2(1-\beta)\beta\left[(1+\beta) + 2((1-p)\beta^2 + p\beta)\right]}{(1+\beta)^2(1+3\beta)} f_1(\theta_1),$$

$$b_2^{IB}(\theta_2) = \frac{2(1-\beta)\beta\left[(1+\beta) + 2(p\beta^2 + (1-p)\beta)\right]}{(1+\beta)^2(1+3\beta)} f_2(\theta_2).$$

□

D Other Extensions: Discussions and Proofs

D.1 Coordination Sensitivity and Proposition 7

In this extension, we consider an alternative delegation game in which the IO proposes (d_1, d_2, T), which has to be accepted by both states for it to pass. If it is rejected, states play the no-delegation game. We let the IO's policy payoff be the following (except for transfers):

$$u_{IO}(d_1, d_2, \theta_1, \theta_2) = -\alpha\left(d_1 - d_2\right)^2.$$

The parameter $\alpha > 0$ measures the IO's coordination motive. We study how this affects states' signaling incentives. We obtain the following:

Proposition 7. *In equilibrium, $b_i^\alpha(\theta_i)$ is increasing in the IO's coordination motive α.*

Proof. Denote as $U_i^{ND} \equiv -(1-\beta)\left(d_i^{ND} - \theta_i\right)^2 - \beta\left(d_i^{ND} - d_j^{ND}\right)^2$ the outside option of state

i. The IO solves

$$\max_{d_i,d_j,T_i,T_j} \quad \mathbb{E}_{IO}\left[-\alpha(d_i-d_j)^2\right]+T_i+T_j$$

s.t.

$$\mathbb{E}_{IO}\mathbb{E}_i\left[-(1-\beta)(d_i-\theta_i)^2-\beta(d_i-d_j)^2\right]-T_i \geq \mathbb{E}_{IO}\mathbb{E}_i\left[U_i^{ND}\right]$$
$$\mathbb{E}_{IO}\mathbb{E}_j\left[-(1-\beta)(d_j-\theta_j)^2-\beta(d_i-d_j)^2\right]-T_j \geq \mathbb{E}_{IO}\mathbb{E}_j\left[U_j^{ND}\right].$$

In the optimum the restrictions are binding. The problem becomes:

$$\begin{aligned}\max_{d_i,d_j} \quad & \mathbb{E}_{IO}\left[-(1-\beta)((d_i-\theta_i)^2+(d_j-\theta_j)^2)-(\alpha+2\beta)(d_i-d_j)^2\right] \\ & -\mathbb{E}_{IO}\left[\mathbb{E}_i\left[U_i^{ND}\right]+\mathbb{E}_j\left[U_j^{ND}\right]\right],\end{aligned}$$

with

$$\begin{aligned}T_i &= \mathbb{E}_{IO}\mathbb{E}_i\left[-(1-\beta)(d_i-\theta_i)^2-\beta(d_i-d_j)^2\right]-\mathbb{E}_{IO}\mathbb{E}_i\left[U_i^{ND}\right] \\ T_j &= \mathbb{E}_{IO}\mathbb{E}_j\left[-(1-\beta)(d_j-\theta_j)^2-\beta(d_i-d_j)^2\right]-\mathbb{E}_{IO}\mathbb{E}_j\left[U_j^{ND}\right].\end{aligned}$$

The optimum is the following:

$$\begin{aligned}d_i^{IO} &= \frac{1+\alpha+\beta}{1+2\alpha+3\beta}\mathbb{E}_{IO}\left[\theta_i\right]+\frac{\alpha+2\beta}{1+2\alpha+3\beta}\mathbb{E}_{IO}\left[\theta_j\right], \\ d_j^{IO} &= \frac{1+\alpha}{1+2\alpha+3\beta}\mathbb{E}_{IO}\left[\theta_j\right]+\frac{\alpha+2\beta}{1+2\alpha+3\beta}\mathbb{E}_{IO}\left[\theta_i\right], \\ T_i^{IO} &= \mathbb{E}_{IO}\mathbb{E}_i\left[-(1-\beta)(d_i^{IO}-\theta_i)^2-\beta(d_i^{IO}-d_j^{IO})^2\right]-\mathbb{E}_{IO}\mathbb{E}_i\left[U_i^{ND}\right], \\ T_j^{IO} &= \mathbb{E}_{IO}\mathbb{E}_j\left[-(1-\beta)(d_j^{IO}-\theta_j)^2-\beta(d_i^{IO}-d_j^{IO})^2\right]-\mathbb{E}_{IO}\mathbb{E}_j\left[U_j^{ND}\right].\end{aligned}$$

The proposal is accepted by each state since both are indifferent between the offer and outside option. In this case

$$b_i^{\alpha}(\theta_i)=\frac{2(1-\beta)(\alpha+\beta+\beta^2+\alpha\beta^2+2\beta^3)}{(1+\beta)^2(1+2\alpha+3\beta)}f_i(\theta_i).$$

Note that $b_i^{\alpha}(\theta_i)$ is increasing in α. Also $b_i^{\alpha}(\theta_i) > b_i^{ND}(\theta_i)$. □

D.2 Limited Discretion and Proposition 8

How does limiting the IO's discretion affects states' gains from delegation? We assume states delegate a symmetric interval $[-\ell/2, \ell/2]$, which has length $\ell \geq 0$ in which discretion is parameterized by ℓ.[18] The IO is restricted to choose the same decision for both states $d = d_1 = d_2$.

The IO's ideal policy based on its beliefs is equal to $\hat{d}_{IO} \equiv \frac{1}{2}\left[\mathbb{E}_{IO}(\theta_1|b_1, m_1) + \mathbb{E}_{IO}(\theta_2|b_2, m_2)\right]$. If this ideal policy falls within the IO's delegation interval, then it is the outcome, otherwise the policy is its lower $(-\ell/2)$ or upper bound $(\ell/2)$.

[18]Without discretion ($\ell = 0$), the IO is forced to select $d = 0$, while if $\ell = \infty$, the IO has unlimited discretion.

Changing the IO's discretion not only alters decisions but also countries' signals. When the IO has relatively more discretion, signals have a greater influence on decisions, which increases incentives to burn money. In selecting the IO's level of discretion, there is a trade-off between getting decisions that are more tailored to countries' domestic circumstances, and incurring money burning costs to transmit information.

The results further emphasize our earlier findings about how IOs negatively impact signaling. With even more coordination after delegation, incentives to burn money are even stronger than in the baseline model. This makes it necessary to limit the IO's discretion to dampen money burning incentives. Further, we show that each state's most preferred length of the delegation interval increases in s because it increases the potential for both countries to have the same type $\theta_1 = \theta_2$. When type spaces do not overlap ($s < \Delta/2$), it is never optimal to give the IO any discretion. Also, as shown earlier, greater disagreement (a great disagreement between the states' types' expected value) makes money burning incentives stronger because there is more to gain from influencing the IO's decision, further increasing the benefits of limited discretion. Formally:

Proposition 8. *In each state's ex ante most preferred institution, the length of the delegation interval ℓ increases in the level of uncertainty s and decreases in the amount of disagreement Δ, where*

$$\ell\,(s, \Delta) = \begin{cases} 0 & \text{if } 0 \leq s \leq \sqrt{3}(\Delta/2), \\ \frac{s}{3} - \frac{\Delta^2}{4s} & \text{if } \sqrt{3}(\Delta/2) < s \leq \Delta. \end{cases}$$

Proof. We impose the restriction that $\ell \in [0, 2s]$, because when $\ell > 2s$, policies are the same as with $\ell = 2s$. The reason is that the highest and lowest policy the IO ever takes are

$$d^{max} = \frac{\max \theta_1 + \max \theta_2}{2} = \frac{-\Delta + s + \Delta + s}{2},$$
$$d^{min} = \frac{\min \theta_1 + \min \theta_2}{2} = \frac{-\Delta - s + \Delta - s}{2}.$$

The difference between the two is the set of policies that the IO will possibly take in equilibrium, which equals $\ell = d^{max} - d^{min} = 2s$. Further, taking expected values we obtain the following ex ante political payoff for both states:

$$\Pi^D = -\frac{1}{3}(1-\beta)\left(\frac{3}{4}\left(\ell^2 + \Delta^2\right) - \ell s + s^2\right).$$

The money burning functions are the following

$$b_1^D(\theta_1) = \frac{(1-\beta)\ell}{2s}\left[\theta_1^2 - \min\{-\Delta/2 + s, 0\}^2\right],$$
$$b_2^D(\theta_2) = \frac{(1-\beta)\ell}{2s}\left[\theta_2^2 - \max\{\Delta/2 - s, 0\}^2\right].$$

Taking an expectation leads to the following ex ante informational payoff for each state:

$$B^D = \frac{\ell}{6s}(1-\beta)\left(s^2 + 3(\Delta/2)^2 - 3\min\{-\Delta/2 + s, 0\}^2\right).$$

Consider the function $(\Pi^D - B^D)$. If $\ell \geq 2s$, then $(\Pi^D - B^D) < (\Pi^{ND} - B^{ND})$. If we optimize the expression $(\Pi^D - B^D)$ restricted to $\ell \in [0, 2s]$ we obtain:

$$\ell(s,\Delta) = \begin{cases} 0 & \text{if } 0 \leq s \leq \sqrt{3}(\Delta/2) \\ \frac{s}{3} - \frac{\Delta^2}{4s} & \text{if } \sqrt{3}(\Delta/2) < s \leq \Delta. \end{cases}$$

□

D.3 Heterogeneous Value of Coordination and Proposition 9

We now study how the gains from delegation depend on states' potentially heterogeneous values of coordination β. This extension serves to capture a situation with a large state that cares more about adjusting decisions to domestic conditions and a smaller state that cares more about coordinated decisions.

We analyze the extreme case when state 1 does not value coordination and has policy preferences of $\pi_1(d_1, \theta_1) = -(d_1 - \theta_1)^2$. State 2, however, still values coordination with weight $\beta_2 \in (0,1)$ and has preferences as in the main model. We assume preferences of the IO that are still a weighted average of both countries' interests:

$$u_{IO}(\theta_1, \theta_2) = \alpha\left[-(d_1 - \theta_1)^2\right] + (1-\alpha)\left[-(1-\beta_2)(d_2 - \theta_2)^2 - \beta_2(d_1 - d_2)^2\right].$$

In the benchmark we assume $\alpha = \frac{1}{2}$ and provide the following result.

Lemma 3. *In ex-ante terms, state 1 never prefers to delegate while state 2 always prefers to delegate. There exists an inverted u-shaped function $\tilde{s}(\beta)$ such that, if $s \leq \tilde{s}$, then delegation generates joint benefits.*

We prove this lemma below together with Proposition 11. This result implies that although state 1 would lose from delegation, state 2 gains more, and could compensate for state 1's loss by sending transfers as long as the level of uncertainty is sufficiently low. The reason for the non-monotonic effect becomes apparent by contrasting two extreme situations. If state 2 cares very little about coordination, with $\beta_2 \approx 0$, then the positive effects of the increase in coordination due to delegation is unlikely to outweigh the costs of money burning, even with little uncertainty. In the other extreme where state 2 finds coordination highly important, with $\beta_2 \approx 1$, state 2 is already willing to coordinate to a large extent with the other state, again implying that delegation is barely beneficial. Increased coordination is only valuable for intermediate values of β_2, and may lead to beneficial delegation for a wider range of uncertainty.

Another way to compensate state 1 is to alter the allocation of authority in the IO. We now study how the joint benefits from delegation can be maximized by selecting $\alpha \in [0,1]$, which is state 1's weight in the IO. An increase in α grants state 1 more authority, shifting decisions in 1's favor, and also affects the signals that countries send. Proposition 9 establishes our result. The results depend crucially on the amount of uncertainty, s, and the importance that the small state places on coordination, β_2. If the goal of the IO is to generate the largest amount of ex-ante joint benefits, then the share of authority by state 1 is increasing in the level of uncertainty.

There are two factors that affect the total gains from delegation. First, each state's payoffs that are determined by equilibrium decisions. Given that institutions that maximize the total gains for both countries weigh the welfare of them both equally, it implies that if there is no uncertainty, countries should have equal authority. This guarantees that decisions are taken that weigh both countries' interests equally. The second factor is informational welfare. state 1 never burns money because it has no interest in changing state 2's behavior. As a result, the only costly signals are sent by state 2. With more uncertainty, this part affects the gains and losses from delegation the most, and by giving state 1 more authority, state 2 knows that its signals have less influence on decisions, reducing incentives to send costly signals. With too much uncertainty, it is optimal to give all authority to state 1. Formally:

Proposition 9. *The institution that maximizes ex-ante joint benefits always gives weakly more authority to state* 1 *with* $\alpha \geq \frac{1}{2}$. *Further, there exists a function* $s^*(\beta_2) < 1$ *such that if there is more uncertainty than* $s^*(\beta_2)$, *then all authority is in the hands of state* 1 $(\alpha = 1)$. *If there is less uncertainty than* $s^*(\beta_2)$, *then* α *is increasing in the amount of uncertainty.*

Proof. We calculate equilibrium payoffs as a function of α and then we study the case $\alpha = 1/2$. Now, policy payoffs are as follows:

$$\pi_1(d_1, d_2, \theta_1) = -(d_1 - \theta_1)^2,$$
$$\pi_2(d_2, d_1, \theta_2) = -(1-\beta_2)(d_2 - \theta_2)^2 - \beta_2 (d_2 - d_1)^2.$$

In the case of delegation the IO maximizes the following

$$u_{IO}(d_1, d_2, \theta_1, \theta_2) = \alpha\left[-(d_1 - \theta_1)^2\right] + (1-\alpha)\left[-(1-\beta_2)(d_2-\theta_2)^2 - \beta_2(d_2 - d_1)^2\right].$$

We need to obtain ex ante political and informational payoffs for both cases.

Without delegation, states take the following decisions as a function of θ_1 and θ_2

$$d_1^{ND} = \theta_1,$$
$$d_2^{ND} = \beta_2\theta_1 + (1-\beta_2)\theta_2.$$

Thus, ex ante political payoffs are as follows

$$\Pi_1^{ND} = 0,$$
$$\Pi_2^{ND} = -\frac{2}{3}(1-\beta_2)\beta_2(6+s^2).$$

Finally, states have no incentives to burn money since state 1 does not benefit from manipulation and state 2 can not influence. Thus

$$B_1^{ND} = B_2^{ND} = 0.$$

With delegation, the IO chooses the following decisions as a function of θ_1 and θ_2:

$$d_1^D = \frac{\theta_1\alpha + \theta_2(1-\alpha)(1-\beta_2)\beta_2}{\alpha + \beta_2 - \alpha\beta_2 - \beta_2^2 + \alpha\beta_2^2},$$
$$d_2^D = \frac{\theta_1\alpha\beta_2 + \theta_2(1-\beta_2)(\beta_2 + \alpha(1-\beta_2))}{\alpha + \beta_2 - \alpha\beta_2 - \beta_2^2 + \alpha\beta_2^2}.$$

Ex ante political payoffs are as follows

$$\Pi_1^D = -\frac{2}{3}\frac{(1-\alpha)^2(1-\beta_2)^2\beta_2^2(6+s^2)}{(\alpha + \beta_2 - \alpha\beta_2 - (1-\alpha)\beta_2^2)^2},$$
$$\Pi_2^D = -\frac{2}{3}\frac{\alpha^2(1-\beta_2)\beta_2(6+s^2)}{(\alpha + \beta_2 - \alpha\beta_2 - (1-\alpha)\beta_2^2)^2}.$$

Money burning functions

$$b_1^D(\theta_1) = \frac{2(1-\alpha)\alpha(1-\beta_2)\beta_2}{(\alpha + \beta_2 - \alpha\beta_2 - \beta_2^2 + \alpha\beta_2^2)^2}\left(\theta_1 - \min\{\overline{\theta}_1, 1\}\right)\left(\frac{\theta_1 + \min\{\overline{\theta}_1, 1\}}{2} - 1\right),$$

$$b_2^D(\theta_2) = \frac{2(1-\alpha)\alpha(1-\beta_2)^2\beta_2^2}{(\alpha + \beta_2 - \alpha\beta_2 - \beta_2^2 + \alpha\beta_2^2)^2}\left(\theta_2 - \max\{\underline{\theta}_2, -1\}\right)\left(\frac{\theta_2 + \max\{\underline{\theta}_2, -1\}}{2} + 1\right).$$

Then, ex ante informational payoff

$$B_1^D = \frac{2}{3}\frac{(1-\alpha)\alpha(1-\beta_2)\beta_2(6-s)s}{(\alpha + \beta_2 - \alpha\beta_2 - (1-\alpha)\beta_2^2)^2},$$

$$B_2^D = \frac{2}{3}\frac{(1-\alpha)\alpha(1-\beta_2)^2\beta_2^2(6-s)s}{(\alpha + \beta_2 - \alpha\beta_2 - (1-\alpha)\beta_2^2)^2}.$$

The rest of the proof assume $\alpha = 1/2$. After some algebra, we obtain $(\Pi_1^{ND} - B_1^{ND}) > (\Pi_1^D - B_1^D)$ and $(\Pi_2^{ND} - B_2^{ND}) < (\Pi_2^D - B_2^D)$. Thus state 1 prefers not to delegate while state 2 prefers to delegate. If we consider instead

$$\left(\Pi_1^{ND} - B_1^{ND} + \Pi_2^{ND} - B_2^{ND}\right) - \left(\Pi_1^D - B_1^D + \Pi_2^D - B_2^D\right),$$

We obtain that there is $\tilde{s}$ such that

If $s \leq \tilde{s}$, then $\left(\Pi_1^{ND} - B_1^{ND} + \Pi_2^{ND} - B_2^{ND}\right) \leq \left(\Pi_1^D - B_1^D + \Pi_2^D - B_2^D\right)$,

If $s > \tilde{s}$, then $\left(\Pi_1^{ND} - B_1^{ND} + \Pi_2^{ND} - B_2^{ND}\right) > \left(\Pi_1^D - B_1^D + \Pi_2^D - B_2^D\right)$.

The cutoff $\tilde{s}$ is the following

$$\tilde{s} \equiv \frac{3 - (9 - 6\beta_2 + 12\beta_2^3 - 6\beta_2^4)^{\frac{1}{2}}}{1 + \beta_2 - \beta_2^2}.$$

We now use the previous results and study the case of general α. Consider $(\Pi_1^D - B_1^D + \Pi_2^D - B_2^D)$ as a function of α. Define

$$\alpha(\beta, s) \equiv \frac{-(1-\beta)\beta(s^2(\beta^2 - \beta - 3) + s(-6\beta^2 + 6\beta + 6) - 12)}{s(s-6) - 2\beta^3 s(s-6) + \beta^4 s(s-6) - \beta^2(5s^2 - 6s + 24) + 6\beta(s^2 - 2s + 4)},$$

and

$$s^* \equiv \frac{3 + 3\beta_2 - 3\beta_2^2 - (9 + 6\beta_2 - 33\beta_2^2 + 54\beta_2^3 - 27\beta_2^4)^{1/2}}{1 + 3\beta_2 - 3\beta_2^2}.$$

Denote $\hat{\alpha}$ the maximizer of $\left(\Pi_1^D - B_1^D + \Pi_2^D - B_2^D\right)$ restricted to $0 \leq \alpha \leq 1$. A simple first-order condition analysis implies the following

If $0 < s \leq s^*$, then $\hat{\alpha} = \alpha\left(\beta_2, s\right)$,

If $s^* < s$, then $\hat{\alpha} = 1$.

It is direct to see that $\frac{\partial \alpha(\beta_2, s)}{\partial s} > 0$ and $\alpha\left(\beta_2, 0\right) = \frac{1}{2}$.

□

D.4 One-sided Incomplete Information and Proposition 10

To investigate the role of asymmetric uncertainty, we study an extreme version where state 1's type is known while state 2's type is drawn as in the main model. We show that state 1 always prefers to delegate as it has no signaling cost, while state 2 would only prefer delegation under the same conditions as in the baseline model. Hence, our results are robust to the introduction of asymmetries in terms of states' domestic conditions.

Proposition 10. *Delegation is ex-ante jointly beneficial if and only if the level of uncertainty is sufficiently low such that $s < \check{s}(\beta)$, where $\check{s}(\beta) > \hat{s}(\beta)$.*

Proof. Formally, we assume state 1's type is publicly observable. Since state 1 can not influence beliefs through its signals, it does not burn money. Political payoffs are the same as in the previous results.

Without delegation, the money burning functions are the following

$$\begin{aligned} b_1^{ND}(\theta_1) &= 0, \\ b_2^{ND}(\theta_2) &= \frac{2(1-\beta)\beta^2}{(1+\beta)^2}\left(\theta_2 - \max\{1-s, \theta_1\}\right)\left(\frac{\theta_2 + \max\{1-s, \theta_1\}}{2} - \theta_1\right). \end{aligned}$$

And then

$$B_2^{ND} = \begin{cases} \frac{2}{3}\frac{(1-\beta)\beta^2(s^3+6s-2)}{(1+\beta)^2 s} & \text{if } s \geq 1 \\ \frac{2}{3}\frac{(1-\beta)\beta^2(6-s)s}{(1+\beta)^2} & \text{if } s < 1, \end{cases}$$

With delegation, the money burning functions are the following

$$b_1^D(\theta_1) = 0,$$
$$b_2^D(\theta_2) = \frac{2(1-\beta)\beta}{(1+3\beta)}(\theta_2 - \max\{1-s,\theta_1\})\left(\frac{\theta_2 + \max\{1-s,\theta_1\}}{2} - \theta_1\right).$$

And then

$$B_2^{D} = \begin{cases} \frac{2}{3}\frac{(1-\beta)\beta(s^3+6s-2)}{(1+3\beta)s} & \text{if } s \geq 1 \\ \frac{2}{3}\frac{(1-\beta)\beta(6-s)s}{(1+3\beta)} & \text{if } s < 1. \end{cases}$$

We have that $B_2^{ND} < B_2^D$. Comparing terms, we obtain that $\Pi_1^D > \Pi_1^{ND}$, thus state 1 always prefers to delegate. In the other side, after some algebra we obtain

If $s \leq \hat{s}$, then $\left(\Pi_2^{ND} - B_2^{ND}\right) \leq \left(\Pi_2^D - B_2^D\right)$,

If $s > \hat{s}$, then $\left(\Pi_2^{ND} - B_2^{ND}\right) > \left(\Pi_2^D - B_2^D\right)$.

If we consider $\left(\Pi_1^{ND} + \Pi_2^{ND} - B_2^{ND}\right) - \left(\Pi_1^D + \Pi_2^D - B_2^D\right)$, there is $\check{s}$ with $\check{s} > \hat{s}$ such that:

If $s \leq \check{s}$, then $\left(\Pi_1^{ND} + \Pi_2^{ND} - B_2^{ND}\right) \leq \left(\Pi_1^D + \Pi_2^D - B_2^D\right)$,

If $s > \check{s}$, then $\left(\Pi_1^{ND} + \Pi_2^{ND} - B_2^{ND}\right) > \left(\Pi_1^D + \Pi_2^D - B_2^D\right)$.

$$s < \check{s} \equiv \frac{\left(3 + 6\beta - (9 + 24\beta - 12\beta^2)^{\frac{1}{2}}\right)}{(1+4\beta)}.$$

□