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

Online Appendix

Table of Contents

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 (b_j, m_j) 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 (b_i, m_i, b_j, m_j). 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 $P=\Theta_i$. □

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:

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

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

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

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.

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

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:

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

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

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

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 $\mathbb{E}_i^0\left[\theta_j\right]$. This can be observed by noting that $\frac{\partial \tilde{\pi}_i\left(\theta_i,\theta_i',\theta_j\right)}{\partial \theta_i'}$ is a linear function of θ_j. Hence,

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

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 (Π^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 $[(6-s)s-3\beta\left(s^2-4s+2\right)]$ 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 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 $\theta_2 \leq \hat{\theta}_2$, 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 $\mathbb{E}_2^0[\Delta u] \leq 0$ 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 $\Pr\left(\theta_2 \leq \hat{\theta}_2\right) = F_s\left(\hat{\theta}_2\right)$ 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 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:

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

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. □

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 deviates 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 does 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 $c \geq \bar{c}$, both states do not deviate for any possible types. The IO will propose its most preferred policies, which are derived in the delegation game, and these policies are 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 $c \geq \bar{c}$, 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

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

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 1. *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:

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

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:

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

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

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

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, it obtains the following payoff:

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

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

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

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 1. *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

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

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.[1] 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 $(\ell/2)$.

[1] 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 2. *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 possibly takes 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 2. *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 9. 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 3 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 3. *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:

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

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

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 $\hat{\alpha} = \alpha(\beta_2, s)$,

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

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

□

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 4. *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

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

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

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

And then

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

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

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

□

References

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

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.

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

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

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