A Theory of Firm Boundaries and Knowledge Sharing*

Evgenii Fadeev[†] Michael Powell[‡]

April 6, 2026

The latest version is here

Abstract

We study asset ownership as an instrument of knowledge protection in collaborations where production requires sharing private knowledge but exposes parties to non-contractible expropriation. Ownership confers control over production and the ability to expropriate. It can protect the owner's knowledge by reorganizing production so that certain parties need not be informed, or by ensuring that some parties who must be informed lack the means to expropriate — allowing information to be shared safely. However, because ownership also gives the owner the ability to expropriate, other parties may withhold their knowledge from the owner. We apply the theory to explain the role of third-party intermediaries that facilitate knowledge sharing among competitors, and to determine when a vertically integrated firm must spin off a subsidiary to encourage knowledge sharing from rivals versus when internal "Chinese Walls" suffice.

*We are grateful for comments from Charles Angelucci, Ashish Arora, Daniel Barron, Wouter Dessein, Bob Gibbons, Oliver Hart, Andrew Koh, Niko Matouschek, Andrea Prat, Andrei Shleifer, Alexander Wolitzky, and participants at various conferences and seminars.

[†]Duke University, Fuqua School of Business, email: evgenii.fadeev@duke.edu.

[‡]Northwestern University, Kellogg School of Management, email: mike-powell@kellogg.northwestern.edu.

1 Introduction

Firm boundaries are often drawn to protect proprietary knowledge. This view is reflected in the management literature (Teece, 1980; Liebeskind, 1996) and among practitioners.[1] It is also consistent with empirical work arguing that a primary purpose of ownership is to facilitate knowledge transfers (Atalay, Hortaçsu, and Syverson, 2014; Demirer and Karaduman, 2025). This paper studies when and why firm boundaries shape how knowledge is transferred and used.

Our starting point is a fundamental property of knowledge, emphasized by Arrow (1962), that distinguishes it from ordinary inputs. Once knowledge is shared with another party, it becomes difficult to contractually restrict how the recipient uses it. In particular, the recipient may use it for private gain at the expense of the original knowledge holder. We refer to such non-contractible follow-on use as knowledge expropriation. Following the property-rights tradition (Grossman and Hart, 1986; Hart and Moore, 1990; Hart, 1995), we study firm boundaries through the allocation of ownership of productive assets, and hence the residual control rights associated with them. We argue that ownership protects knowledge from expropriation through two related but distinct channels.

To illustrate these two channels, consider two variants of a stylized example. In each, two automakers collaborate on a joint car model. Each has a secret engineering technique, and if combined, they can produce a superior vehicle. Their techniques can be freely shared, but disclosure creates expropriation risk: once one automaker learns the other's secret, it can use that knowledge outside the collaboration to improve other models in its own product line (Bai et al., 2025). Suppose also that production requires a specialized manufacturing plant.

Variant 1. Suppose ownership confers control over production. The jointly developed car must be produced by a single automaker that owns and operates the plant, so only the owner can take the key productive actions. Expropriation, however, does not require access to the plant. The owner's knowledge is therefore protected because it can be used without being disclosed, whereas the non-owner's knowledge is exposed to expropriation because it must be revealed to the owner for the collaboration to succeed. Anticipating expropriation, the non-owner may be unwilling to disclose its knowledge fully. Ownership of the plant therefore determines whose knowledge can be used directly and whose knowledge is partially withheld.

Variant 2 differs from Variant 1 in two respects. First, while the owner still takes productive decisions that require the plant, both parties must now be involved in production, so mutual disclosure of both techniques is required for the collaboration to succeed. Second, expropriation now requires control over the plant and is thus possible only for the owner. The owner's knowledge is therefore protected because it can be safely disclosed to the non-owner without

[1] For example, Steve Jobs argued that Apple's decision to design the iPad chip in-house rather than collaborate with Intel was driven by concern that Intel could use Apple's knowledge to serve competitors (Isaacson, 2011).

expropriation risk. The non-owner's knowledge, by contrast, remains exposed to expropriation once disclosed to the owner. Anticipating expropriation, the non-owner may be unwilling to disclose its knowledge fully. Ownership of the plant therefore determines whose knowledge can be safely transferred and whose knowledge is partially withheld.

These examples illustrate two broader lessons. First, because knowledge expropriation is non-contractible, ownership matters: it determines who can apply knowledge without disclosing it and who can disclose knowledge without fear of expropriation. Second, ownership is a blunt instrument for achieving the efficient use of knowledge because it bundles multiple inextricable decision rights into a single asset. In Variant 1, if productive control over each party's knowledge could be allocated separately, each party could use its own knowledge without disclosing it. In Variant 2, if expropriation control could be separated from productive control, it could be assigned to a third party outside production. Either separation would eliminate expropriation risk and allow the efficient use of knowledge. Instead, in both cases, a single owner must receive the non-owner's knowledge and can expropriate it.

The same logic extends beyond bilateral collaborations to richer interfirm arrangements. In practice, competing firms often share knowledge through common partners such as suppliers rather than directly with one another (Küpper et al., 2020). Whenever they do so, how those partners are organized becomes an important issue. For instance, automakers were reluctant to share sensitive information with the auto supplier Delphi Automotive when it was a subsidiary of their competitor, General Motors, prompting GM to spin it off.[2] Rather than spinning off the subsidiary, firms sometimes argue that internal informational "Chinese Walls" between a parent and a subsidiary can serve a similar purpose, assuring competitors that their knowledge will not reach the parent (Gawer and Cusumano, 2002; Martin and Mickle, 2017).

In our model, multiple parties collaborate on projects and can take two kinds of non-contractible actions: cooperative actions, which contribute to project success, and expropriation actions, which use another party's knowledge for private gain. Projects exhibit O-ring complementarities (Kremer, 1993): a project succeeds only if all required cooperative actions are correctly matched to privately observed states. To match actions to the relevant states, parties may need to receive others' information. But receiving such information also enables knowledge expropriation. There is a single productive asset whose ownership confers control over asset-dependent cooperative and expropriation actions, although the owner is not necessarily the only party required for production or capable of expropriation.

Our goal is to isolate how the non-contractibility of actions shapes the optimal ownership

[2]See CNN Money (1999): "Delphi... said the spinoff will allow the company to pursue supply contracts with other automakers, which previously were reluctant to use Delphi because it was a unit of GM". Bradsher (1998) notes that rivals avoided Delphi "for fear that it would share parts designs and other competitive information with G.M."

structure. To do this, we abstract from the strategic sharing of private information and instead evaluate collaboration from the top down. For a fixed ownership structure, we ask what outcomes can be sustained when an omniscient mediator allocates information across parties, anticipating that they will act in their own self-interest. Following Bergemann and Morris (2019), we formalize this by characterizing feasible outcomes as those induced by a decision rule mapping states into private action recommendations. Because actions are non-contractible, these recommendations must be *obedient*: each party must weakly prefer to follow its recommendation given the information it conveys. We focus on feasible outcomes that maximize total surplus.

Our main analysis focuses on the case in which expropriation is sufficiently costly that optimal disclosure deters it. In that case, obedience requires both following recommended cooperative actions and refraining from expropriation. The resulting no-expropriation constraints can be summarized by safety thresholds. To define them, note that each project requires a collection of tasks, where a task specifies that one party must match some party's state. When an acting party is capable of expropriation, deterring it places an upper bound on how precisely information about the target's state can be disclosed to the acting party. We call this upper bound the task's safety threshold. The safety threshold is strictly decreasing in the acting party's expected private gain from expropriation.

Our first main result characterizes optimal information disclosure for any fixed ownership structure. For each project, rank the required tasks by their safety thresholds. An optimal decision rule is nested: whenever a low-threshold ("hard") task is successfully matched, all higher-threshold ("easier") tasks are successfully matched as well. This nesting is optimal because tasks are complementary and precision is limited by the no-expropriation constraints. Additional precision on an easier task is valuable only in states where the harder tasks also succeed. Optimal disclosure therefore aligns precision across tasks. As a result, each project is constrained by an information bottleneck: the required task with the lowest safety threshold.

This characterization implies that ownership matters only through its effect on project bottlenecks. Changing ownership changes a project's bottleneck by altering who must be informed for production and who can expropriate once informed. Our second main result shows that optimal ownership structures maximize the value-weighted sum of bottleneck safety thresholds across projects.

One implication of this result is that asset ownership can be an imperfect instrument when it bundles control over multiple tasks required for production. If the asset confers control over at most one such task, ownership can be assigned so that the party performing that task already knows the relevant state, and no disclosure about this task is needed. If the asset confers control over two or more such tasks that depend on different parties' information, the owner must rely on another party's information in order to produce. Because the owner can also expropriate that information, it cannot be fully informed without creating expropriation risk. Some knowledge must therefore be withheld, and inefficiency is unavoidable.

We apply these results to explain several prominent industry configurations. First, we show how common suppliers and other intermediaries can act as knowledge hubs: when direct sharing among competitors would make a rival with strong expropriation incentives the bottleneck, routing information through a supplier that is safer to inform improves cooperation.

Second, we analyze a setting with two downstream competitors and a common upstream supplier. Each downstream firm has its own project, and the supplier provides an input to both. The supplier can expropriate only if it owns the asset, so assigning the asset to one downstream firm removes the supplier's expropriation capability and makes it that firm's subsidiary. The parent can then safely disclose information to its subsidiary, but the rival remains exposed to expropriation by the parent. Whether this risk can be avoided depends on whether the asset bundles decision rights required in both projects. When both projects require the asset, the parent must be informed about the rival's knowledge and can also expropriate it, so inefficiency is unavoidable. In that case, the organization faces a trade-off between protecting the parent's knowledge and facilitating the rival's project; if the latter is sufficiently valuable, it is optimal to spin off the supplier, as in GM's spin-off of Delphi. When only one project requires the asset, however, the firm associated with that project can own the asset without needing to learn the rival's knowledge. The subsidiary can then serve both sides while the parent remains uninformed — an arrangement resembling an internal Chinese Wall.

We conclude by considering several variations of the model that clarify our main results. First, we show that the optimal outcome under the information-design approach can be implemented as an equilibrium of a broader contracting game. In this game, the contract is a *mediated mechanism* specifying how transfers, ownership, and communication depend on parties' messages to a neutral *uninformed* mediator. This result clarifies that the central friction in our setting is the non-contractibility of actions rather than private information.

Second, we show that, among the various non-contractible actions in our setting, it is the *non-contractibility of expropriation* that makes ownership matter for knowledge sharing. If parties could sign a binding no-expropriation agreement before disclosing knowledge, they could achieve the first best regardless of ownership. By contrast, contracting over cooperative actions or the division of collaborative profits does not resolve the underlying incentive problem.

Third, we introduce an additional friction: parties do not have access to a neutral mediator and instead communicate directly via cheap talk. We show that parties can still limit how much information is disclosed and thereby deter expropriation. However, unlike a mediator, independent parties generally cannot coordinate disclosure so that "easier" tasks always succeed whenever "hard" tasks succeed. As a result, the logic of optimal ownership remains unchanged, but total surplus may be lower in environments where efficient cooperation requires coordinated disclosure. This result clarifies when parties benefit from mediated communication (e.g., the knowledge-hub setting) and when they do not need it (e.g., spin-off or Chinese Wall settings).

Fourth, we consider an extension in which control rights can be separated across multiple assets rather than bundled in a single asset. We show that, with sufficient separability of control rights, parties can achieve the first best. This result clarifies why ownership often fails to substitute for complete contracts: it confers "too many" *inextricable* control rights.

Finally, we discuss the role of expropriation costs. In the main analysis, expropriation is sufficiently costly that optimal information disclosure deters it: temptation to expropriate shapes disclosure, but expropriation costs are never incurred. When expropriation is costly but not prohibitive, it can be optimal to tolerate some expropriation to enable greater disclosure and raise the probability of project success. As a result, optimal ownership depends not only on the temptation to expropriate but also on the level of expropriation costs. We characterize the resulting information-design problem via a connection to optimal transport theory.

Related Literature. We build on the incomplete-contracts approach to firm boundaries (Grossman and Hart, 1986; Hart and Moore, 1990; Hart, 1995). Grossman and Hart (1986) define ownership as the allocation of residual rights of control: when specific rights are too costly to enumerate contractually, ownership assigns the rights left unspecified. Much of this literature focuses on the allocation of control. We follow it in this respect but also emphasize the *residual* nature of ownership: it typically transfers *multiple* decision rights as a bundle.

Within the firm-boundaries literature, our paper is closest methodologically to work that studies ownership structures through a mechanism-design approach (Klemperer, Cramton, and Gibbons, 1987; Matouschek, 2004; Segal and Whinston, 2016; Baliga and Sjöström, 2018). Whereas these papers emphasize how ownership affects information rents when ex post actions are contractible, we study optimal ownership when ex post actions are non-contractible. The central issue is therefore about obedience rather than truth-telling.

To characterize the optimal information-sharing and ownership arrangement in this environment, we use modern information-design tools (Rayo and Segal, 2010; Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019; Mathevet, Perego, and Taneva, 2020). This approach to study organizational-design questions has also been taken by Kolotilin and Li (2021), who study relational communication in repeated relationships with voluntary transfers and show that selective pooling can be used to manage obedience constraints. All these papers take control structures as given.

A related literature studies how information structures and mediated recommendations shape cooperation in cartels. Sugaya and Wolitzky (2018) show that coarser observability can make collusion easier to sustain because finer information lets firms tailor profitable deviations to market conditions. Marshall and Marx (2007) and Ortner, Sugaya, and Wolitzky (2024) study mediated arrangements that help parties collude, in part through private recommendations and sometimes through correlated randomization. In our paper, related instruments serve as building blocks for a theory of the firm.

Finally, we also connect to the literature on communication and delegation in organizations (Dessein, 2002; Alonso, Dessein, and Matouschek, 2008; Rantakari, 2008), which studies how the allocation of control shapes communication. Because ours is a paper on firm boundaries rather than internal organization, we emphasize two features of the environment that delegation models typically abstract from: parties have access to rich contracting instruments (i.e., they can contract jointly on decision rights, information-sharing protocols, and transfers), and the relevant choice is ownership, which allocates a bundle of inextricable control rights.

2 Model

There are N parties, $i \in \mathcal{N} = \{1, \ldots, N\}$, who collaborate on a collection of projects. Each project succeeds only if a set of required actions is correctly tailored to private information held by the parties. Sharing information improves cooperation but creates a risk: parties may use what they learn to expropriate others' knowledge for their private benefit.

Information. The state is $\theta = (\theta_1, \ldots, \theta_N)$, where $\theta_i \in \{-1, 1\}$ is privately observed by party i. The components of θ are independent, and each realization is equally likely, $p(\theta) = \left(\frac{1}{2}\right)^N$.

Actions and Asset Ownership. Each party $i \in \mathcal{N}$ chooses a vector of actions. We distinguish actions by whether they are (i) *cooperative* or *expropriation*, and (ii) *alienable* or *inalienable*. Cooperative actions are productive: when correctly matched to the relevant component of the state, they contribute to project success. Expropriation actions are privately beneficial but socially destructive: matching them to others' information generates a private gain for the acting party while imposing a cost on others. Alienable actions require access to a productive asset, whereas inalienable actions reflect specific human capital and require no such asset.

There is a single productive asset owned by party $g \in \mathcal{N}$. We model control over alienable actions in the following way. Each party $i \in \mathcal{N}$ selects a plan for *alienable cooperative actions*, denoted $x_{ij} \in \{-1, 1\}$ for all $j \in \mathcal{N}$, and for *alienable expropriation actions*, denoted $z_{ij} \in \{-1, 0, 1\}$ for all $j \neq i$. While every party specifies a plan, only the choices of the asset owner g can be payoff-relevant.

Each party $i \in \mathcal{N}$ also chooses *inalienable cooperation actions*, denoted by $\bar{x}_{ij} \in \{-1, 1\}$ for all $j \in \mathcal{N}$, and *inalienable expropriation actions*, denoted by $\bar{z}_{ij} \in \{-1, 0, 1\}$ for $j \neq i$.

For each party $i \in \mathcal{N}$, let $a_i = ((x_{ij})_{j\in\mathcal{N}}, (\bar{x}_{ij})_{j\in\mathcal{N}}, (z_{ij})_{j\neq i}, (\bar{z}_{ij})_{j\neq i})$ denote party i's *action plan*. Let $\mathcal{A}_i$ denote the set of possible action plans available to party i and write $\mathcal{A} = \prod_{i\in\mathcal{N}} \mathcal{A}_i$ and $a = (a_i)_{i\in\mathcal{N}} \in \mathcal{A}$. Equivalently, we may write $a = (x, \bar{x}, z, \bar{z})$.

Some parties can expropriate only if they own the asset, while others can expropriate even if they do not. Let $\mathcal{E} \subseteq \mathcal{N}$ denote the set of parties whose expropriation opportunities are tied to the asset. We define party i's effective expropriation action against party j as ζ_{ij}. For parties who can intrinsically expropriate ($i \notin \mathcal{E}$), $\zeta_{ij} = \bar{z}_{ij}$. For parties requiring the asset ($i \in \mathcal{E}$), $\zeta_{ij} = z_{ij}$ for the asset owner ($g = i$), and $\zeta_{ij} = 0$ for non-owners ($g \neq i$).

Projects. Production is organized into a set of projects $\mathcal{K}$. Each project $k \in \mathcal{K}$ has value $V_k \geqslant 0$ if successfully implemented and value 0 otherwise. Project success requires that a set of cooperative actions be correctly matched to the relevant components of the state.

Formally, each project k has two sets of requirements. First, $\mathcal{R}_k \subseteq \mathcal{N}$ is the set of alienable requirements: for each $j \in \mathcal{R}_k$, the asset owner g must set $x_{gj} = \theta_j$. Second, $\bar{\mathcal{R}}_k \subseteq \mathcal{N} \times \mathcal{N}$ is the set of inalienable requirements: for each $(i, j) \in \bar{\mathcal{R}}_k$, party i must choose $\bar{x}_{ij} = \theta_j$. Given asset owner g, define the project-success indicator $w_k \in \{0, 1\}$ by

$$w_k\left(x, \bar{x}, g; \theta\right) = \left(\prod_{j\in\mathcal{R}_k} \mathbf{1}\{x_{gj} = \theta_j\}\right) \cdot \left(\prod_{(i,j)\in\bar{\mathcal{R}}_k} \mathbf{1}\{\bar{x}_{ij} = \theta_j\}\right).$$

Project k succeeds if and only if all of its alienable and inalienable requirements are met.

Payoffs. Party i's payoff is given by $u_i(a, g; \theta) + t_i$, where t_i is a monetary transfer and

$$u_i(a, g; \theta) = u_i^c(x, \bar{x}, g; \theta) + u_i^e(z, \bar{z}, g; \theta).$$

The term u_i^c represents the *cooperative payoff* from project success, and u_i^e represents the *expropriation payoff*. Transfers satisfy the budget-balance condition, $\sum_{i\in\mathcal{N}} t_i = 0$. We next describe the cooperative and expropriation payoff terms.

The cooperative payoff depends only on which projects succeed. We assume party i receives an exogenous share $\alpha_{ik} \geqslant 0$ of project k's value, with $\sum_{i\in\mathcal{N}} \alpha_{ik} = 1$. Its cooperative payoff is

$$u_i^c(x, \bar{x}, g; \theta) = \sum_{k\in\mathcal{K}} \alpha_{ik} V_k w_k(x, \bar{x}, g; \theta).$$

Thus, cooperative payoffs depend on actions only through project success and do not depend directly on asset ownership beyond its effect on which actions are payoff-relevant.

Expropriation payoffs depend on the effective expropriation actions ζ_{ij}. If party i attempts to expropriate party j's knowledge, and the attempt is successful (i.e., $\zeta_{ij} = \theta_j$), then party

i obtains a private benefit $b_{ij} > 0$, while party j incurs a loss $\lambda_{ij} > 0$. If the attempt is unsuccessful (i.e., $\zeta_{ij} = -\theta_j$), party i incurs a cost $c_{ij} > 0$. If $\zeta_{ij} = 0$, both parties receive 0. Party i's expropriation payoff is therefore

$$u_i^e(z, \bar{z}, g; \theta) = \sum_{j \neq i} [b_{ij} \mathbf{1}\{\zeta_{ij} = \theta_j\} - c_{ij} \mathbf{1}\{\zeta_{ij} = -\theta_j\}] - \sum_{j \neq i} \lambda_{ji} \mathbf{1}\{\zeta_{ji} = \theta_i\},$$

where $\zeta_{ij} = \zeta_{ij}(z_{ij}, \bar{z}_{ij}; g)$ is the effective expropriation action.

We define total surplus as $TS(a, g; \theta) := \sum_{i \in \mathcal{N}} u_i(a, g; \theta)$. Because transfers are budget-balanced, they do not directly enter this expression.

We impose the following two assumptions on expropriation payoffs:

Assumption 1. *For each party $i \in \mathcal{N}$ and target $j \neq i$, $\lambda_{ij} - b_{ij} > V_{\text{total}} := \sum_{k \in \mathcal{K}} V_k$.*

Assumption 2. *For each party $i \in \mathcal{N}$ and target $j \neq i$, $b_{ij} < c_{ij}$.*

The first assumption ensures any successful expropriation destroys more total surplus than can be generated by all projects. The second assumption ensures expropriation is risky and thus attractive only when a party has sufficiently precise information about the target's state.

Imperfect Appropriability. Following Arrow (1962), we assume that once knowledge is shared, its subsequent use is difficult to restrict contractually. We capture this friction by assuming that actions are *non-contractible*.

Solution Concept. Fix an owner $g \in \mathcal{N}$. Consider an *omniscient* mediator who decides how information is disclosed to the parties. A *decision rule* is a mapping $\sigma(\cdot \mid \theta) \in \Delta(\mathcal{A})$ for each $\theta \in \Theta$, which assigns to each realized state θ a distribution over action profiles $a = (a_i)_{i \in \mathcal{N}}$.

The timing is as follows: (i) the mediator commits to a decision rule σ; (ii) the state θ is realized; (iii) each party i observes θ_i; (iv) the mediator draws an action profile a according to $\sigma(\cdot \mid \theta)$ and privately recommends a_i to each party i; (v) the parties choose their actions; and (vi) payoffs are realized.

Recommendations are not enforceable. Instead, a decision rule must be designed so that parties optimally choose to follow their recommendations given their information. Accordingly, we restrict attention to *obedient* decision rules, under which no party wishes to deviate from the recommended action after observing its private recommendation and private state. We denote the set of obedient decision rules by Σ. Formally, $\sigma \in \Sigma$ if

$$\sum_{\theta_{-i} \in \Theta_{-i}} \sum_{a_{-i} \in \mathcal{A}_{-i}} p(\theta_{-i} \mid \theta_i) \sigma(a_i, a_{-i} \mid \theta_i, \theta_{-i}) \Big[u_i\big((a_i, a_{-i}), g; \theta_i, \theta_{-i}\big) - u_i\big((a_i', a_{-i}), g; \theta_i, \theta_{-i}\big) \Big] \geqslant 0$$

$$\text{for all } i \in \mathcal{N},\ \theta_i \in \Theta_i,\ a_i \in \mathcal{A}_i,\ a_i' \in \mathcal{A}_i. \qquad (2.1)$$

This formulation follows the information-design-in-games approach of Bergemann and Morris (2019). The environment consists of a basic game (parties, a prior over states, actions, and payoffs) and an initial information structure, under which each party observes only its own component of the state. For any fixed ownership structure within this environment, the obedience constraints in (2.1) characterize the set of Bayes correlated equilibrium outcomes, where an outcome is a joint distribution over actions and states, $\psi(a, \theta) := p(\theta)\sigma(a \mid \theta)$. This set coincides with the set of Bayes Nash equilibrium (BNE) outcomes that can arise under some expansion of the parties' initial information structure (Bergemann and Morris, 2016).

We study *optimal decision rules* that maximize total surplus subject to obedience. Let

$$TS(g) := \max_{\sigma \in \Sigma} \mathbb{E}\left[TS(a, g; \theta)\right], \tag{2.2}$$

where the expectation is taken with respect to the joint distribution $\psi(a, \theta) := p(\theta)\sigma(a \mid \theta)$. An *optimal asset owner* is any g^* that solves $\max_{g \in \mathcal{N}} TS(g)$.

Discussion of Assumptions. Our solution concept is best understood as a benchmark that isolates a single friction — the non-contractibility of actions — while abstracting from other frictions. Because actions are non-contractible, parties choose what is optimal for them given the information they possess. We then ask what is the best feasible outcome when parties may receive additional information beyond what they initially know. Once we establish this benchmark, we study whether it can be supported as an equilibrium outcome of an underlying game and how it changes when additional frictions are introduced or assumptions are relaxed.

We formally analyze these variations in Section 5. We show that the baseline outcome can be implemented in a broader contracting game (Section 5.1) and that the key friction making ownership matter is the non-contractibility of expropriation rather than cooperative actions or the division of cooperative surplus (Section 5.2). We also explore how our results change when we introduce unmediated cheap-talk communication (Section 5.3), separable control rights across multiple assets (Section 5.4), and arbitrary expropriation costs (Section 5.5).

3 Analysis

The fundamental constraint on total surplus is the threat of expropriation. Party i can profitably attempt to expropriate party j's knowledge only if it simultaneously has (i) the *means* to expropriate (access to an effective expropriation action under the current ownership structure) and (ii) sufficiently precise *information* about θ_j. As we show below, it is optimal to ensure that no party is both *armed* (has expropriation means) and sufficiently *well-informed* about a target to engage in expropriation.

This goal can be achieved through two channels. First, *information design*: holding ownership fixed, any party that is both armed and required for a project's success must receive capped information (up to its safety threshold), while parties without expropriation means can be fully informed safely. Second, *ownership and production organization*: changing ownership affects who has asset-dependent expropriation means and who must be informed for production. Thus, expropriation can be deterred either by removing a party's means of expropriation or by structuring production so it need not learn others' information.

3.1 Optimal Information Disclosure

Fix an owner $g \in \mathcal{N}$ and an obedient decision rule $\sigma(\cdot \mid \theta) \in \Delta(\mathcal{A})$. After observing private state θ_i and recommendation a_i, party i forms posterior beliefs about others' states. For $j \neq i$, let

$$\tau_{ij}(\theta_i, a_i) := \mathbb{P}_\psi\big(\theta_j = 1 \,|\, \theta_i, a_i\big),$$

where the posterior is defined under the joint distribution $\psi(\theta, a) = p(\theta)\sigma(a|\theta)$. When the dependence on (θ_i, a_i) is clear from context, we write τ_{ij} for $\tau_{ij}(\theta_i, a_i)$.

If party i can engage in expropriation, it will choose $\zeta_{ij} = 1$ if $\tau_{ij}b_{ij} - (1 - \tau_{ij})c_{ij} > 0$ and $\zeta_{ij} = -1$ if $(1 - \tau_{ij})b_{ij} - \tau_{ij}c_{ij} > 0$. Combining these inequalities, expropriation occurs if and only if the belief is sufficiently precise:

$$\max\{\tau_{ij}, 1 - \tau_{ij}\} > \tau^\star_{ij} := \frac{c_{ij}}{b_{ij} + c_{ij}} > \frac{1}{2},$$

where we assume that in the case of indifference, parties do not expropriate.

Before stating the general results on the structure of the optimal decision rule, it is useful to build intuition with two examples.

Example 1: Optimal Decision Rule For One Party. Suppose there are two parties, $\mathcal{N} = \{1, 2\}$, and one project, $\mathcal{K} = \{1\}$. The project does not require an asset, $\mathcal{R}_1 = \varnothing$, but requires party 1 to match the state of party 2 via its inalienable action, $\bar{\mathcal{R}}_1 = \{(1, 2)\}$. That is, the project succeeds if and only if $\bar{x}_{12} = \theta_2$. Parties do not require the asset to engage in expropriation, $\mathcal{E} = \varnothing$. In this example, the asset ownership is irrelevant. Suppose both parties receive half of the project's value, $\alpha_{11} = \alpha_{21} = 1/2$.

The optimal decision rule can be found using the techniques from Kamenica and Gentzkow (2011). Specifically, denote by τ the posterior belief of party 1 that $\theta_2 = 1$ after receiving the recommendation from the mediator. Figure 1 illustrates the expected total surplus as a function τ. It is optimal for party 1 to select $\bar{x}_{12} = 1$ if $\tau > 1/2$ and $\bar{x}_{12} = -1$ if $\tau < 1/2$. If party 1's information about θ_2 is sufficiently precise, it also optimally engages in expropriation.

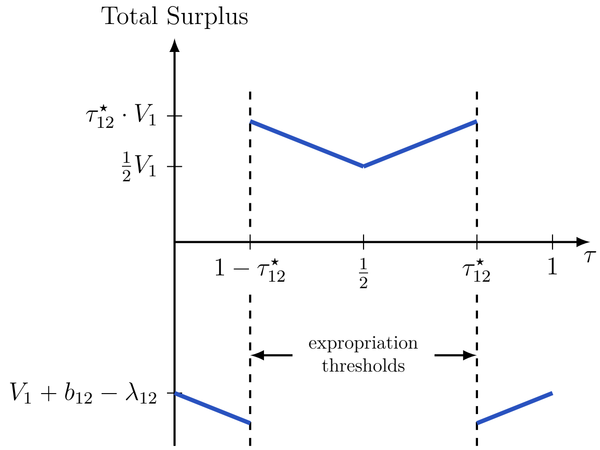

Figure 1: Total Surplus As a Function of Beliefs (Example 1)

This figure illustrates the relationship between total expected surplus and party 1's belief that $\theta_2 = 1$ for Example 1.

Specifically, if $\tau > \tau^\star_{12}$, party 1 selects $\zeta_{12} = 1$, and $\tau < 1 - \tau^\star_{12}$, party 1 selects $\zeta_{12} = -1$.

Under expropriation, the total expected surplus is negative. Thus, the optimal decision rule is to randomize between posteriors $\tau^\star_{12}$ and $1 - \tau^\star_{12}$. It can be constructed as follows. Suppose there is a random variable $\xi \sim \text{Unif}[0, 1]$ independent of θ. The recommendations to party 1 are

$$x_{12} = \bar{x}_{12} = \begin{cases} \theta_2 & \text{if } \xi \leqslant \tau^\star_{12}, \\ -\theta_2 & \text{if } \xi > \tau^\star_{12}, \end{cases} \quad \text{and} \quad z_{12} = \bar{z}_{12} = 0.$$

Example 2: Nested Decision Rule For Multiple Parties. Suppose there are three parties, $\mathcal{N} = \{1, 2, 3\}$, and two projects, $\mathcal{K} = \{1, 2\}$. Neither project has alienable requirements, so $\mathcal{R}_1 = \mathcal{R}_2 = \varnothing$. Project 1 requires two inalienable actions from parties 1 and 2 to match the state of party 3, so $\bar{\mathcal{R}}_1 = \{(1, 3), (2, 3)\}$. Project 2 requires only an inalienable action from party 2 to match the state of party 3, so $\bar{\mathcal{R}}_2 = \{(2, 3)\}$. All parties possess inalienable means of expropriation ($\mathcal{E} = \varnothing$), so asset ownership plays no role in this example.

Suppose expropriation parameters are $b_{13} = 2$, $c_{13} = 3$, $b_{23} = 1$, and $c_{23} = 9$. Then

$$\tau^\star_{13} = \frac{3}{2+3} = 0.6 \qquad \text{and} \qquad \tau^\star_{23} = \frac{9}{1+9} = 0.9.$$

To deter expropriation, party i's recommendation for the action $\bar{x}_{i3}$ must be such that its posterior belief τ_{i3} about θ_3 lies in the safety region $[1 - \tau^\star_{i3}, \tau^\star_{i3}]$. Following Example 1, a natural obedient rule is to recommend the correct match to party i with probability $\tau^\star_{i3}$ and the incorrect match with probability $1-\tau^\star_{i3}$. However, with two parties required for Project 1, the key question is whether these recommendations should be generated independently.

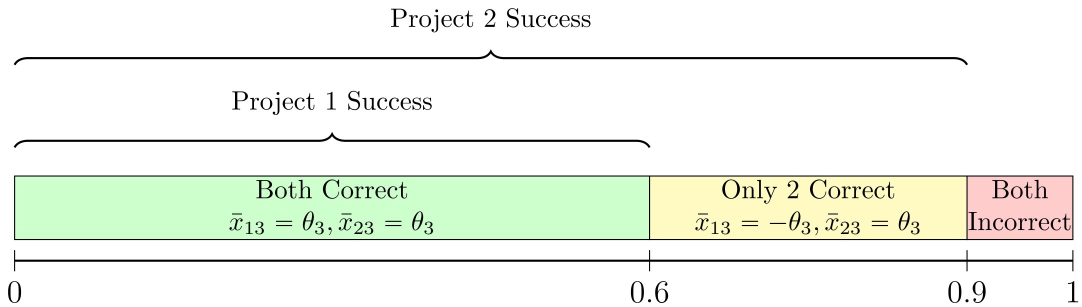

Figure 2: Nested Decision Rule For Example 2

The figure illustrates the realization of recommendations based on the random variable $\xi \sim \text{Unif}[0,1]$. Party 1 receives the correct recommendation only when $\xi \leqslant 0.6$. Party 2 receives the correct recommendation when $\xi \leqslant 0.9$. This decision rule ensures that Party 2 is always informed whenever Party 1 is informed.

If parties 1 and 2 receive independent recommendations satisfying their respective safety thresholds, then Project 1 succeeds with probability $\tau_{13}^\star \cdot \tau_{23}^\star$, while Project 2 succeeds with probability $\tau_{23}^\star$. Hence the expected total surplus is

$$V_1 \cdot \tau_{13}^\star \cdot \tau_{23}^\star + V_2 \cdot \tau_{23}^\star = V_1 \cdot 0.54 + V_2 \cdot 0.9$$

The mediator can do strictly better by *positively correlating* recommendations. Intuitively, party 1 is the bottleneck in Project 1, so whenever party 1 is recommended the correct action, party 2 should also be recommended the correct action.

Let $\xi \sim \text{Unif}[0,1]$ be a *common* randomization device. Consider the nested rule in which, for each $i \in \{1,2\}$, the mediator privately recommends $\bar{x}_{i3} = \theta_3$ if $\xi \leqslant \tau_{i3}^\star$ and $\bar{x}_{i3} = -\theta_3$ otherwise. This rule is illustrated in Figure 2. Under it, Project 1 succeeds with probability $\min\{\tau_{13}^\star, \tau_{23}^\star\} = \tau_{13}^\star$, while Project 2 succeeds with probability $\tau_{23}^\star$. The expected total surplus is

$$V_1 \cdot \min\{\tau_{13}^\star, \tau_{23}^\star\} + V_2 \cdot \tau_{23}^\star = V_1 \cdot 0.6 + V_2 \cdot 0.9,$$

which strictly exceeds the independent benchmark. Proposition 1 below shows that such nested decision rules — implemented via a common randomization device — are optimal more generally.

General Results. Fix an owner $g \in \mathcal{N}$. Let $\text{Arm}(g)$ denote the set of "armed" parties, that is parties, who can engage in expropriation under owner $g \in \mathcal{N}$:

$$\text{Arm}(g) := \{i \in \mathcal{N} : i = g \text{ or } i \notin \mathcal{E}\}.$$

This set includes the owner and all parties with inalienable means for expropriation ($i \notin \mathcal{E}$).

Disarmed parties can be fully informed about others because they cannot expropriate. For armed parties, by contrast, information must be capped at the expropriation threshold. Let $\tau_{ij}^{\text{safe}}(g)$ denote the maximum precision of party i's belief about θ_j that is consistent with no expropriation:

$$\tau_{ij}^{\text{safe}}(g) := \begin{cases} \tau_{ij}^{\star} = \frac{c_{ij}}{b_{ij}+c_{ij}} & \text{if } i \in \text{Arm}(g) \\ 1 & \text{if } i \notin \text{Arm}(g) \end{cases}.$$

For convenience, we set $\tau_{ii}^{\text{safe}}(g) \equiv 1$ for all $i \in \mathcal{N}$ and $g \in \mathcal{N}$.

For project k, it is useful to represent the production requirements as *matching tasks* of the form "party i must match party j's state." Let

$$\mathcal{Q}_k(g) := \{(g, j) : j \in \mathcal{R}_k\} \ \cup \ \bar{\mathcal{R}}_k \ \subseteq \ \mathcal{N} \times \mathcal{N}$$

denote the set of such tasks that must be performed for project k to succeed under ownership g.

The next proposition establishes that for each project $k \in \mathcal{K}$ the maximum attainable expected surplus is determined by the *bottleneck task*: the lowest safety threshold $\tau_{ij}^{\text{safe}}(g)$ among tasks in $(i, j) \in \mathcal{Q}_k(g)$.

Proposition 1 (Optimal Decision Rule)**.** *Fix an asset owner $g \in \mathcal{N}$. Under Assumptions 1–2, the maximum attainable expected total surplus is*

$$TS(g) = \sum_{k \in \mathcal{K}} V_k \cdot \left(\min_{(i,j) \in \mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) \right). \tag{3.1}$$

Moreover, there exists an optimal obedient decision rule with a nested structure: whenever a task (i, j) with lower safety threshold succeeds, all tasks with higher safety thresholds succeed as well. The mediator draws a single common random variable $\xi \sim \text{Unif}[0, 1]$ and, conditional on (θ, ξ), privately recommends:

(i) No expropriation: each party i is recommended to choose $z_{ij} = \bar{z}_{ij} = 0$ for all $j \neq i$.

(ii) Cooperation: party i is recommended the correct match ($x_{ij} = \bar{x}_{ij} = \theta_j$) if and only if $\xi \leqslant \tau_{ij}^{\text{safe}}(g)$, and is recommended the opposite match otherwise ($x_{ij} = \bar{x}_{ij} = -\theta_j$).

Proof. See Section A.1 in the appendix. □

The intuition behind Proposition 1 is as follows. The mediator prefers to avoid expropriation because it destroys total surplus. To do so, information must be partially withheld from *armed* parties (those with the means to expropriate). Therefore, the mediator discloses information to each party i about state θ_j only up to its safety threshold $\tau_{ij}^{\text{safe}}(g)$.

Since cooperative actions are complementary, the probability of successful cooperation in a project is bounded by the bottleneck task: the party-target pair with the lowest safety threshold, captured by the term $\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g)$. The nested decision rule described in the proposition achieves this bound by correlating agents' information through a common randomization device. This decision rule creates a strict hierarchy of information: if a less trustworthy agent is correctly informed about a particular state, then all more trustworthy agents are also correctly informed about this state.

3.2 Optimal Asset Ownership

Proposition 1 characterizes the maximum total surplus as a function of ownership structure $g \in \mathcal{N}$. Optimal asset ownership solves the following problem:

$$g^* \in \arg\max_{g\in\mathcal{N}} TS(g) = \arg\max_{g\in\mathcal{N}} \left\{ \sum_{k\in\mathcal{K}} V_k \cdot \left(\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) \right) \right\}.$$

Ownership and Bundling of Control Rights. We start with a result on how the bundling of multiple decision rights into a single asset creates inefficiency.

Corollary 1. *The first best is achieved only when there is at most one alienable requirement:*

$$\max_g TS(g) = V_{\text{total}} \quad \Rightarrow \quad \Big| \bigcup_{k\in\mathcal{K}} \mathcal{R}_k \Big| \leqslant 1$$

Proof. If there is more than one alienable requirement, then for any ownership assignment $g \in \mathcal{N}$, the asset owner must match the state of at least one party $j \neq g$. Because g is armed, the total surplus in projects where $j \in \mathcal{R}_k$ is at most $\tau_{gj}^{\star} V_k < V_k$. □

If there are at least two asset-dependent tasks, the owner must rely on information from at least one non-owner. Because the owner is armed, it cannot be fully informed, and this necessarily creates inefficiency.

The Two Roles of Ownership. To further characterize how optimal asset ownership depends on the model's primitives, let $\mathcal{P}_k(g)$ denote the set of *active* participants in project k under ownership $g \in \mathcal{N}$:

$$\mathcal{P}_k(g) := \left\{ i \in \mathcal{N} : \text{there exists some } j \neq i \text{ such that } (i,j) \in \mathcal{Q}_k(g) \right\}. \tag{3.2}$$

That is, $\mathcal{P}_k(g)$ consists of parties who must match the state of some *other* party for project k to succeed.

We provide two examples that clarify the two roles of ownership. First, control over production operates through the task set $\mathcal{Q}_k(g)$ by determining the set of active participants

$\mathcal{P}_k(g)$. Second, control over expropriation operates through the safety thresholds $\tau_{ij}^{\text{safe}}(g)$ by determining who has the means to expropriate.

Example 3: Ownership as Control Over Production. Suppose there are two parties, $\mathcal{N} = \{1, 2\}$, and a single project, $\mathcal{K} = \{1\}$. The project requires an asset owner to match the states of both parties to succeed, so $\mathcal{R}_1 = \{1, 2\}$. No inalienable actions are required ($\bar{\mathcal{R}}_1 = \varnothing$), and all parties possess the inalienable means to expropriate ($\mathcal{E} = \varnothing$).

Given an owner $g \in \{1, 2\}$, the project succeeds if and only if $x_{g1} = \theta_1$ and $x_{g2} = \theta_2$. The owner can always match its own state, $x_{gg} = \theta_g$, without disclosing it. By contrast, the non-owner's state must be disclosed to the owner for the project to succeed. Since the owner is armed, the information disclosed about the non-owner's state must be capped to deter expropriation.

Formally, ownership affects the set of active participants, $\mathcal{P}_1(g) = \{g\}$. However, it does not affect safety thresholds ($\tau_{ij}^{\text{safe}}(g) = \tau_{ij}^{\star}$) because expropriation is inalienable ($\mathcal{E} = \varnothing$). The asset should be owned by the party with the weaker temptation to expropriate (equivalently, the higher safety threshold):

$$g^* \in \arg\max_{g \in \mathcal{N}} \left\{\tau_g^{\star} \cdot V_1\right\} \quad \Leftrightarrow \quad g^* \in \arg\min_{g \in \mathcal{N}} \left\{\frac{b_g}{c_g}\right\}, \tag{3.3}$$

where $\tau_g^{\star} := \tau_{gj}^{\star}$, $b_g := b_{gj}$, $c_g := c_{gj}$ for $j \neq g$.

Example 4: Ownership as Control Over Expropriation. Suppose again there are two parties, $\mathcal{N} = \{1, 2\}$, and a single project, $\mathcal{K} = \{1\}$. The project requires an asset owner to match the states of both parties ($\mathcal{R}_1 = \{1, 2\}$). Unlike the previous case, cooperation now requires complementary inalienable actions: each party must use their human capital to match the other party's state ($\bar{\mathcal{R}}_1 = \{(1, 2), (2, 1)\}$). Furthermore, we assume that both parties require the asset to engage in expropriation ($\mathcal{E} = \{1, 2\}$).

There are two key differences from Example 3. First, cooperation now requires inalienable actions by both parties, and therefore ownership does not affect the set of active participants, $\mathcal{P}_1(g) = \{1, 2\}$ for all $g \in \mathcal{N}$. Second, expropriation is alienable and controlled by the owner: the non-owner is disarmed and has no effective expropriation action. Therefore, ownership affects safety thresholds, $\tau_{ij}^{\text{safe}}(g) = \tau_{ij}^{\star}$ if $i = g$ and $\tau_{ij}^{\text{safe}}(g) = 1$ if $i \neq g$.

As in the previous example, the optimal asset owner is the party with the lowest temptation to expropriate (see (3.3)). The logic, however, differs across the two settings. In Example 3, ownership determines who participates in production. In Example 4, ownership determines who lacks the means to expropriate.

Additional Results. Next, we characterize the optimal ownership structure for several special cases. Let $\mathcal{Q}_{k,i}(g) := \{j \in \mathcal{N} : (i, j) \in \mathcal{Q}_k(g)\}$ denote the set of targets party i must match for project k to succeed. Let $\tau_{k,i}^{\min}(g) := \min_{j \in \mathcal{Q}_{k,i}(g)} \tau_{ij}^{\text{safe}}(g)$ denote party i's bottleneck in project k (the "hardest" task). For notational convenience, we set $\min \varnothing := 1$.

For the next result, we make the following assumption (in addition to Assumptions 1–2):

Assumption 3. *For every ownership structure $g \in \mathcal{N}$ and project $k \in \mathcal{K}$, non-owner active participants are disarmed: $\mathcal{P}_k(g) \cap \operatorname{Arm}(g) \subseteq \{g\}$.*

Under this assumption, any project in which the owner is *not* an active participant can be implemented with certainty: all of its active participants are disarmed and can safely be fully informed. The only projects that are "throttled" by an information cap are those in which the owner must act. Thus, the formula for maximum attainable surplus can be simplified.

Corollary 2. *Under Assumption 3, the maximum attainable surplus under an owner $g \in \mathcal{N}$ is*

$$TS(g) = V_{\text{total}} - \sum_{k \in \mathcal{K}} V_k \cdot \mathbf{1}\{g \in \mathcal{P}_k(g)\} \cdot (1 - \tau_{k,g}^{\min}(g)). \tag{3.4}$$

Proof. Follows directly from the formula in (3.1) under Assumption 3 and the definitions of $\mathcal{P}_k(g)$ and $\tau_{k,g}^{\min}(g)$. □

This corollary generalizes the logic of Examples 3 and 4. Specifically, if there is only one project, and an asset owner is the only armed active participant, then $TS(g) = \tau_{gj}^{\star} \cdot V_1$ for $j \neq g$, and the asset should be assigned to the party with the lowest temptation to expropriate. More generally, suppose there are many projects, and an asset owner is an active participant in all projects ($g \in \mathcal{P}_k(g)$ for all $k \in \mathcal{K}$). Then

$$TS(g) = \sum_{k \in \mathcal{K}} V_k \cdot \tau_{k,g}^{\min}(g),$$

so the optimal ownership trades off project values against the owner's worst bilateral safety threshold toward the targets it must match in each project.

The corollary also shows that the optimal ownership depends on parties' *involvement* in projects. Specifically, suppose that all parties have the same worst bilateral safety thresholds: $\tau_{k,g}^{\min}(g) := \tau_k^{\min}$ for all $g \in \mathcal{N}$. Then, the optimal owner is

$$g^* \in \arg\min_{g \in \mathcal{N}} \sum_{k \in \mathcal{K}} V_k \cdot \mathbf{1}\{g \in \mathcal{P}_k(g)\} \cdot (1 - \tau_k^{\min}),$$

which depends on g only through $\mathbf{1}\{g \in \mathcal{P}_k(g)\}$. Intuitively, if the owner is the only armed participant, ownership should not be assigned to a party whose human capital (inalienable actions) is required in many high-value projects. Making such a highly involved party the owner would "arm" it, necessitating restricted information flows across all those projects. Assigning ownership to the party who is *least* involved across projects can increase total surplus by allowing the more involved party to be fully informed.

In the next corollary, we focus on ownership as a tool for reallocating the means of expropriation by assuming it does not affect the production requirements.

Corollary 3. *Suppose that for all projects ownership does not affect the production requirements ($\mathcal{Q}_k(g) = \mathcal{Q}_k$ for all $g \in \mathcal{N}$ and all $k \in \mathcal{K}$). Then, it is weakly optimal to assign the asset to a party with inalienable means for expropriation, if one exists. That is, if $\mathcal{N} \backslash \mathcal{E} \neq \varnothing$, then for any $g \in \mathcal{E}$ and $h \in \mathcal{N} \backslash \mathcal{E}$, $TS(h) \geqslant TS(g)$.*

Proof. Fix $g \in \mathcal{E}$ and $h \in \mathcal{N} \backslash \mathcal{E}$. By definition of safety thresholds,

$$\tau_{ij}^{\text{safe}}(g) = \begin{cases} \tau_{gj}^{\star} & \text{if } i = g, \\ 1 & \text{if } i \in \mathcal{E} \backslash \{g\}, \\ \tau_{ij}^{\star} & \text{if } i \notin \mathcal{E}, \end{cases} \qquad \tau_{ij}^{\text{safe}}(h) = \begin{cases} 1 & \text{if } i \in \mathcal{E}, \\ \tau_{ij}^{\star} & \text{if } i \notin \mathcal{E}. \end{cases}$$

Thus, $\tau_{ij}^{\text{safe}}(h) \geqslant \tau_{ij}^{\text{safe}}(g)$ for all i, j. Condition $\mathcal{Q}_k(g) = \mathcal{Q}_k(h)$ for all $k \in \mathcal{K}$ implies $TS(h) \geqslant TS(g)$. □

Corollary 3 states that when ownership does not change the set of active participants, it is optimal to allocate the asset to a party that possesses inalienable means of expropriation (i.e., a party that can expropriate regardless of ownership). Since this party is dangerous irrespective of whether it owns the asset, giving it the asset does not create a new expropriation risk. Conversely, giving the asset to a party with alienable expropriation means would turn a previously safe party into a dangerous one.

4 Applications

Our framework highlights that asset ownership may protect knowledge through either control over production or control over expropriation. For example, Apple chose to design the iPad chip in-house rather than partner with Intel because it "just didn't want to teach them everything, which they could go and sell to our competitors" (Steve Jobs, quoted in Isaacson, 2011, p. 493). Here, ownership served to control production and limit knowledge disclosure. In contrast, Apple's ownership choices in its China operations are better interpreted as control over expropriation. Specifically, Apple actively shared knowledge with its Chinese suppliers. Although it did not own most production facilities, it owned key production machines, which limited suppliers' ability to expropriate Apple's know-how (McGee, 2025).

We also show that ownership has costs: non-owners may not be willing to fully disclose their information to the owner. For example, China's quid pro quo policy requires multinationals to form joint ventures with local partners to access the Chinese market. This structure gives the local partner control over some manufacturing decisions, forcing the multinational to disclose production-relevant information. One implication of this policy is that it may discourage automakers from bringing their most advanced technology to China due to concerns about expropriation (Bai et al., 2025).

In this section, we illustrate how our results translate into predictions about optimal ownership and information sharing in several industry environments where two competing downstream firms interact with a common upstream supplier.

Section 4.1 shows how the upstream party can act as a *knowledge hub*. Two competing downstream parties can benefit from pooling knowledge, but they are too tempted to expropriate each other's secrets. A third party that contributes neither private information nor human capital may nonetheless optimally hold the asset solely because its temptation to expropriate is lower. We use this case to rationalize knowledge sharing through third-party suppliers and solution providers, as commonly observed in the auto industry (Küpper et al., 2020).

Section 4.2 studies a different environment in which downstream firms do not need to share knowledge with each other, but each requires an input that only a common upstream supplier can provide. We derive simple conditions under which the supplier must remain independent to maximize total surplus, and when the optimal arrangement instead resembles an internal Chinese Wall: the supplier is a subsidiary of one downstream firm, the rival's project succeeds as often as if the supplier were independent, and the parent remains insufficiently informed to expropriate. This application helps explain GM's spin-off of Delphi and provides a rationale for separating Intel's foundry business from its chip-design operations (Fitch, 2025).

4.1 Knowledge Hubs

Consider an extension of Example 3 with three parties, $\mathcal{N} = \{1, 2, 3\}$, and a single project, $\mathcal{K} = \{1\}$. Parties 1 and 2 represent downstream competitors, and party 3 is a common supplier. The project requires the asset owner to match the states of parties 1 and 2 to succeed, so $\mathcal{R}_1 = \{1, 2\}$. No inalienable actions are required, $\bar{\mathcal{R}}_1 = \varnothing$, and all parties possess the inalienable means to expropriate, so $\mathcal{E} = \varnothing$.

In this environment, only the owner's alienable cooperative actions are payoff-relevant. The mediator can therefore keep the two non-owners uninformed, eliminating their expropriation incentives. The remaining constraint is the owner's own temptation: to implement the project, the owner must hold enough information about θ_1 and θ_2 to match them, but not so much as to trigger expropriation. By Proposition 1, total surplus under owner $g \in \{1, 2, 3\}$ is

$$TS(g = 1) = \tau_{12}^\star \cdot V_1, \qquad TS(g = 2) = \tau_{21}^\star \cdot V_1, \qquad TS(g = 3) = \min\{\tau_{31}^\star, \tau_{32}^\star\} \cdot V_1.$$

If the supplier is less tempted than either competitor,

$$\min\{\tau_{31}^\star, \tau_{32}^\star\} \geqslant \max\{\tau_{12}^\star, \tau_{21}^\star\},$$

then it is optimal to assign the asset to the supplier. The distinctive feature of this example is that the optimal owner (party 3) contributes neither private information nor inalienable inputs to production: θ_3 is payoff-irrelevant and $\bar{\mathcal{R}}_1 = \varnothing$. Party 3 is valuable solely because its temptation to expropriate is weaker. In this sense, it functions as a *knowledge hub*, enabling greater information pooling among competing firms.

4.2 Spin-Offs versus Chinese Walls

Suppose there are three parties, $\mathcal{N} = \{1, 2, 3\}$, and two projects, $\mathcal{K} = \{1, 2\}$. We interpret parties 1 and 2 as downstream competitors and party 3 as a common upstream supplier. Project 1 is associated with operations of party 1, and Project 2 is associated with party 2. Both projects require an inalienable input from the supplier: $\bar{\mathcal{R}}_1 = \{(3, 1)\}$ and $\bar{\mathcal{R}}_2 = \{(3, 2)\}$. Project 1 also requires an alienable action by an asset owner: $\mathcal{R}_1 = \{1\}$. For Project 2, we consider two cases: it either requires an alienable action, $\mathcal{R}_2 = \{2\}$, or it does not, $\mathcal{R}_2 = \varnothing$. Expropriation capabilities differ by party: the supplier requires the asset to expropriate, while the downstream parties can expropriate using inalienable means ($\mathcal{E} = \{3\}$).

Cooperative Surplus. Given an asset owner $g \in \{1, 2, 3\}$, the cooperative surplus is

$$V_1 \cdot \mathbf{1}\{\bar{x}_{31} = \theta_1\} \cdot \mathbf{1}\{x_{g1} = \theta_1\} + V_2 \cdot \begin{cases} \mathbf{1}\{\bar{x}_{32} = \theta_2\} \cdot \mathbf{1}\{x_{g2} = \theta_2\} & \text{if} \quad \mathcal{R}_2 = \{2\}, \\ \mathbf{1}\{\bar{x}_{32} = \theta_2\} & \text{if} \quad \mathcal{R}_2 = \varnothing. \end{cases}$$

Bundling of Control Rights. When Project 2 requires the asset ($\mathcal{R}_2 = \{2\}$), the owner has two alienable requirements ($|\mathcal{R}_1 \cup \mathcal{R}_2| = 2$). Thus, by Corollary 1, the first best cannot be achieved. This inextricability lies at the heart of the trade-off between vertical integration and a spin-off. By contrast, when $\mathcal{R}_2 = \varnothing$, the owner is responsible for only one alienable requirement. In this case, the first best can be achieved and resembles an internal Chinese Wall.

Subsidiary. If a downstream party owns the asset ($g \in \{1, 2\}$), we call the supplier a subsidiary of that party (vertical integration). If the supplier owns the asset ($g = 3$), we call it an independent upstream firm.

Chinese Wall. We say that an *internal Chinese Wall* is feasible if there exists an optimal decision rule such that a downstream party's probability of project success when the supplier is a subsidiary of its rival is at least as high as when the supplier is independent. This definition implies that the subsidiary must receive as much information from the downstream non-owner as it would if it were independent. At the same time, the optimality of this rule ensures that the downstream owner (parent) remains insufficiently informed to expropriate.

Case 1: Party 2 Requires the Asset, $\mathcal{R}_2 = \{2\}$. If party 1 owns the asset, then Project 1 can be implemented with certainty: party 1 knows θ_1 and the supplier is disarmed (so θ_1 can be safely disclosed to party 3 to satisfy $\bar{x}_{31} = \theta_1$). Project 2 is different. To make the payoff-relevant alienable decision for Project 2, the owner must match θ_2. Since party 1 is armed, information about θ_2 must be capped at party 1's safety threshold $\tau^\star_{12}$. The same logic applies symmetrically if party 2 owns the asset. If instead the supplier owns the asset, then the supplier is armed and must receive capped information about both downstream parties' states, so Projects 1 and 2 succeed with probabilities $\tau^\star_{31}$ and $\tau^\star_{32}$, respectively. Consequently, the maximal attainable surplus under each ownership assignment is

$$TS(g=1) = V_1 + \tau^\star_{12} V_2, \qquad TS(g=2) = \tau^\star_{21} V_1 + V_2, \qquad TS(g=3) = \tau^\star_{31} V_1 + \tau^\star_{32} V_2.$$

If the supplier is a subsidiary of party 1 ($g = 1$), party 2's project succeeds with probability $\tau^\star_{12}$. If the supplier is independent ($g = 3$), this probability is $\tau^\star_{32}$. Because we assume the supplier is less tempted to expropriate ($\tau^\star_{32} > \tau^\star_{12}$), an internal Chinese Wall is not feasible: party 2's project is strictly more likely to succeed under an independent supplier. A symmetric argument applies if party 2 owns the asset.

To simplify the derivation of the optimal ownership structure, suppose that $\tau^\star_D := \tau^\star_{12} = \tau^\star_{21}$ and $\tau^\star_U := \tau^\star_{31} = \tau^\star_{32}$. Recall that the supplier is less tempted to expropriate than a direct competitor, $\tau^\star_U > \tau^\star_D$. Define the threshold

$$\kappa := \frac{1 - \tau^\star_U}{\tau^\star_U - \tau^\star_D}$$

and assume that $\kappa < 1$. Then vertical integration by party 1 ($g = 1$) is optimal when its project is sufficiently more valuable than party 2's project, $V_2/V_1 < \kappa$. Symmetrically, if $V_2/V_1 > 1/\kappa$, vertical integration by party 2 ($g = 2$) is optimal. For intermediate values, $V_2/V_1 \in [\kappa, 1/\kappa]$, an independent supplier ($g = 3$) is optimal.[3] Figure 3 summarizes the resulting pattern.

[3]If $\kappa > 1$, the optimal ownership is $g^* = 1$ if $V_2/V_1 \leqslant 1$ and $g^* = 2$ if $V_2/V_1 > 1$.

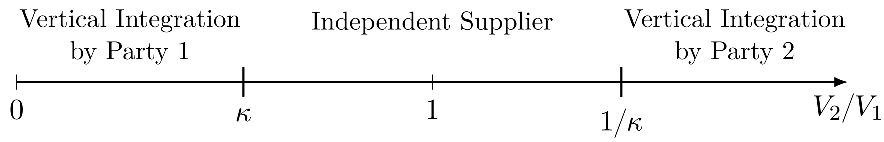

Figure 3: Optimal Ownership When Both Projects Require the Asset

The optimal owner varies with the relative market size V_2/V_1. For this figure, we assume that $\kappa < 1$.

Case 2: Party 2 Does Not Require the Asset, $\mathcal{R}_2 = \varnothing$. Relative to Case 1, Project 2 no longer requires the owner to match θ_2. If party 1 owns the asset ($g = 1$), the supplier is disarmed and can be fully informed about both θ_1 and θ_2. Meanwhile, the parent can remain uninformed about θ_2 and match its own state with certainty. Thus, the first-best outcome is achieved.

The resulting attainable surplus levels for $g \in \{1, 2, 3\}$ are

$$TS(g = 1) = V_1 + V_2, \qquad TS(g = 2) = \tau^\star_{21} V_1 + V_2, \qquad TS(g = 3) = \tau^\star_{31} V_1 + \tau^\star_{32} V_2.$$

When $g = 1$, Project 2 succeeds with probability 1. When $g = 3$, Project 2 succeeds with probability $\tau^\star_{32} < 1$. Since $1 > \tau^\star_{32}$, an internal Chinese Wall is feasible.

Interpretation. Suppose that party 1 is the only downstream firm (incumbent) serving its market (Project 1): $V_1 > 0$ and $V_2 = 0$. Then, it is optimal for this party to own the asset to prevent expropriation by the supplier and safely share knowledge with it. Now assume that a new downstream party enters and serves a different market ($V_2 > 0$). Ideally, they should serve their respective markets, but they can also engage in expropriation. The optimal ownership structure depends on whether the entrant requires the asset that the incumbent party owns to control expropriation by the supplier.

If the entrant does require alienable actions based on this asset, then as the new market V_2 grows in size, the incumbent faces a trade-off: keeping the asset to protect its knowledge versus spinning it off to realize the higher potential of the new market. This interpretation is consistent with GM's spin-off of Delphi. If the entrant does not require actions based on the asset that controls the supplier's expropriation, then it is optimal for the incumbent to keep the asset, and the optimal information structure resembles a Chinese Wall between the parent and the subsidiary.

We note that establishing the theoretical possibility of a Chinese Wall is distinct from the question of how to build it. Questions regarding the specific internal governance mechanisms and incentive systems required to enforce separation between a parent and its subsidiary lie beyond the scope of this framework and represent an exciting avenue for future research.

5 Discussion

This section clarifies the mechanisms driving our main results by exploring several variations of the baseline model. We establish that the outcome of our information-design approach can implemented in a broader contracting game (Section 5.1), isolate the specific contracting frictions that make ownership matter (Section 5.2), and relax the assumptions of mediated communication (Section 5.3), bundled control rights (Section 5.4), and prohibitive expropriation costs (Section 5.5).

5.1 Implementation Through a Mediated Contract

In the baseline model, we assume that an omniscient mediator can condition recommendations on the true state. Suppose instead that the mediator does not observe the state directly, but parties can use the mediator to exchange information.

Communication Game. Consider the following multi-stage communication game. (i) The state θ is realized, and each party i privately observes θ_i. (ii) Parties confidentially submit reports from some finite message spaces to a neutral mediator, who then implements ownership and transfers and sends private output messages to each party according to a pre-specified mechanism. (iii) Each party privately observes its own output message, together with the implemented ownership and transfers. (iv) Parties simultaneously choose their non-contractible actions. (v) Payoffs are realized.

Mechanism Design. To characterize the set of outcomes that can be supported as a Bayesian Nash equilibrium of some communication game, it is without loss of generality to restrict attention to *direct* mediated mechanisms (Forges, 1986). Specifically, a direct mediated mechanism consists of an allocation rule $\varphi : \Theta \to \Delta(\mathcal{A} \times \mathcal{N})$ and a transfer rule $t : \mathcal{A} \times \mathcal{N} \times \Theta \to \mathcal{T}$. Each party i submits a report $\hat{\theta}_i \in \{-1, 1\}$ to the mediator. Given the report profile $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_N) \in \Theta := \{-1, 1\}^N$, the mediator implements asset ownership $g \in \mathcal{N}$ and sends private action recommendations $a = (a_1, \ldots, a_N) \in \mathcal{A}$ according to $\varphi(a, g \mid \hat{\theta})$. The mediator also enforces monetary transfers according to $t(a, g, \hat{\theta})$. Transfers must be budget balanced, so $\mathcal{T}$ is the set of budget-balanced transfer vectors.

An outcome is a joint distribution over actions, ownership, transfers, and states. Feasible outcomes must satisfy incentive-compatibility constraints: no party should gain by misreporting its state, disobeying its recommendation, or combining the two deviations. Our first result shows that the maximal total surplus in this communication game coincides with the value of the information-design problem.

Proposition 2. *The maximum attainable surplus in the communication game is* $\max_{g \in \mathcal{N}} TS(g)$. *This value is attained by a direct mediated mechanism.*

The intuition behind this proposition is as follows. For fixed ownership, the set of feasible outcomes is characterized by obedience, truth-telling, and no-double-deviation constraints. Therefore, the maximum attainable surplus cannot exceed the value of the benchmark mediator problem, which imposes only obedience. Randomizing over ownership does not help relax these constraints because parties observe ownership when choosing their non-contractible actions (a party that operates the asset knows that it operates the asset).

The allocation rule that achieves the maximum possible surplus is given by

$$\varphi^*(a, g^* \mid \theta) = \sigma^*(a \mid \theta) \qquad \text{and} \qquad \varphi^*(a, g \mid \theta) = 0 \text{ for } g \neq g^*, \tag{5.1}$$

where $g^* \in \arg\max_g TS(g)$ and $\sigma^*(a \mid \theta)$ is the optimal decision rule under g^*. We show that, under constant transfers, this mechanism is incentive compatible. Intuitively, once no armed party receives enough information to profitably expropriate, the remaining incentives are aligned, so truthful reporting can be sustained. Thus, the central friction in our setting is the non-contractibility of actions rather than private information per se.

Contracting Game. We now embed the communication game in a broader bargaining environment. Fix an initial owner $g_0 \in \mathcal{N}$. A *contract* is a direct mediated mechanism (φ, t), where φ is an allocation rule and t is a budget-balanced transfer rule, as defined above.

The contracting game has two additional stages before the communication game. First, party 1 makes a take-it-or-leave-it offer of a contract (φ, t) to the other parties. Second, each party $i \in \{2, \ldots, N\}$ either accepts or rejects. If all parties accept, the communication game induced by (φ, t) is played. If at least one party rejects, the contract is void, the asset owner remains g_0, transfers are zero, no information is shared, and parties choose actions in the disagreement environment.

In Appendix A.2, we show that there exists a Perfect Bayesian Equilibrium in which party 1 offers the direct mediated contract (φ^*, t^*), where φ^* is given by (5.1) and t^* ensures participation. All other parties accept, and, conditional on acceptance, parties report truthfully and obey their recommendations.

5.2 Contractibility of Actions

In the baseline model, we assume that all actions are non-contractible. If some actions are contractible, then whenever the mediator recommends those actions they are enforced, which relaxes the obedience constraints. Thus, contractibility weakly increases the maximum attainable total surplus. This section outlines the results; formal proofs are provided in Appendix A.3.

If *expropriation* actions $(z, \bar{z})$ are contractible, then the mediator can directly enforce $z_{ij} = \bar{z}_{ij} = 0$ before any information is disclosed. Once expropriation is ruled out, the mediator can fully reveal information and achieve the first best under *any* ownership structure.

If *cooperative* actions $(x, \bar{x})$ and project shares (α_{ik}) are contractible while expropriation remains non-contractible, then the fundamental constraint is unchanged: after learning information, an armed party can still deviate toward expropriation. Therefore, information disclosure must still satisfy the same *expropriation obedience* (safety-threshold) constraints as in the baseline model. As a result, the maximal attainable total surplus is governed by the same obedience-based bound as in the baseline problem (i.e., it cannot exceed $\max_g TS(g)$).

If only the cooperative shares (α_{ik}) are contractible (without making $(x, \bar{x})$ contractible), this cannot increase attainable total surplus either. The "shares-only" environment is a restriction of the "contractible cooperation and shares" environment, so its value is weakly smaller. Since the latter yields no improvement over the baseline and the baseline policy remains feasible when shares are contractible, all three values coincide.

5.3 Non-Contractible Communication: Cheap Talk

In the baseline setting, the mediator plays two distinct roles: it garbles information to deter expropriation (limiting the posterior precision of armed receivers), and it coordinates disclosure by correlating recommendations across parties. In this subsection, we examine the role of the omniscient mediator by contrasting it with unmediated *cheap talk*, where each sender can condition its message only on its own state.

We show that with a sufficiently rich state and action space, direct cheap talk can replicate the mediator's information-garbling role without any need for commitment to randomization. Thus, in environments that require only information garbling, such as the spin-off or Chinese Wall applications, the results remain the same under cheap-talk communication. However, the mediator's coordination role generally cannot be replicated. Because parties send messages independently rather than through a central coordinator, total surplus under the O-ring technology is reduced by uncoordinated disclosure.

Cheap Talk with Binary State. Suppose party j must disclose its information to exactly one armed party i. To deter expropriation, i's posterior precision regarding θ_j must not exceed the threshold $\tau^\star_{ij} \in (\frac{1}{2}, 1)$. Under the optimal mediated decision rule, the mediator ensures this constraint binds by committing to a noisy disclosure: it recommends the correct action with probability $\tau^\star_{ij}$ and the incorrect action with probability $1 - \tau^\star_{ij}$. This rule maximizes the probability of successfully completing task (i, j) while deterring expropriation.

Under direct communication, party j lacks the ex-ante commitment power to implement this garbling. Any informative pure strategy profile must be fully revealing, in which case the armed receiver i would expropriate the information, giving sender j a profitable deviation. A babbling equilibrium always exists, but it yields a strictly lower probability of task success than the mediated benchmark.

To attain the mediator's success probability $\tau_{ij}^{\star}$, the sender must therefore randomize across messages. However, because j lacks commitment, it must evaluate its messages ex-post (after observing the true state). Party j is willing to randomize its disclosure only if it is indifferent between reporting the true state and lying. For this indifference condition to hold, the receiver i must also play a mixed strategy, randomizing between attempting expropriation and no expropriation to keep the sender indifferent. Thus, any cheap-talk equilibrium that matches the mediator's probability of cooperative success must involve strictly positive expropriation on the equilibrium path, decreasing total expected surplus.

Rich State Space: Continuum Analogue. The failure of unmediated cheap talk to achieve optimal garbling is an artifact of the binary state space. With a binary state, any *pure* reporting strategy induces a posterior belief of either $\frac{1}{2}$ or 1. Therefore, reaching the indifference threshold $\tau_{ij}^{\star} \in (\frac{1}{2}, 1)$ requires the sender to randomize. However, by enriching the state space to a continuum, the sender can induce any posterior between $\frac{1}{2}$ and 1 through *pure* strategies. Thus, the lack of commitment to randomize is not a fundamental economic friction in our environment.

Formally, suppose instead of a binary state each party i privately observes an independent state θ_i uniformly distributed on a unit circle $\mathbb{T} = [0, 1)$. Cooperative actions x_{ij} and $\bar{x}_{ij}$ are chosen from $\mathbb{T}$. Expropriation actions z_{ij} and $\bar{z}_{ij}$ are chosen from $\mathbb{T} \cup \{\varnothing\}$, where $\varnothing$ denotes the safe action of not attempting expropriation.

Actions succeed if they fall within a target-specific circular tolerance $\delta_j > 0$ of the true state, using the circular distance $d(x, y) := \min\{|x - y|,\ 1 - |x - y|\}$. Specifically, for an owner $g \in \mathcal{N}$, an alienable task $j \in \mathcal{R}_k$ succeeds if $d(x_{gj}, \theta_j) \leqslant \delta_j$, and an inalienable task $(i, j) \in \bar{\mathcal{R}}_k$ succeeds if $d(\bar{x}_{ij}, \theta_j) \leqslant \delta_j$. As in the baseline model, party i's effective expropriation action is ζ_{ij}. Expropriation succeeds if $\zeta_{ij} \in \mathbb{T}$ and $d(\zeta_{ij}, \theta_j) \leqslant \delta_j$, yielding the same payoffs for successful and unsuccessful attempts as the baseline model.

We impose the following continuous analogue of Assumption 2:

Assumption 2⋆. *For each party $i \in \mathcal{N}$ and target $j \neq i$, $b_{ij} < c_{ij}$ and $\delta_j \in (0, 1/4]$.*

This assumption ensures that an uninformed party has no incentive to expropriate. Specifically, under the uniform prior, if an uninformed party i guesses the state θ_j, it succeeds with probability $2\delta_j$. Because the benefit of successful expropriation is strictly lower than the cost of an unsuccessful attempt ($b_{ij} < c_{ij}$), it follows that the expropriation threshold is $\tau^{\star}_{ij} = \frac{c_{ij}}{b_{ij}+c_{ij}} > 2\delta_j$. Therefore, without additional information, the party will not expropriate.

The Mediator's Benchmark. In the continuum analogue, the omniscient mediator problem has the same bottleneck logic as the baseline binary model (see Appendix A.4). To rule out expropriation, each task (i, j) cannot succeed with a probability higher than its safety threshold $\tau^{\text{safe}}_{ij}(g)$. Therefore, project k cannot succeed with a probability higher than the minimum of these safety thresholds across all required tasks. We construct an obedient decision rule that achieves this upper bound, delivering the same total surplus formula as in Proposition 1 (see Equation 3.1).

Consider the continuum adaptation of the optimal decision rule in the binary case. Define the *safety arc length* as

$$L_{ij}(g) := \frac{2\delta_j}{\tau^{\text{safe}}_{ij}(g)},$$

and let $\xi \sim \text{Unif}[0, 1]$ be a common randomization device. The mediator recommends

$$x_{ij} = \bar{x}_{ij} = \theta_j + L_{ij}(g)\Big(\frac{1}{2} - \xi\Big) \pmod 1, \qquad z_{ij} = \bar{z}_{ij} = \varnothing.$$

Geometrically, the mediator creates an arc of length $L_{ij}(g)$ centered at the true state θ_j, and then draws a recommendation uniformly at random from this arc. Conditional on receiving this recommendation, the posterior of an armed party is uniform over an interval of length $L_{ij}(g)$ with the recommended action as its midpoint. Therefore, the maximal probability of success for either the productive task or expropriation is $2\delta_j/L_{ij}(g) = \tau^{\text{safe}}_{ij}(g)$, which deters expropriation.

Cheap Talk. Consider the cheap talk under a continuous state and action space. For simplicity, we restrict attention to environments in which each sender's state must be communicated to at most one *armed* party. Formally, fix an owner g and suppose that for every target $j \in \mathcal{N}$, there is at most one party $i \in \text{Arm}(g)$ such that $(i, j) \in \bigcup_{k\in\mathcal{K}} \mathcal{Q}_k(g)$. This assumption covers Examples 3 and 4 and all applications from Section 4. We prove the following result.

Proposition 3 (Direct Garbling)**.** *Consider the continuum analogue of our environment. Suppose that each target state must be communicated to exactly one armed party. Under Assumptions 1 and 2***, there exists a pure-strategy Perfect Bayesian Equilibrium such that every task* (i, j) *succeeds with probability* $\tau^{\text{safe}}_{ij}(g)$*, and there is no expropriation on the equilibrium path.*

Proof. See Appendix A.4 □

The PBE is constructed via deterministic pooling of states, as illustrated in Figure 4. For each task (i, j) with an armed receiver i, the state space is partitioned into a success region

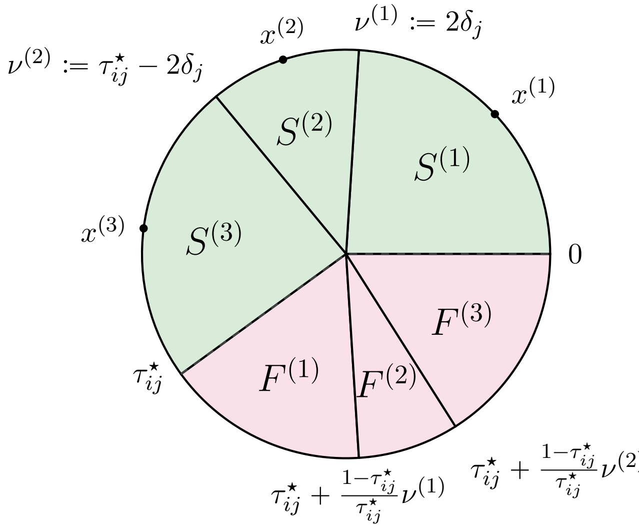

Figure 4: State space partition for the cheap-talk PBE

Party j sends message $m \in \{1, \ldots, M\}$ to armed party i if $\theta_j \in S^{(m)} \cup F^{(m)}$. This example is constructed under $\delta_j = 0.12$ and $\tau^\star_{ij} = 0.6$, where $M = \lceil \frac{\tau^\star_{ij}}{2\delta_j} \rceil = 3$. The cooperative actions corresponding to message $m \in \{1, 2, 3\}$ are $x^{(1)} = \delta_j$, $x^{(2)} = \tau^\star_{ij}/2$, and $x^{(3)} = \tau^\star_{ij} - \delta_j$.

$[0, \tau^\star_{ij}]$ and a failure region $(\tau^\star_{ij}, 1)$. Both regions are divided into M intervals $\{S^{(m)}\}_{m=1}^M$ and $\{F^{(m)}\}_{m=1}^M$ such that $\bigcup_{m=1}^M S^{(m)} = [0, \tau^\star_{ij}]$ and $\bigcup_{m=1}^M F^{(m)} = (\tau^\star_{ij}, 1)$, where $M = \lceil \frac{\tau^\star_{ij}}{2\delta_j} \rceil$ is the number of messages j can send to i.

In equilibrium, party j sends message $m \in \{1, 2 \ldots, M\}$ to i if $\theta_j \in S^{(m)} \cup F^{(m)}$. The lengths of the success and failure fragments are proportional such that

$$\frac{|S^{(m)}|}{|S^{(m)}| + |F^{(m)}|} = \tau^\star_{ij}.$$

Under this partition, the armed receiver i is indifferent between expropriating and not expropriating, and optimally selects the intended cooperative action $x^{(m)}$ that covers the success region $S^{(m)}$. Crucially, the sender j has no incentive to deviate: if $\theta_j \in S^{(m)}$, then sending m ensures the task (i, j) succeeds; if $\theta_j \in F^{(m)}$, the task fails regardless of the message sent, making the sender indifferent across all messages.

Proposition 3 implies that the cheap-talk environment delivers the same total surplus as the mediated benchmark in Examples 3 and 4, as well as in the spin-off and Chinese-Wall applications. Because each project in those environments relies on at most one task with an armed receiver, the optimal information disclosure (the mediator's solution) requires only information garbling, which can be replicated through direct communication with a rich enough state and action space.

The knowledge-hub application, however, is different. When the supplier acts as the hub ($g = 3$), it must learn both θ_1 and θ_2 from parties 1 and 2 to implement the project. Proposition 3 ensures that both tasks $(3, 1)$ and $(3, 2)$ can be independently implemented via cheap talk with success probabilities $\tau^\star_{31}$ and $\tau^\star_{32}$. However, because parties can condition their messages only on their own private states, the disclosures are independent. This lack of correlation yields a strictly lower expected surplus:

$$TS^{\text{cheap talk}}(g = 3) = \tau^\star_{31} \cdot \tau^\star_{32} \cdot V_1 < \min\{\tau^\star_{31}, \tau^\star_{32}\} \cdot V_1 = TS(g = 3).$$

Thus, the main benefit of being able to contractually specify a communication protocol through a neutral mediator is the coordination of information disclosure from various independent sources.

5.4 Multiple Assets and Separation of Control Rights

In the baseline model, a single asset bundles multiple alienable decision rights. This bundling generates the central trade-off: ownership protects the owner's knowledge but leaves others exposed to expropriation, discouraging disclosure. This subsection extends the model to multiple assets and allows control rights to be unbundled.

Assets. Let $\mathcal{M}$ be a finite set of assets. Two assignment functions determine which asset controls each *alienable* decision:

$$\phi^c : \mathcal{N} \to \mathcal{M}, \qquad \phi^e : \mathcal{N} \to \mathcal{M}.$$

Here $\phi^c(j)$ is the asset whose owner controls the alienable *cooperative* action targeting state θ_j, while $\phi^e(j)$ is the asset whose owner controls the alienable *expropriation* action targeting θ_j (for asset-dependent expropriators). The baseline model with one asset corresponds to $|\mathcal{M}| = 1$, so a single party controls *all* alienable actions. In general, control over production decisions is bundled when $\phi^c(i) = \phi^c(j)$ for $i \neq j$; control over production and expropriation is bundled when $\phi^c(j) = \phi^e(j)$.

An ownership profile is $g = (g_i)_{i \in \mathcal{M}} \in \mathcal{N}^{\mathcal{M}}$. For each target j, define the induced owners

$$g^c(j) := g_{\phi^c(j)}, \qquad g^e(j) := g_{\phi^e(j)}.$$

Production. Project k succeeds if and only if the relevant alienable and inalienable matching requirements are satisfied: $x_{g^c(j)\,j} = \theta_j$ for all $j \in \mathcal{R}_k$, and $\bar{x}_{\ell j} = \theta_j$ for all $(\ell, j) \in \bar{\mathcal{R}}_k$. The set of matching tasks for project k under ownership profile g is

$$\mathcal{Q}_k(g) := \{(g^c(j), j) : j \in \mathcal{R}_k\} \ \cup \ \bar{\mathcal{R}}_k. \tag{5.2}$$

Expropriation. A party i can expropriate target $j \neq i$ either via inalienable means ($i \notin \mathcal{E}$) or via an asset if $i \in \mathcal{E}$ and $i = g^e(j)$. Thus the set of parties who are armed *against target* j is

$$\text{Arm}_j(g) := \{i \in \mathcal{N} : i = g^e(j) \text{ or } i \notin \mathcal{E}\}.$$

Define the target-specific safety threshold (and set $\tau_{ii}^{\text{safe}}(g) := 1$)

$$\tau_{ij}^{\text{safe}}(g) \ := \ \begin{cases} \tau_{ij}^{\star} = \dfrac{c_{ij}}{b_{ij} + c_{ij}} & \text{if } i \in \text{Arm}_j(g) \text{ and } i \neq j, \\ 1 & \text{otherwise.} \end{cases} \tag{5.3}$$

Total Surplus. Under the same parameter restrictions as in Proposition 1 (Assumptions 1–2), the characterization of the maximum attainable total surplus carries over:

$$TS(g) = \sum_{k \in \mathcal{K}} V_k \cdot \min_{(i,j) \in \mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g),$$

where $\mathcal{Q}_k(g)$ and $\tau_{ij}^{\text{safe}}(g)$ are defined in (5.2) and (5.3), respectively.

Since the baseline model corresponds to $|\mathcal{M}| = 1$ and $\phi^c := \phi^e$, allowing separable control rights weakly increases maximum attainable total surplus. The following proposition shows how unbundling control rights across assets can achieve the first best.

Proposition 4 (First best under separable control rights)**.** *The first best V_{total} is achievable in each of the following cases.*

(i) All cooperative requirements are alienable ($\bar{\mathcal{R}}_k = \varnothing$ for all k), and distinct targets are controlled by distinct cooperative assets (ϕ^c is injective).

(ii) All expropriation is asset-dependent ($\mathcal{E} = \mathcal{N}$), and distinct targets are governed by distinct expropriation assets (ϕ^e is injective).

Proof. (i) Because ϕ^c is injective, we can choose an ownership profile g such that $g^c(j) = j$ for all $j \in \mathcal{N}$. Since $\bar{\mathcal{R}}_k = \varnothing$, each project has task set $\mathcal{Q}_k(g) = \{(j,j) : j \in \mathcal{R}_k\}$, which is trivial because party j observes θ_j. Hence $\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) = 1$ for every k and therefore $TS(g) = V_{\text{total}}$.

(ii) Because ϕ^e is injective, choose g such that $g^e(j) = j$ for all $j \in \mathcal{N}$. Fix any $i \neq j$. Since $\mathcal{E} = \mathcal{N}$, party i has no inalienable expropriation; and because $g^e(j) = j \neq i$, party i also does not control the expropriation asset for target j. Hence $i \notin \mathrm{Arm}_j(g)$ and $\tau_{ij}^{\mathrm{safe}}(g) = 1$ for all $i \neq j$. It follows that for every project k, $\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\mathrm{safe}}(g) = 1$, so again $TS(g) = V_{\mathrm{total}}$. □

The intuition behind Proposition 4 is as follows. When cooperation requires only alienable actions and ownership can be arranged so that each action that depends on party i's information is controlled by party i, then no private information needs to be disclosed. Parties can therefore avoid expropriation because they can remain silent while still implementing the required cooperative decisions. This case shows how unbundling production decision rights can achieve the first best in Example 3 (ownership as control over production).

When expropriation requires access to assets and ownership can be arranged so that each party controls the asset needed to expropriate *its own* information, then parties can block all expropriation attempts. In that case, the first best is achievable because the means of expropriation are held by the original knowledge holder rather than the potential expropriator. This case shows how unbundling production and expropriation decision rights can achieve the first best in Example 4 (ownership as control over expropriation).

5.5 Optimal Decision Rule Under Arbitrary Expropriation Costs

In the baseline model, we assume successful expropriation is prohibitively costly, $\lambda_{ij} - b_{ij} > V_{\mathrm{total}}$, so the mediator always prefers to deter expropriation. Here we provide intuition for the general results under arbitrary expropriation costs; Appendix A.5 provides formal details.

When expropriation is costly but not prohibitive, the mediator faces the following tradeoff. Information is "free" up to a party's safety threshold, but pushing accuracy beyond that point entails expected expropriation costs. At the same time, production is O-ring: accuracy for one task is wasted in states where some other required tasks fail. Thus, the mediator wants to "buy" accuracy only when it can be coordinated into joint success.

The general solution admits a two-step characterization: (i) *within-party design*, and (ii) *across-party coupling*. The first step determines for each party i how informative i's private recommendations are across targets (subject to expropriation obedience). This step pins down what i can achieve *on its own* — in particular, the *marginal* distribution of i's success across projects. In the second step, given the induced marginals from Step (i), the mediator chooses how to correlate different parties' project successes to maximize joint (O-ring) output.

We explain these steps in the reverse order: conditional on Step (i), Step (ii) is an optimal transport problem with no incentive constraints. Because recommendations are private, correlating parties' latent randomization affects outcomes but does not change any party's posterior or incentives. Step (ii) can be interpreted as moving probability mass toward states of *simultaneous* success, and Step (i) can then be interpreted as choosing the "supply" and "demand" that feed into that transport problem.

To build intuition, consider a case with two parties $i \in \{1, 2\}$ and two projects $k \in \mathcal{K} = \{1, 2\}$. Assume that both parties are required for both projects. The general case with an arbitrary number of projects, parties, and matching requirements is described in Appendix A.5.

Step (ii): Across-Party Coupling. Any obedient decision rule induces, for each party $i \in \{1, 2\}$, a random *success type* $T_i \subseteq \mathcal{K} = \{1, 2\}$, where $k \in T_i$ means that party i successfully completes *all* tasks required of i for project k, while $k \notin T_i$ means that i fails in at least one of the tasks in project k. Thus, T_i can take four values: $T_i \in \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. For notational convenience, denote $00 := \emptyset$, $10 := \{1\}$, $01 := \{2\}$, and $11 := \{1, 2\}$.

Let the *marginal* distribution of T_i be $\beta_i(h) := \mathbb{P}(T_i = h)$ for $h \in \{00, 10, 01, 11\}$. Let the *joint* distribution (a coupling) be $\gamma(h_1, h_2) := \mathbb{P}(T_1 = h_1, T_2 = h_2)$, which can be represented as a 4×4 matrix, so the row sums of γ are β_1 and the column sums are β_2.

Conditional on (h_1, h_2), project k succeeds if and only if $k \in h_1 \cap h_2$. Hence, the cooperative payoff from (h_1, h_2) is $R(h_1, h_2) := V_1 \cdot \mathbf{1}\{1 \in h_1 \cap h_2\} + V_2 \cdot \mathbf{1}\{2 \in h_1 \cap h_2\}$, which can be represented as the following matrix:

$$R(h_1, h_2) := \begin{array}{c|cccc} h_1 \backslash h_2 & 00 & 10 & 01 & 11 \\ \hline 00 & 0 & 0 & 0 & 0 \\ 10 & 0 & V_1 & 0 & V_1 \\ 01 & 0 & 0 & V_2 & V_2 \\ 11 & 0 & V_1 & V_2 & V_1 + V_2 \end{array} .$$

Fix the marginals (β_1, β_2). The mediator's best possible cooperative surplus is obtained by choosing γ to solve

$$\max_{\gamma \geqslant 0} \sum_{h_1, h_2} \gamma(h_1, h_2)\, R(h_1, h_2) \quad \text{s.t.} \quad \sum_{h_2} \gamma(h_1, h_2) = \beta_1(h_1), \quad \sum_{h_1} \gamma(h_1, h_2) = \beta_2(h_2),$$

which is a finite Kantorovich optimal transport problem: β_1 is "supply," β_2 is "demand," and γ is the "shipping plan." Because $R(11, 11) = V_1 + V_2$ is the largest entry, the transport optimizer tries to match "strong" types with "strong" types (positive assortative matching), moving mass toward the diagonal (especially $(11, 11)$), subject to the row and column sums. Intuitively, Step (ii) chooses *correlation*: holding fixed what each party can achieve on its own (the marginals), the mediator aligns success profiles to maximize simultaneous project success.

Step (i): Within-Party Design. In our setting, the marginals (β_1, β_2) are not exogenous:

they are chosen via information design and the expropriation tradeoff. At a high level, the mediator chooses how often party i is recommended to attempt expropriation and how accurate i's recommendations are conditional on that expropriation pattern. Obedience generates a "free" region and a "paid" region: if i is not recommended to expropriate a target j, posterior accuracy about j cannot exceed the safety threshold $\tau_{ij}^{\text{safe}}(g)$; if i is recommended to attempt expropriation, accuracy must be between $\tau_{ij}^{\text{safe}}(g)$ and 1. Thus, increasing accuracy shifts β_i away from 00 and toward 10, 01, and 11, but doing so beyond safety thresholds requires incurring expected expropriation costs.

Optimal Decision Rule. To sum up, in Step (i), the mediator chooses the distribution of expropriation attempt patterns (which targets each party attempts to expropriate and how often) and the accuracy of recommendations (the probability that they are correct) conditional on an attempt pattern. In Step (ii), the mediator chooses the correlation of success across parties (γ) conditional on marginal success (β_i) determined in Step (i).

In optimal transport language, Step (i) chooses the "supply" and "demand" by reshaping each β_i (incurring expropriation costs when necessary), and Step (ii) then solves the transport problem with the payoff matrix R to align these marginals across parties. For an arbitrary number of projects and parties, Step (ii) is a multi-marginal optimal transport problem.

When expropriation is sufficiently costly (the baseline model), it is optimally deterred, so accuracies are capped at safety thresholds. Then, in an optimal decision rule, each party's feasible success profiles form a chain (a totally ordered set): in the two-project case, the support of T_i is either $\{00, 01, 11\}$ or $\{00, 10, 11\}$ (one of 10 and 01 has zero probability). This ranking simplifies Step (ii), as the optimal coupling γ admits a one-dimensional comonotone representation: there exists a common $\xi \sim \text{Unif}[0, 1]$ such that each T_i is a deterministic function of ξ. Visually, the support of the optimal coupling traces a one-dimensional monotone path through matrix R. This insight generalizes to cases with many projects and parties.

With arbitrary expropriation costs, the mediator may optimally randomize over expropriation attempt patterns. Conditional on a fixed pattern (i.e., the set of targets for which party i is recommended expropriation), it is still optimal to make task-level correctness comonotone within the party, so the *conditional* T_i remains a chain. However, different expropriation patterns can induce different chain orderings across projects, and mixing across patterns can therefore generate crossing types (both 10 and 01 with positive probability).

A simple sufficient condition that rules out this mixing problem is an ordered (nested) production technology: if projects can be indexed so that $\mathcal{Q}_1(g) \supseteq \mathcal{Q}_2(g) \supseteq \cdots \supseteq \mathcal{Q}_{|\mathcal{K}|}(g)$ (Project 1 is the "hardest" and Project $|\mathcal{K}|$ is the "easiest"), then the induced ordering of project success is the same across all attempt patterns, so the unconditional T_i retains a totally ordered support. In this case, the optimal coupling γ again admits a one-dimensional comonotone (single-uniform) representation; without such monotonicity, crossing types can be optimal and the full multi-marginal transport problem is required.

Optimal Ownership. Up to this point, we have described how to find the optimal decision rule for a given ownership structure $g \in \mathcal{N}$. Once this rule is characterized, the optimal owner is the party that maximizes total surplus. A key difference relative to the baseline model is that this assignment may now depend on the absolute social cost of expropriation.

To build intuition, consider the extension of Example 3 (where an owner requires a non-owner's knowledge for production) under arbitrary expropriation costs: two parties $\mathcal{N} = \{1, 2\}$ and a single project $\mathcal{K} = \{1\}$. The project requires only alienable actions ($\mathcal{R}_1 = \{1, 2\}$ and $\bar{\mathcal{R}}_1 = \varnothing$), and each party has the ability to expropriate ($\mathcal{E} = \varnothing$).

Since the non-owner does not participate in production, only the owner must be informed about the non-owner's information. This problem can be analyzed using Bayesian Persuasion techniques as in Example 1. Applying the logic of Figure 1, the optimal decision rule takes one of two forms: either partial disclosure up to a safety threshold (optimal when expropriation costs are high) or full disclosure (optimal when expropriation costs are low). Thus, the maximum total surplus under an owner g is

$$TS(g) = \max\{\tau_g^\star \cdot V_1, V_1 - (\lambda_g - b_g)\}$$

In the baseline model, the social cost of expropriation $\lambda_g - b_g$ is high for both $g \in \{1, 2\}$. Consequently, the second term is always smaller, and the optimal owner is the party with the lower temptation to expropriate (higher $\tau_g^\star$).

Suppose now that party 1 has a higher temptation to expropriate ($\tau_1^\star < \tau_2^\star$), but its expropriation is "cheaper" ($\lambda_1 - b_1 < V_1 < \lambda_2 - b_2$). If $\lambda_1 - b_1$ is sufficiently close to zero, it becomes optimal to assign the asset to party 1 — enabling full efficiency at a small cost — even though it has a higher temptation to expropriate.

6 Conclusion

This paper studies how knowledge is allocated across parties when it can be used both productively and opportunistically. In the absence of contractual tools to prevent opportunistic behavior, ownership determines who must be informed and who has the means to expropriate. However, reallocating ownership is an imperfect substitute for complete contracts: it moves a coarse bundle of decision rights rather than the targeted control needed to achieve efficient cooperation without expropriation risk. This framework helps explain a number of industry practices, such as organizing knowledge pooling through common partners, spinning off suppliers to facilitate greater knowledge disclosure from rivals, and the theoretical possibility of Chinese Walls within organizations.

This framework also suggests a few directions for future work. First, our model takes a step toward studying how transactions across firms are shaped by their internal structures. In particular, we provide a simple condition on the production technology that makes internal "Chinese Walls" within organizations impossible. However, even if such walls are theoretically possible, their credibility still depends on governance and internal incentives. Incorporating richer organizational structure into this framework is a natural next step.

Second, in our mechanism-design formulation, parties commit to a mediated contract before observing their private information. If contracting is instead initiated by an informed party, the contract offer itself might convey information. In standard economic models, private information is usually represented as the realization of a random variable observed by some parties but not others, drawn from a common-knowledge distribution. In such settings, information can in principle be concealed through the design of the mechanism — an insight known as the *inscrutability principle* (Myerson, 1983).

The inscrutability logic can fail when information is "eye-opening," in the sense that it changes how parties perceive the relevant state space or feasible actions (Tirole, 2009). Then the mere act of specifying previously unforeseen contingencies expands the other party's awareness and shifts beliefs. Even if parties can contractually rule out expropriation, the disclosure required to initiate contracting changes outside options relative to no contracting, so the initial ownership structure can shape knowledge flows even when expropriation is contractible. Formalizing this setting likely requires an alternative model of information (e.g., Akerlof, Holden, and Li, 2025) and is a promising direction for future research.

References

Akerlof, R., Holden, R., & Li, H. (2025). Getting the Picture. *Working Paper*.

Alonso, R., Dessein, W., & Matouschek, N. (2008). When Does Coordination Require Centralization? *American Economic Review*, *98*(1), 145–179.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In R. Nelson (Ed.), *The Rate and Direction of Incentive Activity: Economic and Social Factors* (pp. 609–626). Princeton University Press.

Atalay, E., Hortaçsu, A., & Syverson, C. (2014). Vertical Integration and Input Flows. *American Economic Review*, *104*(4), 1120–1148.

Bai, J., Barwick, P. J., Cao, S., & Li, S. (2025). Quid Pro Quo, Knowledge Spillovers, and Industrial Quality Upgrading: Evidence from the Chinese Auto Industry. *American Economic Review*, *115*(11), 3825–3852.

Baliga, S., & Sjöström, T. (2018). A Theory of the Firm Based on Haggling, Coordination, and Rent-Seeking.

Bergemann, D., & Morris, S. (2016). Bayes Correlated Equilibrium and the Comparison of Information Structures in Games. *Theoretical Economics*, *11*(2), 487–522.

Bergemann, D., & Morris, S. (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, *57*(1), 44–95.

Bradsher, K. (1998). G.M. Plans to Spin Off Parts Division. *The New York Times*.

CNN Money. (1999, May). Delphi Spinoff Completed.

Demirer, M., & Karaduman, Ö. (2025). Do Mergers and Acquisitions Improve Efficiency? Evidence from Power Plants. *Working Paper*.

Dessein, W. (2002). Authority and Communication in Organizations. *Review of Economic Studies*, *69*(4), 811–838.

Fitch, A. (2025). Is Nvidia Intel's Savior? Not Quite. *The Wall Street Journal*.

Forges, F. (1986). An Approach to Communication Equilibria. *Econometrica*, *54*(6), 1375–1385.

Gawer, A., & Cusumano, M. A. (2002). *Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation*. Harvard Business School Press.

Grossman, S., & Hart, O. (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy*, *94*(4), 691–719.

Hart, O. (1995). *Firms, Contracts, and Financial Structure (Clarendon Lectures in Economics)*. Oxford University Press.

Hart, O., & Moore, J. (1990). Property Rights and the Nature of the Firm. *Journal of Political Economy*, *98*(6), 1119–1158.

Isaacson, W. (2011). *Steve Jobs*. Simon & Schuster.

Kamenica, E., & Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, *101*(6), 2590–2615.

Klemperer, P., Cramton, P., & Gibbons, R. (1987). Dissolving a Partnership Efficiently. *Econometrica*, *55*(3), 615–632.

Kolotilin, A., & Li, H. (2021). Relational Communication. *Theoretical Economics*, *16*(4), 1391–1430.

Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, *108*, 551–575.

Kreps, D. M., & Wilson, R. (1982). Sequential Equilibria. *Econometrica*, *50*(4), 863–894.

Küpper, D., Okur, A., Betti, F., Bezamat, F., Fendri, M., & Fernandez, B. (2020). How Manufacturers Can Unlock Value from Data Sharing.

Liebeskind, J. (1996). Knowledge, Strategy, and the Theory of the Firm. *Strategic Management Journal*, *17*, 93–107.

Marshall, R. C., & Marx, L. M. (2007). Bidder Collusion. *Journal of Economic Theory*, *133*(1), 374–402.

Martin, T. W., & Mickle, T. (2017). Why 'Apple Rival' Samsung Also Wins If iPhone X Is a Hit. *The Wall Street Journal*.

Mathevet, L., Perego, J., & Taneva, I. (2020). On Information Design in Games. *Journal of Political Economy*, *128*(4), 1370–1404.

Matouschek, N. (2004). Ex Post Inefficiencies in a Property Rights Theory of the Firm. *The Journal of Law, Economics, and Organization*, *20*(1), 125–147.

McGee, P. (2025). *Apple in China: The Capture of the World's Greatest Company*. Simon; Schuster.

Myerson, R. (1983). Mechanism Design by an Informed Principal. *Econometrica*, *51*(6), 1767–1797.

Ortner, J., Sugaya, T., & Wolitzky, A. (2024). Mediated Collusion. *Journal of Political Economy*, *132*(4), 1247–1289.

Rantakari, H. (2008). Governing Adaptation. *Review of Economic Studies*, *75*(4), 1257–1285.

Rayo, L., & Segal, I. (2010). Optimal Information Disclosure. *Journal of Political Economy*, *118*(5), 949–987.

Segal, I., & Whinston, M. D. (2016). Property Rights and the Efficiency of Bargaining. *Journal of the European Economic Association*, *14*(6), 1287–1318.

Sugaya, T., & Wolitzky, A. (2018). Maintaining Privacy in Cartels. *Journal of Political Economy*, *126*(6), 2569–2607.

Teece, D. (1980). Economies of Scope, and the Scope of the Enterprise. *Journal of Economic Behavior and Organization*, *1*(3), 223–247.

Tirole, J. (2009). Cognition and Incomplete Contracts. *American Economic Review*, *99*(1), 265–294.

A Appendix

A.1 Proof of Proposition 1

Recall that $\mathcal{R}_k$ and $\bar{\mathcal{R}}_k$ are alienable and inalienable requirements, respectively, and $\mathcal{Q}_k(g) := \{(g,j) : j \in \mathcal{R}_k\} \cup \bar{\mathcal{R}}_k \subseteq \mathcal{N} \times \mathcal{N}$ is the set of production requirements under owner $g \in \mathcal{N}$. Proposition 1 states that the maximum attainable expected total surplus under owner $g \in \mathcal{N}$ is

$$TS(g) = \sum_{k \in \mathcal{K}} V_k \cdot \left(\min_{(i,j) \in \mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) \right). \tag{A.1}$$

Moreover, there exists an optimal obedient decision rule that is *nested*. Let $\xi \sim \text{Unif}[0,1]$ be a common randomization device. For each (ξ, θ, g) define an action profile $a(\xi, \theta, g) = (x, \bar{x}, z, \bar{z})$ as follows:

$$z_{ij} = 0 \ \text{ and } \ \bar{z}_{ij} = 0 \qquad \text{for all } i \in \mathcal{N}, \ j \neq i, \tag{A.2}$$

$$x_{ij} = \bar{x}_{ij} = \begin{cases} \theta_j & \text{if } \xi \leqslant \tau_{ij}^{\text{safe}}(g), \\ -\theta_j & \text{if } \xi > \tau_{ij}^{\text{safe}}(g), \end{cases} \qquad \text{for all } i \in \mathcal{N}. \tag{A.3}$$

Define $\sigma^g(\cdot \mid \theta) \in \Delta(\mathcal{A})$ as the distribution induced by $\xi \sim \text{Unif}[0,1]$ and $a = a(\xi, \theta; g)$; that is,

$$\sigma^g(a \mid \theta) = \mathbb{P}_\xi\big[a(\xi, \theta; g) = a\big].$$

We prove the proposition in two steps. Lemma 1 shows that this decision rule is obedient and achieves $TS(g)$. Lemma 2 shows that for any obedient decision rule, the expected total surplus does not exceed $TS(g)$.

Lemma 1 (Nested rule is obedient and achieves $TS(g)$)**.** *The decision rule σ^g is obedient under ownership $g \in \mathcal{N}$. Moreover, each project k succeeds with probability $\min_{(i,j) \in \mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g)$, and hence*

$$\mathbb{E}_\psi\Big[TS(a, g; \theta)\Big] = \sum_{k \in \mathcal{K}} V_k \cdot \min_{(i,j) \in \mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) = TS(g),$$

where $\psi(\theta, a) = p(\theta)\sigma^g(a \mid \theta)$ is the induced distribution over actions and states.

Proof. We verify obedience and compute the induced success probabilities.

Posteriors. Party i observes (θ_i, a_i) before choosing actions. The expropriation recommendations in (A.2) are constant and uninformative. Since $x_{ij} = \bar{x}_{ij}$ for all i, j in (A.3), for notational convenience we will keep track of the vector $x_i := (x_{i1}, \ldots, x_{iN})$ only.

Fix i and write $x_{i,-i} := (x_{ij})_{j \neq i}$. For each $j \neq i$, let $s_{ij}(\xi) \in \{-1, 1\}$ be a deterministic sign defined by $s_{ij}(\xi) = 1$ if $\xi \leqslant \tau_{ij}^{\text{safe}}(g)$ and $s_{ij}(\xi) = -1$ otherwise, so that $x_{ij} = s_{ij}(\xi) \cdot \theta_j$. For any vector $v \in \{-1, 1\}^{N-1}$ and any $\xi \in [0, 1]$, independence and uniformity of $(\theta_j)_{j \neq i}$ imply

$$\begin{aligned} \mathbb{P}_\psi(x_{i,-i} = v \mid \xi) &= \mathbb{P}_\psi\big(s_{ij}(\xi) \cdot \theta_j = v_j \ \forall j \neq i \,\big|\, \xi\big) \\ &= \prod_{j \neq i} \mathbb{P}_\psi\big(\theta_j = s_{ij}(\xi) \cdot v_j\big) = \prod_{j \neq i} \tfrac{1}{2} = 2^{-(N-1)}. \end{aligned} \tag{A.4}$$

This conditional probability does *not* depend on ξ. Hence $x_{i,-i}$ is independent of ξ. Moreover, since θ_i is independent of $(\xi, x_{i,-i})$, ξ is also independent of $(\theta_i, x_{i,-i})$. Equivalently, Bayes' rule yields that for any $v \in \{-1, 1\}^{N-1}$, the conditional density of $\xi \mid x_{i,-i}$ is

$$f_{\xi \mid x_{i,-i}}(\xi \mid v) = \frac{\mathbb{P}_\psi(x_{i,-i} = v \mid \xi)\, f_\xi(\xi)}{\mathbb{P}_\psi(x_{i,-i} = v)} = \frac{2^{-(N-1)} \cdot 1}{2^{-(N-1)}} = 1,$$

so $\xi \mid x_{i,-i}$ is uniform on $[0, 1]$. Since θ_i is independent of $(\xi, x_{i,-i})$, we also have $f_{\xi \mid \theta_i, x_{i,-i}}(\cdot \mid \theta_i, x_{i,-i}) := 1$.

Fix $j \neq i$. Under the nested rule, $\theta_j = x_{ij}$ if and only if $\xi \leqslant \tau_{ij}^{\text{safe}}(g)$. Therefore, using iterated expectations and the posterior in (A.4),

$$\begin{aligned} \mathbb{P}_\psi(\theta_j = x_{ij} \mid \theta_i, x_{i,-i}) &= \mathbb{P}_\psi(\xi \leqslant \tau_{ij}^{\text{safe}}(g) \mid \theta_i, x_{i,-i}) \\ &= \int_0^{\tau_{ij}^{\text{safe}}(g)} f_{\xi \mid \theta_i, x_{i,-i}}(\xi \mid \theta_i, x_{i,-i})\, d\xi = \int_0^{\tau_{ij}^{\text{safe}}(g)} 1\, d\xi = \tau_{ij}^{\text{safe}}(g). \end{aligned}$$

No expropriation. The recommendation rule always prescribes no expropriation (A.2). If i is disarmed, then every effective expropriation action is 0 and there is no deviation. If i is armed, then $\tau_{ij}^{\text{safe}}(g) = \tau_{ij}^{\star} = c_{ij}/(b_{ij} + c_{ij})$ by definition, so conditional on (θ_i, a_i) the maximal success probability from attempting expropriation against any target is at most $\tau_{ij}^{\star}$, implying that the expected private gain from expropriation is weakly non-positive. With the tie-breaking rule favoring the safe action, obeying (A.2) is optimal.

Obedient Cooperation. Now consider deviations that change i's cooperative actions $(x_i, \bar{x}_i)$. For any project k, let $J_{i,k}(g)$ denote the set of indices $j \neq i$ such that $(i, j) \in \mathcal{Q}_k(g)$, i.e., the states that i must match (via either x_{ij} if $i = g$ and $j \in \mathcal{R}_k$, or via $\bar{x}_{ij}$ if $(i, j) \in \bar{\mathcal{R}}_k$). If $J_{i,k}(g) = \varnothing$, then i's cooperative action does not affect project k.

Fix a project k where $J_{i,k}(g) \neq \varnothing$. Under the nested rule, if all parties follow their recommendations, project k succeeds if and only if $\xi \leqslant \tau_{ij}^{\text{safe}}(g)$ for all required tasks (i, j), which is equivalent to:

$$\xi \leqslant \min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g). \tag{A.5}$$

Since $\tau_{ij}^{\text{safe}}(g) > 1/2$ for all pairs, this probability is strictly greater than 1/2.

Now suppose party i deviates from its recommendation for a subset of targets $\tilde{J}_{i,k}(g) \subseteq J_{i,k}(g)$. Party i successfully matches the states if and only if the recommendation was correct for the non-deviated targets ($j \notin \tilde{J}_{i,k}(g)$) and incorrect for the deviated targets ($j \in \tilde{J}_{i,k}(g)$). In terms of the randomization variable ξ, this requires:

$$\xi \in \Big(\max_{j\in\tilde{J}_{i,k}(g)} \tau_{ij}^{\text{safe}}(g), \min_{j\in J_{i,k}(g)\setminus\tilde{J}_{i,k}(g)} \tau_{ij}^{\text{safe}}(g) \Big], \quad \text{with the convention } \min\varnothing := 1.$$

Since $\tau_{ij}^{\text{safe}}(g) > 1/2$, the lower bound of this interval is strictly greater than 1/2. Therefore, the length of the interval — and thus the success probability — is strictly less than 1/2. Consequently, any deviation yields a strictly lower probability of success than obedience.

Total surplus. Multiplying the probability of the event in (A.5) by V_k and summing over all projects yields the stated expression for $TS(g)$. □

Lemma 2 (Upper bound for any obedient rule)**.** *Fix ownership $g \in \mathcal{N}$. Let $\sigma(\cdot \mid \theta)$ be any obedient decision rule under g, and let $\psi(\theta, a) = p(\theta)\sigma(a \mid \theta)$ be the induced distribution. Then*

$$\mathbb{E}_\psi [TS(a, g; \theta)] \leqslant TS(g).$$

Proof. For each task $(i, j) \in \mathcal{Q}_k(g)$ define the event that this task is satisfied:

$$C_{k,(i,j)} := \begin{cases} \{(\theta, a) : x_{ij} = \theta_j\} & \text{if } (i, j) = (g, j) \in \{(g, j) : j \in \mathcal{R}_k\} \text{ and } (i, j) \notin \bar{\mathcal{R}}_k, \\ \{(\theta, a) : \bar{x}_{ij} = \theta_j\} & \text{if } (i, j) \in \bar{\mathcal{R}}_k \text{ and } (i, j) \notin \{(g, j) : j \in \mathcal{R}_k\}, \\ \{(\theta, a) : x_{ij} = \bar{x}_{ij} = \theta_j\} & \text{if } (i, j) = (g, j) \in \{(g, j) : j \in \mathcal{R}_k\} \cap \bar{\mathcal{R}}_k. \end{cases} \tag{A.6}$$

Let E_k denote the event that project k succeeds:

$$E_k := \bigcap_{(i,j)\in\mathcal{Q}_k(g)} C_{k,(i,j)}.$$

We write total surplus as

$$TS(a, g; \theta) = TS^{\text{coop}}(a, g; \theta) + TS^{\text{exp}}(a, g; \theta),$$

where $TS^{\text{coop}}(a, g; \theta) = \sum_{k\in\mathcal{K}} V_k \mathbf{1}\{E_k\}$, and TS^{exp} is the sum of all expropriation terms.

Upper Bound on Task Probabilities. Let

$$r_{ij} \coloneqq \mathbb{P}_\psi\left(\zeta_{ij} \neq 0\right)$$

be the probability that party i attempts an expropriation of party j's information, $j \neq i$. If $i \notin \mathrm{Arm}(g)$, then party i cannot engage in expropriation and $r_{ij} = 0$.

Decompose $\mathbb{P}_\psi(C_{k,(i,j)})$ according to whether i attempts to expropriate j:

$$\mathbb{P}_\psi(C_{k,(i,j)}) = \mathbb{P}_\psi(C_{k,(i,j)}, \zeta_{ij} = 0) + \mathbb{P}_\psi(C_{k,(i,j)}, \zeta_{ij} \neq 0).$$

On the event $\{\zeta_{ij} = 0\}$, obedience implies that attempting expropriation is not strictly profitable for party i. Therefore, conditional on (θ_i, a_i), the probability that any sign choice satisfies the matching task is at most $\tau_{ij}^{\text{safe}}(g)$, so

$$\mathbb{P}_\psi\left(C_{k,(i,j)} \mid \zeta_{ij} = 0\right) \leqslant \tau_{ij}^{\text{safe}}(g) \Rightarrow \mathbb{P}_\psi(C_{k,(i,j)}, \zeta_{ij} = 0) \leqslant \tau_{ij}^{\text{safe}}(g)\mathbb{P}_\psi\left(\zeta_{ij} = 0\right) = \tau_{ij}^{\text{safe}}(g)(1 - r_{ij}).$$

In addition,

$$\mathbb{P}_\psi(C_{k,(i,j)}, \zeta_{ij} \neq 0) \leqslant \mathbb{P}_\psi\left(\zeta_{ij} \neq 0\right) = r_{ij}.$$

Combining these two inequalities, we get

$$\mathbb{P}_\psi(C_{k,(i,j)}) \leqslant \tau_{ij}^{\text{safe}}(g)(1 - r_{ij}) + r_{ij}.$$

Upper bound on cooperative surplus. Fix project $k \in \mathcal{K}$. If $\mathcal{Q}_k(g)$ contains only trivial tasks of the form (i,i), then $\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) = 1$, so the desired bound holds.

Fix any project k with non-trivial tasks: $(i,j) \in \mathcal{Q}_k(g)$ with $i \neq j$. Pick one bottleneck task:

$$(i_k, j_k) \in \arg\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g) \quad \text{with } j_k \neq i_k.$$

Since $E_k \subseteq C_{k,i_k,j_k}$, we have

$$\mathbb{P}_\psi(E_k) \leqslant \tau_{i_k j_k}^{\text{safe}}(g) + \big(1 - \tau_{i_k j_k}^{\text{safe}}(g)\big) r_{i_k j_k},$$

Multiplying by V_k and summing over projects yields

$$\mathbb{E}_\psi[TS^{\text{coop}}] \leqslant TS(g) + \sum_{i\in\mathrm{Arm}(g)} \sum_{j\neq i} \big(1 - \tau_{ij}^{\text{safe}}(g)\big) V_{ij}^{\min} r_{ij}, \tag{A.7}$$

where

$$V_{ij}^{\min} := \sum_{k:\ (i_k,j_k)=(i,j)} V_k.$$

Upper bound on expropriation surplus. Fix a pair (i,j) with $i \in \mathrm{Arm}(g)$ and $j \neq i$. On the event $\{\zeta_{ij} \neq 0\}$, obedience implies that attempting expropriation is (weakly) optimal, hence the conditional success probability is at least the private threshold:

$$\mathbb{P}_\psi(\zeta_{ij} = \theta_j \mid \zeta_{ij} \neq 0) \;\geqslant\; \tau_{ij}^\star.$$

Therefore the expected welfare contribution from this ordered pair satisfies

$$\mathbb{E}_\psi\Big[(b_{ij} - \lambda_{ij})\,\mathbf{1}\{\zeta_{ij} = \theta_j\} - c_{ij}\,\mathbf{1}\{\zeta_{ij} = -\theta_j\}\Big] \leqslant \mathbb{E}_\psi\Big[(b_{ij} - \lambda_{ij})\,\mathbf{1}\{\zeta_{ij} = \theta_j\}\Big] \leqslant (b_{ij} - \lambda_{ij})\,\tau_{ij}^\star\, r_{ij},$$

where the first inequality follows from $-c_{ij} < 0$ and the second inequality follows from the fact that the probability of successful expropriation should be at least $\tau_{ij}^\star$. Summing over (i,j) gives

$$\mathbb{E}_\psi[TS^{\mathrm{exp}}] \leqslant \sum_{i \in \mathrm{Arm}(g)} \sum_{j \neq i} \tau_{ij}^\star\,(b_{ij} - \lambda_{ij})\, r_{ij}. \tag{A.8}$$

Upper bound on total surplus. Combining (A.7) and (A.8) gives

$$\begin{aligned}\mathbb{E}_\psi[TS(a,g;\theta)] &\leqslant TS(g) + \sum_{i \in \mathrm{Arm}(g)} \sum_{j \neq i} r_{ij}\Big[(1-\tau_{ij}^\star)V_{ij}^{\min} + \tau_{ij}^\star(b_{ij} - \lambda_{ij})\Big] \\ &= TS(g) - \sum_{i \in \mathrm{Arm}(g)} \sum_{j \neq i} r_{ij}\Big[\underbrace{\tau_{ij}^\star(\lambda_{ij} - b_{ij}) - \big(1-\tau_{ij}^\star\big)V_{ij}^{\min}}_{>0}\Big].\end{aligned}$$

By Assumption 1, $\lambda_{ij} - b_{ij} > V_{\mathrm{total}} \geqslant V_{ij}^{\min}$. Also $b_{ij} < c_{ij}$ implies $\tau_{ij}^\star = \frac{c_{ij}}{b_{ij}+c_{ij}} > \frac{1}{2}$, hence $\tau_{ij}^\star > (1-\tau_{ij}^\star)$. Therefore, for every $i \in \mathrm{Arm}(g)$,

$$\tau_{ij}^\star(\lambda_{ij} - b_{ij}) \;>\; \tau_{ij}^\star V_{ij}^{\min} \;>\; (1-\tau_{ij}^\star)V_{ij}^{\min},$$

so the bracketed term is strictly positive. Therefore, the maximum attainable expected total surplus is $TS(g)$ and can be attained only when there is no expropriation, $r_{ij} = 0$ for all $i \in \mathcal{N}$. □

Conclusion. Lemma 1 shows $TS(g)$ is attainable by the nested no-expropriation rule. Lemma 2 shows no obedient rule can exceed it under Assumption 1. This completes the proof of Proposition 1.

A.2 Implementation Through a Mediated Contract

This appendix proves the results stated in Section 5.1. We first analyze the communication game and prove Proposition 2. We then analyze the contracting game and show how the outcome from the baseline solution concept can be supported as a Perfect Bayesian Equilibrium (PBE).

A.2.1 Communication Game and Proof of Proposition 2

Incentive Constraints. Fix a direct mechanism (φ, t). When all parties report truthfully and follow their recommendations, party i's interim expected utility is

$$U_i(\varphi, t \mid \theta_i) \;:=\; \sum_{\theta_{-i}\in\Theta_{-i}} p(\theta_{-i} \mid \theta_i) \sum_{a\in\mathcal{A}} \sum_{g\in\mathcal{N}} \varphi(a, g \mid \theta_i, \theta_{-i}) \Big[u_i(a, g; \theta_i, \theta_{-i}) + t_i\big((a, g), (\theta_i, \theta_{-i})\big)\Big].$$

Parties are not compelled to follow recommendations; a (pure) deviation for party i is a function $\delta_i : D_i \to \mathcal{A}_i$ mapping the available information at the action stage into an action plan, where $D_i := \mathcal{N} \times \mathcal{T} \times \mathcal{A}_i \times \Theta_i$, and party i observes $d_i := (g, t, a_i, \theta_i) \in D_i$. For any alternative report $\theta_i' \in \Theta_i$ and any deviation function $\delta_i : D_i \to \mathcal{A}_i$, define the *double-deviation* payoff

$$\begin{aligned} U_i^\star(\varphi, t, \delta_i, \theta_i' \mid \theta_i) \;:=\; \sum_{\theta_{-i}\in\Theta_{-i}} p(\theta_{-i} \mid \theta_i) \sum_{a\in\mathcal{A}} \sum_{g\in\mathcal{N}} \varphi(a, g \mid \theta_i', \theta_{-i}) \Big[u_i\big((\delta_i(g, t', a_i, \theta_i), a_{-i}), g; \theta_i, \theta_{-i}\big) \\ + t_i\big((a, g), (\theta_i', \theta_{-i})\big)\Big], \end{aligned}$$

where $t' := t((a, g), (\theta_i', \theta_{-i}))$ is the transfer vector implemented after the report (θ_i', θ_{-i}).

The contract (φ, t) is *incentive compatible* if, for every $i \in \mathcal{N}$,

$$U_i(\varphi, t \mid \theta_i) \;\geqslant\; U_i^\star(\varphi, t, \delta_i, \theta_i' \mid \theta_i) \qquad \text{for all } \theta_i, \theta_i' \in \Theta_i \text{ and all } \delta_i : D_i \to \mathcal{A}_i. \tag{A.9}$$

This condition subsumes truth-telling (take $\delta_i(d_i) = a_i$) and obedience (take $\theta_i' = \theta_i$) as special cases.

Communication-Game Problem. The maximal expected total surplus attainable in the communication game is

$$\begin{aligned} TS(\varphi, t) := \max_{\varphi, t} \quad & \sum_{\theta\in\Theta} p(\theta) \sum_{g\in\mathcal{N}} \sum_{a\in\mathcal{A}} \varphi(a, g \mid \theta)\, TS(a, g; \theta) \\ \text{subject to} \quad & \text{(A.9)}, \\ & \varphi(a, g \mid \theta) \geqslant 0, \qquad \sum_{a,g} \varphi(a, g \mid \theta) = 1 \qquad \text{for all } \theta \in \Theta, \\ & \sum_{i\in\mathcal{N}} t_i(a, g, \theta) = 0 \qquad \text{for all } (a, g, \theta) \in \mathcal{A} \times \mathcal{N} \times \Theta. \end{aligned} \tag{A.10}$$

Relaxed Problem: Obedience Only. Consider the relaxation of (A.10) obtained by dropping the truth-telling and the double-deviation constraints. In this relaxation, we evaluate incentives under the *truthful* report profile and impose only obedience of the recommended non-contractible actions.

Because transfers are implemented before actions are chosen and do not depend on realized non-contractible actions, they do not affect obedience. The obedience constraints are: for every $i \in \mathcal{N}$, every $\theta_i \in \Theta_i$, every $g \in \mathcal{N}$, every recommended action plan $a_i \in \mathcal{A}_i$, and every deviation $a_i' \in \mathcal{A}_i$,

$$\sum_{\theta_{-i}} p(\theta_{-i} \mid \theta_i) \sum_{a_{-i}} \varphi\left(a_i, a_{-i}, g \mid \theta_i, \theta_{-i}\right) \left[u_i\left(a_i, a_{-i}, g; \theta\right) - u_i\left(a_i', a_{-i}, g; \theta\right)\right] \geqslant 0. \tag{A.11}$$

Thus, the relaxed problem is

$$\max_{\varphi} \quad \sum_{\theta \in \Theta} p(\theta) \sum_{g \in \mathcal{N}} \sum_{a \in \mathcal{A}} \varphi(a, g \mid \theta)\, TS(a, g; \theta) \tag{A.12}$$

$$\text{subject to} \quad \text{(A.11)}, \qquad \varphi(a, g \mid \theta) \geqslant 0, \quad \sum_{a,g} \varphi(a, g \mid \theta) = 1 \;\; \text{for all } \theta \in \Theta.$$

For any fixed ownership assignment $g \in \mathcal{N}$, if we restrict attention to allocation rules that put probability one on g, then (A.12) coincides with the benchmark mediator problem (2.2) defining $TS(g)$.

Lemma 3 (Deterministic ownership)**.** *In the communication-game problem* (A.12)*, allowing randomization of ownership g does not increase the maximal attainable total surplus. There exists an optimal solution with deterministic ownership.*

Proof. Let φ be any feasible allocation rule for (A.12), and let $G \in \mathcal{N}$ denote the (possibly random) owner induced by φ (which is publicly observed before actions are chosen). Let $\psi(a, g, \theta) := p(\theta)\varphi(a, g \mid \theta)$ denote the joint distribution of (a, g, θ) induced by φ.

Fix any $g \in \mathcal{N}$ with $\mathbb{P}_\psi(G = g) > 0$, and define the conditional joint distribution of actions and states given $G = g$ by

$$\psi^g(a, \theta) := \mathbb{P}_\psi(a, \theta \mid G = g) = \frac{\psi(a, g, \theta)}{\mathbb{P}_\psi(G = g)}.$$

Because the obedience constraints (A.11) are imposed separately for each ownership realization g, and because parties observe $G = g$ before choosing actions, the conditional distribution ψ^g is obedient under owner g. Therefore, by Lemma 2,

$$\mathbb{E}_{\psi^g}\big[TS(a, g; \theta)\big] \leqslant TS(g).$$

Taking expectations over G yields

$$\mathbb{E}_\psi[TS(a,G;\theta)] = \sum_{g\in\mathcal{N}} \mathbb{P}_\psi(G=g)\,\mathbb{E}_{\psi^g}[TS(a,g;\theta)] \leqslant \sum_{g\in\mathcal{N}} \mathbb{P}_\psi(G=g)\,TS(g) \leqslant \max_{g\in\mathcal{N}} TS(g).$$

Finally, choosing deterministic ownership $G \equiv g^* \in \arg\max_{g\in\mathcal{N}} TS(g)$ and an optimal obedient decision rule that attains $TS(g^*)$ achieves the upper bound. Hence an optimal solution to (A.12) exists with deterministic ownership. □

The solution to the relaxed (obedience-only) problem in (2.2) with deterministic ownership provides an upper bound for the communication-game problem (A.10). Let σ^* be an optimal decision rule for the problem in (2.2), and let g^* denote an optimal owner, $g^* \in \arg\max_g TS(g)$.

Lemma 4 (Constant-transfer implementation)**.** *Fix $g^* \in \arg\max_{g\in\mathcal{N}} TS(g)$ and let σ^* be an optimal obedient decision rule under owner g^*. Then for any constant budget-balanced transfer vector $\bar{t} \in \mathcal{T}$, the direct mediated mechanism $(\varphi^*, \bar{t})$ defined by*

$$\varphi^*(a,g \mid \theta) = \begin{cases} \sigma^*(a \mid \theta), & \text{if } g = g^*, \\ 0, & \text{if } g \neq g^*, \end{cases} \tag{A.13}$$

is incentive compatible.

Proof. Because transfers are constant, they cancel from all report and deviation comparisons. Fix a party i, a true type θ_i, an alternative report $\hat{\theta}_i \in \Theta_i$, and an arbitrary action-stage deviation $\delta_i : D_i \to \mathcal{A}_i$.

Step 1: Upper Bound for a Single Task. Fix any target $j \neq i$. Under truthful reporting by party j, we have $\hat{\theta}_j = \theta_j$, and party i receives a recommendation

$$x_{ij} = \bar{x}_{ij} = \begin{cases} \theta_j, & \xi \leqslant \tau_{ij}^{\text{safe}}(g^*), \\ -\theta_j, & \xi > \tau_{ij}^{\text{safe}}(g^*). \end{cases} \tag{A.14}$$

In Lemma 1, we proved that the posterior belief that $\theta_j = x_{ij}$ under (A.14) is equal to $\tau_{ij}^{\text{safe}}(g^*)$. Since $\tau_{ij}^{\text{safe}}(g^*) \geqslant \frac{1}{2}$, x_{ij} is a posterior mode for θ_j given d_i. Thus, among all choices of i's productive action for target j, the maximal achievable success probability in task (i,j) equals $\tau_{ij}^{\text{safe}}(g^*)$. Therefore, for any deviation δ_i and any report $\hat{\theta}_i$, the probability that task (i,j) in project k succeeds (event $C_{k,(ij)}$ defined in (A.6)) has the following upper bound:

$$\mathbb{P}(C_{k,(ij)} \mid \hat{\theta}_i, \delta_i, \theta_i) \;\leqslant\; \tau_{ij}^{\text{safe}}(g^*) \qquad \text{for all } j \neq i. \tag{A.15}$$

A similar bound applies to tasks performed by *other* parties who obey. Fix $\ell \neq i$ and $j \neq \ell$. Since ℓ is held to obey equilibrium recommendations, its match success satisfies

$$\mathbb{P}(C_{k,(\ell j)} \mid \hat{\theta}_i, \delta_i, \theta_i) = \begin{cases} \tau_{\ell j}^{\text{safe}}(g^*), & j \neq i \text{ (since } \hat{\theta}_j = \theta_j), \\ \tau_{\ell i}^{\text{safe}}(g^*), & j = i, \ \hat{\theta}_i = \theta_i, \\ 1 - \tau_{\ell i}^{\text{safe}}(g^*), & j = i, \ \hat{\theta}_i = -\theta_i. \end{cases}$$

Furthermore, since $\tau_{\ell i}^{\text{safe}}(g^*) \geqslant \frac{1}{2}$, in all cases we have

$$\mathbb{P}(C_{k,(\ell j)} \mid \hat{\theta}_i, \delta_i, \theta_i) \leqslant \tau_{\ell j}^{\text{safe}}(g^*). \tag{A.16}$$

Step 2: Upper Bound on Project Success. For each project k, success requires all required tasks to succeed: $E_k = \bigcap_{(\ell,j)\in\mathcal{Q}_k(g^*)} C_{k,(\ell j)}$. Hence for any $\hat{\theta}_i, \delta_i$,

$$\mathbb{P}(E_k \mid \hat{\theta}_i, \delta_i, \theta_i) \leqslant \min_{(\ell,j)\in\mathcal{Q}_k(g^*)} \mathbb{P}(C_{k,(\ell j)} \mid \hat{\theta}_i, \delta_i, \theta_i) \leqslant \min_{(\ell,j)\in\mathcal{Q}_k(g^*)} \tau_{\ell j}^{\text{safe}}(g^*).$$

Under truthful reporting and obedience, each required task (ℓ, j) succeeds if and only if $\xi \leqslant \tau_{\ell j}^{\text{safe}}(g^*)$, so (because the same ξ is shared across tasks)

$$\mathbb{P}(E_k \mid \text{truth, obey}, \theta_i) = \min_{(\ell,j)\in\mathcal{Q}_k(g^*)} \tau_{\ell j}^{\text{safe}}(g^*).$$

That is, truthful reporting and obedience attain the upper bound for every project.

Since $u_i^c = \sum_{k\in\mathcal{K}} \alpha_{ik} V_k \, \mathbf{1}\{E_k\}$ with $\alpha_{ik} V_k \geqslant 0$, maximizing each $\mathbb{P}(E_k)$ implies that truthful reporting and obedience maximize i's expected cooperative payoff.

No Expropriation. Under σ^* the posterior accuracy about any target j is capped at $\tau_{ij}^{\text{safe}}(g^*)$, so by the definition of the safety threshold (see Lemma 1), any expropriation attempt yields a weakly lower expected payoff than $\zeta_{ij} = 0$. The expropriation payoff is unaffected by i's report because i's information about θ_{-i} is independent of $\hat{\theta}_i$.

Conclusion. Combining the cooperative and expropriation parts yields the double-deviation IC constraints (A.9). □

Proposition 2. Lemma 3 shows that $TS(g^*)$ provides an upper bound on the maximal surplus in the communication game. Lemma 4 shows that this surplus is attainable. Together, these results complete the proof of Proposition 2.

A.2.2 Contracting Game

We now embed the communication game into a broader bargaining environment. Fix an initial owner $g_0 \in \mathcal{N}$. A *contract* is a direct mediated mechanism (φ, t). The contracting game adds two stages before the communication game:

1. Party 1 makes a take-it-or-leave-it offer of a contract (φ, t).

2. Each party $i \in \{2, \ldots, N\}$ either accepts or rejects.

If all parties accept, the communication game induced by (φ, t) is played. If at least one party rejects, the contract is void, the owner remains g_0, transfers are zero, no information is shared, and the disagreement outcome is played.

Let (a^0, g_0) denote the disagreement outcome, where $x_{ij} = \bar{x}_{ij} = 1$ and $z_{ij} = \bar{z}_{ij} = 0$ for all feasible pairs (i, j). Let

$$\underline{U}_i := \sum_{\theta \in \Theta} p(\theta)\, u_i(a^0, g_0; \theta)$$

denote party i's ex-ante disagreement payoff. Let

$$\bar{U}_i^* := \sum_{\theta \in \Theta} p(\theta) \sum_{a \in \mathcal{A}} \sigma^*(a \mid \theta)\, u_i(a, g^*; \theta)$$

denote party i's expected non-transfer payoff under the optimal benchmark rule. Define the constant transfer vector $t^* \in \mathcal{T}$ by

$$t_i^* = \underline{U}_i - \bar{U}_i^* \quad (i \neq 1), \qquad t_1^* = -\sum_{i \neq 1} t_i^*.$$

Let (φ^*, t^*) denote the associated mediated contract, where φ^* is defined in (A.13).

Proposition 5 (PBE implementation of the optimal contract)**.** *There exists a Perfect Bayesian Equilibrium of the full bargaining game in which:*

1. *Party 1 offers the contract (φ^*, t^*).*

2. *Every party $i \neq 1$ accepts.*

3. *Conditional on acceptance of (φ^*, t^*), every party reports truthfully and obeys its recommendation on the equilibrium path.*

The induced on-path joint distribution over ownership, states, and actions is

$$p(\theta)\, \mathbf{1}\{g = g^*\} \sigma^*(a \mid \theta),$$

and total surplus equals $\max_{g\in\mathcal{N}} TS(g)$.

Proof. For each (φ, t) among feasible direct mediated contracts, the continuation (communication) game induced by (φ, t) is a finite extensive-form game with perfect recall. Hence it has a sequential equilibrium (Kreps and Wilson, 1982), and therefore at least one PBE. For each feasible direct contract (φ, t), fix one continuation PBE of the induced communication game, and let $W_i(\varphi, t)$ denote party i's ex-ante payoff in that continuation equilibrium. For the optimal contract (φ^*, t^*), choose the truthful-reporting and obedient continuation Bayesian Nash equilibrium from Lemma 4, and we complete it to a continuation PBE using the off-path beliefs specified below.

Define the stage-1 offer strategy of party 1 by offering (φ^*, t^*). For each $i \neq 1$, define the stage-2 acceptance rule by

$$\chi_i(\varphi, t) = \begin{cases} \text{accept}, & \text{if } W_i(\varphi, t) \geqslant \underline{U}_i, \\ \text{reject}, & \text{if } W_i(\varphi, t) < \underline{U}_i. \end{cases}$$

Because t^* was chosen so that

$$W_i(\varphi^*, t^*) = \underline{U}_i \qquad \text{for every } i \neq 1,$$

every party $i \neq 1$ accepts (φ^*, t^*).

Conditional on acceptance of (φ^*, t^*), let every party report truthfully and obey its recommendation on path. By Lemma 4, this continuation behavior is a Bayesian Nash equilibrium of the communication game induced by (φ^*, t^*). Since the prior has full support, every report profile to the mediator occurs with positive probability. Thus, the only zero-probability information sets are histories $d_i = (g, t, a_i, \theta_i) \in D_i$ that are not induced on path by (φ^*, t^*).

Fix such an information set d_i. For each $j \neq i$, define a productive sign $\tilde{s}_{ij}(d_i) \in \{-1, 1\}$ by

$$\tilde{s}_{ij}(d_i) := \begin{cases} x_{ij}, & \text{if } x_{ij} = \bar{x}_{ij} \in \{-1, 1\}, \\ 1, & \text{otherwise.} \end{cases}$$

At d_i, prescribe the continuation actions $z_{ij} = \bar{z}_{ij} = 0$ for all $j \neq i$, $x_{ii} = \bar{x}_{ii} = \theta_i$, and $x_{ij} = \bar{x}_{ij} = \tilde{s}_{ij}(d_i)$ for $j \neq i$. Party i's off-path belief $\mu_i(\cdot \mid d_i)$ over θ_{-i} is generated by a latent common random variable $\xi \sim \text{Unif}[0, 1]$ according to

$$\theta_j = \begin{cases} \tilde{s}_{ij}(d_i), & \text{if } \xi \leqslant \tau_{ij}^{\text{safe}}(g), \\ -\tilde{s}_{ij}(d_i), & \text{if } \xi > \tau_{ij}^{\text{safe}}(g), \end{cases} \qquad \text{for every } j \neq i.$$

Moreover, party i believes that the same latent ξ governs all other parties' productive recommendations under owner g, and that all other parties follow their continuation strategies. Under these beliefs, each sign $\tilde{s}_{ij}(d_i)$ is a posterior mode for θ_j, and the posterior precision is $\tau_{ij}^{\text{safe}}(g)$. Hence no expropriation is weakly optimal, and the prescribed productive action plan is sequentially rational.

Finally, by Proposition 2, no direct contract can induce a continuation surplus strictly above $\max_{g \in \mathcal{N}} TS(g)$, whereas (φ^*, t^*) attains that value. If all parties accept a contract (φ, t), then

$$W_1(\varphi, t) = TS(\varphi, t) - \sum_{i \neq 1} W_i(\varphi, t) \leqslant \max_{g \in \mathcal{N}} TS(g) - \sum_{i \neq 1} \underline{U}_i = W_1(\varphi^*, t^*),$$

where the inequality uses Proposition 2 and the fact that acceptance implies $W_i(\varphi, t) \geqslant \underline{U}_i$ for all $i \neq 1$. If some party rejects (φ, t), party 1 receives the disagreement payoff, which is weakly below $W_1(\varphi^*, t^*)$. Hence party 1 weakly prefers offering (φ^*, t^*). Therefore the offer strategy, acceptance rules, continuation strategies, and belief system constructed above define a PBE of the full bargaining game. □

A.3 Contractible Actions and Cooperative Payoffs

In the baseline model, all components of the action plan $a_i = (x_i, \bar{x}_i, z_i, \bar{z}_i) \in \mathcal{A}_i$ are *non-contractible*: the mediator can only send private recommendations, and obedience must be verified against deviations in *all* components. In addition, project shares $\{\alpha_{ik}\}_{i,k}$ are fixed and exogenously given.

This section studies two counterfactual environments. First, we allow expropriation actions to be contractible; this makes it possible to implement the first best. Second, we allow cooperative actions and project shares to be contractible. We show that making cooperative actions and shares contractible does not increase the maximum attainable surplus relative to the baseline.

For notational convenience, decompose each party's action plan into a *cooperation block* and an *expropriation block*:

$$a_i^c := (x_i, \bar{x}_i) \in \mathcal{A}_i^c = \{-1, 1\}^{2N} \qquad a_i^e := (z_i, \bar{z}_i) \in \mathcal{A}_i^e = \{-1, 0, 1\}^{2(N-1)},$$

so that $\mathcal{A}_i = \mathcal{A}_i^c \times \mathcal{A}_i^e$ and $a_i = (a_i^c, a_i^e)$. The mediator observes θ and draws an action profile $a = (a_i)_{i \in \mathcal{N}} \in \mathcal{A}$ from a direct decision rule $\sigma(\cdot \mid \theta) \in \Delta(\mathcal{A})$.

(i) Contractible expropriation actions. Suppose the mediator can *enforce* expropriation actions $(z, \bar{z})$, while cooperative actions $(x, \bar{x})$ remain non-contractible. Then, after receiving a recommendation $a_i = (a_i^c, a_i^e)$, party i can deviate only in the cooperation block a_i^c; the expropriation block a_i^e is fixed.

Accordingly, σ is *obedient under g with contractible expropriation* if, for every $i \in \mathcal{N}$, every $\theta_i \in \Theta_i$, every recommended $a_i = (a_i^c, a_i^e) \in \mathcal{A}_i$, and every alternative $\tilde{a}_i^c \in \mathcal{A}_i^c$,

$$\sum_{\theta_{-i} \in \Theta_{-i}} p(\theta_{-i} \mid \theta_i) \sum_{a_{-i} \in \mathcal{A}_{-i}} \sigma(a_i, a_{-i} \mid \theta_i, \theta_{-i}) \Big[u_i\big(((a_i^c, a_i^e), a_{-i}), g; \theta\big) - u_i\big(((\tilde{a}_i^c, a_i^e), a_{-i}), g; \theta\big) \Big] \geqslant 0. \tag{A.17}$$

Define the corresponding omniscient benchmark value:

$$TS^e(g) := \max_{\sigma} \sum_{\theta} p(\theta) \sum_{a} \sigma(a \mid \theta) \cdot TS(a, g; \theta) \tag{A.18}$$
$$\text{subject to (A.17)}, \quad \sigma(a \mid \theta) \geqslant 0, \quad \sum_{a} \sigma(a \mid \theta) = 1 \text{ for all } \theta \in \Theta$$

(ii) Contractible actions and cooperative payoffs. Suppose instead that the mediator can *enforce* cooperative actions $(x, \bar{x})$, while expropriation actions $(z, \bar{z})$ remain non-contractible. Then party i can deviate only in the expropriation block a_i^e; the cooperation block a_i^c is fixed.

Accordingly, σ is *obedient under g with contractible cooperation* if, for every $i \in \mathcal{N}$, every $\theta_i \in \Theta_i$, every recommended $a_i = (a_i^c, a_i^e) \in \mathcal{A}_i$, and every alternative $\tilde{a}_i^e \in \mathcal{A}_i^e$,

$$\sum_{\theta_{-i} \in \Theta_{-i}} p(\theta_{-i} \mid \theta_i) \sum_{a_{-i} \in \mathcal{A}_{-i}} \sigma(a_i, a_{-i} \mid \theta_i, \theta_{-i}) \Big[u_i\big(((a_i^c, a_i^e), a_{-i}), g; \theta\big) - u_i\big(((a_i^c, \tilde{a}_i^e), a_{-i}), g; \theta\big) \Big] \geqslant 0. \tag{A.19}$$

Suppose the mediator can also choose and enforce the division of cooperative payoffs $\{\alpha_{ik}\}_{i,k}$. Thus, the corresponding omniscient mediator's benchmark value is

$$TS^c(g) := \max_{\sigma} \sum_{\theta} p(\theta) \sum_{a} \sigma(a \mid \theta) \cdot TS(a, g; \theta) \tag{A.20}$$
$$\text{subject to (A.19)}, \quad \alpha_{ik} \geqslant 0, \quad \sum_{i \in \mathcal{N}} \alpha_{ik} = 1 \text{ for all } k \in \mathcal{K},$$
$$\sigma(a \mid \theta) \geqslant 0, \quad \sum_{a} \sigma(a \mid \theta) = 1 \text{ for all } \theta \in \Theta.$$

Proposition 6 (Contractibility and the attainable surplus)**.** *Fix any ownership assignment $g \in \mathcal{N}$. If expropriation actions $(z, \bar{z})$ are contractible, then*

$$TS^e(g) = V_{\text{total}} \qquad \textit{for every } g \in \mathcal{N}.$$

If cooperative actions $(x, \bar{x})$ and project shares $\{\alpha_{ik}\}_{i,k}$ are contractible, then

$$TS^c(g) = TS(g) \qquad \textit{for every } g \in \mathcal{N},$$

where $TS(g)$ is the value characterized in Proposition 1.

Proof. **(i) Contractible expropriation actions:** $TS^e(g) = V_{\text{total}}$. Define a (deterministic) decision rule σ^{FB} as follows. For every $\theta \in \Theta$:

1. Set *all* expropriation actions to zero: $z_{ij} = 0$ and $\bar{z}_{ij} = 0$ for all i and all $j \neq i$;

2. Recommend $x_{ij} = \bar{x}_{ij} = \theta_j$ for all $i \in \mathcal{N}$ and $j \in \mathcal{N}$.

Under σ^{FB}, there is no expropriation, and every project succeeds surely. Thus the total surplus is V_{total}. Since $(z, \bar{z})$ are enforced, party i can deviate only in $a_i^c = (x_i, \bar{x}_i)$. If i deviates to some $\tilde{a}_i^c \neq a_i^c$, then at least one matching requirement that uses i's action is violated, which can only weakly decrease i's cooperative payoff (since $\alpha_{ik} \geqslant 0$ for all k). Thus, σ^{FB} is obedient.

(ii) Contractible cooperation and shares: $TS^c(g) = TS(g)$. First, we show that $TS^c(g) \geqslant TS(g)$. Let σ^g be the nested no-expropriation decision rule constructed in the proof of Proposition 1 (see (A.2)–(A.3)). Under σ^g we have $\zeta_{ij} = 0$ for all $i \neq j$ almost surely, and each project k succeeds with probability $\min_{(i,j)\in\mathcal{Q}_k(g)} \tau_{ij}^{\text{safe}}(g)$, so it achieves $TS(g)$. This decision rule is obedient and satisfies (A.19).

Second, we show that $TS^c(g) \leqslant TS(g)$. This inequality follows from the same steps as in the proof of Lemma 2. The binding restriction comes from the expropriation obedience constraints. Since expropriation actions remain non-contractible, the same upper bound continues to apply even when cooperative actions are contractible.

Finally, allowing the mediator to choose the shares $\{\alpha_{ik}\}$ does not affect the expected total surplus, which depends only on total project values. It also does not affect the obedience constraints (A.19): because cooperative actions are fixed in both terms, cooperative payoffs cancel when comparing deviations in expropriation. Hence, making shares contractible is immaterial for the welfare bound. □

A.4 Non-Contractible Communication: Cheap Talk

This appendix proves the results stated in Section 5.3. We first show how to adapt the reasoning from the proof of Proposition 1 for the binary-state case to the continuum analogue. We then prove Proposition 3 regarding the cheap-talk game.

A.4.1 The Mediator's Benchmark With Continuous State and Action Space

The proof of Proposition 1 extends to the continuum analogue after replacing the binary posterior precision with the *local posterior precision*:

$$q_{ij}(I_i) := \sup_{y \in \mathbb{T}} \mathbb{P}_\psi\big(d(y, \theta_j) \leqslant \delta_j \mid I_i\big), \qquad I_i = (\theta_i, a_i).$$

For an armed party i, the obedience constraint for the no-expropriation action implies $q_{ij}(I_i) \leqslant \tau^\star_{ij}$ at every information set. Thus, any productive task (i, j) can succeed with probability at most $\tau^{\text{safe}}_{ij}(g)$. Hence, for any obedient decision rule with no expropriation,

$$\mathbb{P}(E_k) \leqslant \min_{(i,j) \in \mathcal{Q}_k(g)} \tau^{\text{safe}}_{ij}(g),$$

and therefore,

$$\mathbb{E}_\psi[TS(a, g; \theta)] \leqslant \sum_{k \in \mathcal{K}} V_k \min_{(i,j) \in \mathcal{Q}_k(g)} \tau^{\text{safe}}_{ij}(g).$$

The decision rule provided in Section 5.3 attains this upper bound. The obedience and optimality of refraining from expropriation follow the same logic as in Appendix A.1.

A.4.2 Proof of Proposition 3

Proof. Fix an owner $g \in \mathcal{N}$. Because each target state is communicated to at most one armed party, every *risky* communication problem can be treated separately.

Consider first any task (i, j) with $i \notin \operatorname{Arm}(g)$. Since party i is disarmed, sender j can truthfully reveal θ_j to i and task (i, j) succeeds with probability $1 = \tau^{\text{safe}}_{ij}(g)$.

Success and Failure Zones. Now consider any task (i, j) with $i \in \operatorname{Arm}(g)$. To simplify notation, let $\tau^\star := \tau^\star_{ij}$ and $\delta := \delta_j$. Define the success zone and failure zone in $\mathbb{T} = [0, 1)$ by $\mathbb{T}^{\text{success}} := [0, \tau^\star]$ and $\mathbb{T}^{\text{fail}} := (\tau^\star, 1)$.

Partition of the Success Zone. Let $s^\circ := \min\{2\delta, \frac{\tau^\star}{2}\}$. We partition $\mathbb{T}^{\text{success}}$ into $M := \lceil \tau^\star/(2\delta) \rceil$ contiguous intervals $S^{(1)}, \ldots, S^{(M)}$ such that

$$S^{(1)} := [0, s^\circ), \qquad S^{(M)} := [\tau^\star - s^\circ, \tau^\star], \qquad |S^{(m)}| \leqslant 2\delta \quad \text{for every } m,$$

and $\bigcup_{m=1}^{M} S^{(m)} = \mathbb{T}^{\text{success}}$. Write $S^{(m)} = [\nu^{(m-1)}, \nu^{(m)})$ with $\nu^{(0)} := 0$ for $m = 1, \ldots, M-1$ and $S^{(m)} = [\nu^{(m-1)}, \nu^{(m)}]$ with $\nu^{(M)} := \tau^\star$. Thus, $|S^{(m)}| = \nu^{(m)} - \nu^{(m-1)}$.

For each interval $S^{(m)}$, choose an intended receiver action $x^{(m)}$ such that

$$x^{(1)} = \delta, \qquad x^{(M)} = \tau^\star - \delta, \qquad x^{(m)} = \frac{\nu^{(m-1)} + \nu^{(m)}}{2} \quad \text{for } m = 2, \ldots, M-1$$

Partition of the Failure Zone. Define the linear mapping:

$$\phi(y) := \tau^\star + \omega \cdot y \quad \text{for } y \in \mathbb{T}^{\text{success}}, \quad \text{where } \omega := \frac{1-\tau^\star}{\tau^\star}.$$

Define failure intervals as follows:

$$F^{(1)} := (\tau^\star, \phi(\nu^{(1)})), \qquad F^{(M)} := [\phi(\nu^{(M-1)}), 1), \qquad F^{(m)} := \phi(S^{(m)}) \quad \text{for } m = 2, \ldots, M-1.$$

The sets $\{F^{(m)}\}_{m=1}^{M}$ are disjoint intervals that partition $\mathbb{T}^{\text{fail}}$, and

$$|F^{(m)}| = \omega |S^{(m)}| = \frac{1-\tau^\star}{\tau^\star} |S^{(m)}|.$$

Strategies. The sender's strategy is truthful with respect to this partition: send message m if and only if $\theta_j \in S^{(m)} \cup F^{(m)}$. Upon receiving message m, the receiver chooses the intended action $x_{ij} = \bar{x}_{ij} = x^{(m)}$ and no expropriation ($\zeta = \varnothing$).

Sender incentive compatibility. If $\theta_j \in S^{(m)}$, then by construction

$$\theta_j \in B(x^{(m)}, \delta) := \{\theta : d(x^{(m)}, \theta) \leqslant \delta\},$$

so the prescribed message yields task (i,j) success without expropriation.

If $\theta_j \in F^{(m)}$, then θ_j lies strictly outside every receiver success neighborhood. Under any message, the task fails:

$$B(x^{(m')}, \delta) \cap \left(\mathbb{T}^{\text{fail}}\right) = \varnothing \quad \text{for all } m'.$$

Thus, sender types in $F^{(m)}$ are indifferent, and the prescribed strategy is sequentially rational.

Receiver incentive compatibility. After receiving message m, the receiver's posterior is uniform on $S^{(m)} \cup F^{(m)}$. By construction of the failure zones, the relative measure of the success interval is

$$\frac{|S^{(m)}|}{|S^{(m)}| + |F^{(m)}|} = \frac{|S^{(m)}|}{|S^{(m)}| + \frac{1-\tau^\star}{\tau^\star}|S^{(m)}|} = \tau^\star.$$

Since the intended action $x^{(m)}$ covers all $S^{(m)}$ and none of $F^{(m)}$, the probability of success from choosing $x^{(m)}$ is $\tau^\star$.

We verify that no deviating action $y \in \mathbb{T}$ can yield a strictly larger probability of success.

The success mass captured by any action y is $|B(y,\delta) \cap (S^{(m)} \cup F^{(m)})|$. If $B(y,\delta)$ intersects at most one of the two intervals, the captured mass is bounded by

$$|B(y,\delta) \cap (S^{(m)} \cup F^{(m)})| \leqslant \max\{|S^{(m)}|, |F^{(m)}|\} = |S^{(m)}|,$$

where the equality holds because $\tau^\star > 1/2$ implies $|F^{(m)}| < |S^{(m)}|$. Such a deviation yields a success probability of at most $\tau^\star$, making it unprofitable.

Suppose $M \geqslant 3$. This implies $\tau^\star > 4\delta$ and therefore $|S^{(1)}| = |S^{(M)}| = 2\delta$. For any inner message $m \in \{2,\ldots,M-1\}$, the distance between $S^{(m)}$ and $F^{(m)}$ in either direction around the circle is bounded below by 2δ. Thus, $B(y,\delta)$ cannot simultaneously intersect both $S^{(m)}$ and $F^{(m)}$, and for any inner cell $m \in \{2,\ldots,M-1\}$, any action captures at most mass $|S^{(m)}|$, meaning there is no profitable deviation from $x^{(m)}$.

Now consider $m \in \{1, M\}$ and a deviation $y \in \mathbb{T}$ such that $B(y,\delta)$ simultaneously intersects both $S^{(1)}$ and $F^{(1)}$. By symmetry, it suffices to analyze $m = 1$.

Intervals $S^{(1)}$ and $F^{(1)}$ are separated by two gaps on the circle $\mathbb{T}$. First, there is the clockwise gap spanning the remainder of the success zone, with length $D_{cw} = \tau^\star - |S^{(1)}|$. An interval of length 2δ covering this gap captures at most $2\delta - D_{cw} = |S^{(1)}| - (\tau^\star - 2\delta)$. Because $2\delta < \tau^\star$, this captured mass is strictly less than $|S^{(1)}|$.

Second, there is the counter-clockwise gap spanning the remainder of the failure zone, with length $D_{ccw} = 1 - \tau^\star - |F^{(1)}|$. An interval of length 2δ covering this gap captures at most $2\delta - D_{ccw} = 2\delta - 1 + \tau^\star + |F^{(1)}|$. For the deviation to be unprofitable, the captured mass must be bounded by the intended mass:

$$2\delta - 1 + \tau^\star + |F^{(1)}| \leqslant |S^{(1)}| \implies |S^{(1)}| - |F^{(1)}| \geqslant \tau^\star + 2\delta - 1.$$

Because $|F^{(1)}| = \omega|S^{(1)}|$, the following inequality must be satisfied:

$$|S^{(1)}| \geqslant \tau^\star \frac{\tau^\star + 2\delta - 1}{2\tau^\star - 1}. \tag{A.21}$$

By construction, $|S^{(1)}| = \min\{2\delta, \tau^\star/2\}$. If $|S^{(1)}| = \tau^\star/2$, then

$$\frac{\tau^\star}{2} \geqslant \tau^\star \frac{\tau^\star + 2\delta - 1}{2\tau^\star - 1} \iff \frac{1}{2} \geqslant \delta,$$

which is satisfied by Assumption 2* ($\delta \leqslant 1/4$). If $|S^{(1)}| = 2\delta$, then

$$2\delta \geqslant \tau^\star \frac{\tau^\star + 2\delta - 1}{2\tau^\star - 1} \iff (1 - \tau^\star)(\tau^\star - 2\delta) \geqslant 0,$$

which is satisfied by Assumption $2^\star$ ($\tau^\star > 1/2$ and $\delta \leqslant 1/4$).

Thus, the inequality (A.21) is satisfied, and no deviation can yield a higher expected success probability than $x^{(1)}$. By symmetric geometric construction, the same holds for $m = M$, concluding the proof of receiver incentive compatibility.

Probability of Success. In the constructed PBE, task (i, j) with $i \in \mathrm{Arm}(g)$ succeeds with probability $\tau_{ij}^\star = \tau_{ij}^{\mathrm{safe}}(g)$. □

A.5 Optimal Decision Rule Under Arbitrary Expropriation Costs

We describe the results as follows. First, for any obedient decision rule we derive an upper bound on the total surplus. Second, we construct an obedient decision rule that attains this bound. Finally, we formulate the mediator's optimization problem.

A.5.1 The Upper Bound on Total Surplus

Information Accuracy. Fix an owner $g \in \mathcal{N}$. Consider an obedient decision rule $\sigma \in \Sigma$ with the induced joint distribution $\psi(a, \theta) = p(\theta)\sigma(a \mid \theta)$ over states and actions. Let $I_i = (\theta_i, a_i)$ be party i's private information at the action stage, and let

$$q_{ij}(I_i) := \max_{\omega \in \{-1,1\}} \mathbb{P}_\psi(\theta_j = \omega \mid I_i)$$

denote the maximum conditional accuracy of matching state θ_j.

The conditional accuracy $q_{ij}(I_i)$ provides an upper bound on the probability that party i successfully matches task j. Let $\mathcal{Q}_{k,i}(g) := \{j : (i, j) \in \mathcal{Q}_k(g)\}$ denote the set of targets party i must match for project k to succeed. Let $E_{k,i}$ denote the event that party i succeeds in all its tasks in project k. Since this event requires *all* tasks to succeed, we have

$$\mathbb{P}_\psi\left(E_{k,i}\right) = \mathbb{E}_\psi\left[\mathbb{P}_\psi(E_{k,i} \mid I_i)\right] \leqslant \mathbb{E}_\psi\left[\min_{j \in \mathcal{Q}_{k,i}(g)} \{q_{ij}(I_i)\}\right].$$

Expropriation Attempt Patterns. For each armed party $i \in \mathrm{Arm}(g)$, we define the random expropriation attempt pattern S_i and its probability π_{iS} as

$$S_i := \{j \in \mathcal{N} \backslash \{i\} : \zeta_{ij} \neq 0\}, \quad \pi_{iS} := \mathbb{P}_\psi\left(S_i = S\right)$$

That is, S_i is the set of targets j for which i is recommended to attempt expropriation. Let $J_i(g) := \bigcup_k \mathcal{Q}_{k,i}(g)$ denote the set of all production targets for party i. Without loss of optimality, we can focus on decision rules in which i receives no expropriation recommendations for irrelevant targets: $\zeta_{ij} = 0$ for all $j \notin J_i(g)$, which implies $S_i \subseteq J_i(g)$.

Upper Bound on a Party's Success. We derive an upper bound on party i's success probability by conditioning on the attempt pattern S_i. By the Law of Iterated Expectations and Jensen's inequality (as the minimum function is concave), we have:

$$\mathbb{P}_\psi\left(E_{k,i}\right) \leqslant \mathbb{E}_S\left[\mathbb{E}_\psi\left[\min_{j\in\mathcal{Q}_{k,i}(g)} q_{ij}(I_i) \,\middle|\, S_i\right]\right] \leqslant \mathbb{E}_S\left[\min_{j\in\mathcal{Q}_{k,i}(g)} \mathbb{E}_\psi\left[q_{ij}(I_i) \mid S_i\right]\right]. \tag{A.22}$$

The obedience constraints impose the following bounds on the accuracy $q_{ij}(I_i)$. If party i's recommendation is safe action $\zeta_{ij} = 0$ (i.e., $j \notin S_i$), obedience requires $q_{ij}(I_i) \leqslant \tau_{ij}^{\text{safe}}(g)$. Conversely, if i is recommended expropriation $\zeta_{ij} \neq 0$ (i.e., $j \in S_i$), it must be sufficiently certain about the target state: $q_{ij}(I_i) \geqslant \tau_{ij}^{\text{safe}}(g)$. Thus, $\mathbb{E}_\psi[q_{ij}(I_i) \mid S_i] \leqslant \tau_{ij}^{\text{safe}}(g)$ for $j \notin S_i$ and $\mathbb{E}_\psi[q_{ij}(I_i) \mid S_i] \geqslant \tau_{ij}^{\text{safe}}(g)$ for $j \in S_i$.

To derive the upper bound, we set the accuracy of safe actions to their maximum permissible value. We define the conditional accuracy variable m_{ij}^S as follows:

$$m_{ij}^S := \begin{cases} \tau_{ij}^{\text{safe}}(g) & \text{if } j \notin S \quad \text{(Safe Action)}, \\ \mathbb{E}_\psi[q_{ij}(I_i) \mid S_i] \in [\tau_{ij}^{\text{safe}}(g), 1] & \text{if } j \in S \quad \text{(Expropriation Attempt)}. \end{cases}$$

Combining the derived inequalities with these definitions, we obtain the following upper bound on the probability that party i succeeds in project k:

$$\mathbb{P}_\psi\left(E_{k,i}\right) \leqslant \sum_{S\subseteq J_i(g)} \pi_{iS}\, \rho_{k,i}^S := P_{k,i}(\pi, m; g), \quad \rho_{k,i}^S := \min_{j\in\mathcal{Q}_{k,i}(g)} m_{ij}^S,$$

where we adopt the convention $\min \varnothing := 1$.

The Party's Success Across Projects. If party i participates in multiple projects, the success events for different projects can be correlated. For any nonempty subset $L \subseteq \mathcal{K}$, define the event that party i succeeds in *all* projects in L: $E_{L,i} := \bigcap_{k\in L} E_{k,i}$. If i succeeds in all projects in L, it must match *every* target in the union $\bigcup_{k\in L} \mathcal{Q}_{k,i}(g)$. Hence, repeating the previous argument in (A.22) with the target set $\bigcup_{k\in L} \mathcal{Q}_{k,i}(g)$ gives

$$\mathbb{P}_\psi(E_{L,i}) \leqslant \sum_{S\subseteq J_i(g)} \pi_{iS} \min_{j\in\cup_{k\in L}\mathcal{Q}_{k,i}(g)} m_{ij}^S = \sum_{S\subseteq J_i(g)} \pi_{iS} \min_{k\in L} \rho_{k,i}^S. \tag{A.23}$$

Up to this point, we have shown that any obedient decision rule σ (with induced distribution ψ) is associated with two statistics: expropriation attempt probabilities π and conditional expected accuracies m. These statistics pin down the upper bound on the probability of i's success in various projects. Specifically, equation (A.23) provides a restriction on the joint distribution of the random vector $(\mathbf{1}\{E_{k,i}\})_{k\in\mathcal{K}}$ under ψ.

To upper bound cooperative surplus, we replace the unknown conditional joint law of $(E_{k,i})_k$ by an *auxiliary* joint law that attains all bounds in (A.23) simultaneously. Let $\xi_i \sim \mathrm{Unif}[0,1]$ be an auxiliary independent random variable. Conditional on $S_i = S$, define auxiliary success events $\widetilde{E}^S_{k,i} := \{\xi_i \leqslant \rho^S_{k,i}\}$ for $k \in \mathcal{K}$. Then, for every nonempty $L \subseteq \mathcal{K}$,

$$\mathbb{P}\Big(\bigcap_{k\in L} \widetilde{E}^S_{k,i}\Big) = \mathbb{P}\Big(\xi_i \leqslant \min_{k\in L} \rho^S_{k,i}\Big) = \min_{k\in L} \rho^S_{k,i},$$

so the auxiliary construction saturates the right-hand side of (A.23) for all L at once.

Now mix over attempt patterns using the *actual* weights $\pi_{iS} := \mathbb{P}_\psi(S_i = S)$: first draw S_i according to π_i, then draw $\xi_i \sim \mathrm{Unif}[0,1]$, and set $\widetilde{E}_{k,i} := \widetilde{E}^{S_i}_{k,i}$. By the law of total probability, for every $L \subseteq \mathcal{K}$, we have

$$\mathbb{P}\Big(\bigcap_{k\in L} \widetilde{E}_{k,i}\Big) = \sum_{S\subseteq J_i(g)} \pi_{iS}\, \mathbb{P}\Big(\bigcap_{k\in L} \widetilde{E}^S_{k,i}\Big) = \sum_{S\subseteq J_i(g)} \pi_{iS} \min_{k\in L} \rho^S_{k,i}.$$

Since cooperative surplus is increasing in each party's project-success indicators, replacing the unknown joint law of $(E_{k,i})_k$ by the "maximal" law $(\widetilde{E}_{k,i})_k$ (which saturates all upper bounds) can only increase the best cooperative surplus consistent with (A.23). Thus, evaluating the cooperative part under $(\widetilde{E}_{k,i})$ yields an upper bound. Later we show that this upper bound is attainable.

Define the auxiliary (maximal) success type of party i as the random set

$$T_i := \{k \in \mathcal{K} :\ \widetilde{E}_{k,i} \text{ occurs}\} \subseteq \mathcal{K}.$$

Let $\beta_{iT}(\pi, m; g)$ denote the marginal distribution of T_i:

$$\beta_{iT}(\pi, m; g) := \mathbb{P}(T_i = T), \qquad T \subseteq \mathcal{K}.$$

Conditional on $S_i = S$, the event $\{T_i = T\}$ is equivalent to

$$\{\xi_i \leqslant \rho^S_{k,i}\ \forall k \in T\} \cap \{\xi_i > \rho^S_{k,i}\ \forall k \notin T\} \quad \Longleftrightarrow \quad \xi_i \in \Big(\max_{k\notin T} \rho^S_{k,i},\ \min_{k\in T} \rho^S_{k,i}\Big],$$

with $\max\varnothing := 0$ and $\min\varnothing := 1$. Since $\xi_i \sim \mathrm{Unif}[0,1]$,

$$\mathbb{P}(T_i = T \mid S_i = S) = \Big[\min_{k\in T} \rho^S_{k,i} - \max_{k\notin T} \rho^S_{k,i}\Big]_+,$$

and therefore

$$\beta_{iT}(\pi, m; g) = \sum_{S \subseteq J_i(g)} \pi_{iS} \left[\min_{k \in T} \rho^S_{k,i} - \max_{k \notin T} \rho^S_{k,i} \right]_+ . \tag{A.24}$$

Fix (π, m) and hence the induced marginals $\big(\beta_{iT}(\pi, m; g)\big)_{T \subseteq \mathcal{K}}$ for each active party i. At this stage, the mediator still has freedom to choose how to *correlate* the success types $(T_i)_i$ across parties. Let $\widetilde{\mathcal{P}}_k(g)$ denote the set of participants in project k (those whose matching tasks appear in $\mathcal{Q}_k(g)$).[4] Project k succeeds if and only if every $i \in \widetilde{\mathcal{P}}_k(g)$ succeeds in k, which in type language is

$$\text{Project } k \text{ succeeds} \quad \Longleftrightarrow \quad k \in T_i \ \text{ for all } i \in \widetilde{\mathcal{P}}_k(g).$$

Let $\mathcal{I}(g) := \bigcup_{k \in \mathcal{K}} \widetilde{\mathcal{P}}_k(g)$ be the set of parties whose actions matter for at least one project. A *coupling* is a joint distribution

$$\gamma \in \Delta\Big((2^{\mathcal{K}})^{\mathcal{I}(g)}\Big) \quad \text{over type profiles} \quad \mathbf{T} = (T_i)_{i \in \mathcal{I}(g)}.$$

The only feasibility restrictions on γ are the marginal constraints

$$\sum_{\mathbf{T}_{-i}} \gamma(\mathbf{T}) = \beta_{iT_i}(\pi, m; g) \qquad \forall i \in \mathcal{I}(g), \ \forall T_i \subseteq \mathcal{K},$$

because (i) each party observes only its own recommendation, so obedience is pinned down by its own marginal signal structure (already encoded in (π, m)), and (ii) the mediator can implement any cross-party correlation by correlating the latent randomization used to generate success events.

Under a coupling γ, cooperative surplus equals

$$\sum_{k \in \mathcal{K}} V_k \, \mathbb{P}_\gamma\big(k \in T_i \ \forall i \in \widetilde{\mathcal{P}}_k(g)\big) = \sum_{k \in \mathcal{K}} V_k \sum_{\mathbf{T}} \gamma(\mathbf{T}) \, \mathbf{1}\{k \in T_i \ \forall i \in \widetilde{\mathcal{P}}_k(g)\}.$$

Therefore, for fixed (π, m), the mediator's *best* cooperative surplus is the value of the following *multi-marginal optimal transport (MMOT)* linear program:

$$\begin{aligned} \mathrm{W}\big(\pi, m; g\big) := \max_{\gamma \geqslant 0} \ & \sum_{k \in \mathcal{K}} V_k \sum_{\mathbf{T}} \gamma(\mathbf{T}) \, \mathbf{1}\{k \in T_i \ \forall i \in \widetilde{\mathcal{P}}_k(g)\} \\ \text{s.t.} \quad & \sum_{\mathbf{T}_{-i}} \gamma(\mathbf{T}) = \beta_{iT_i}(\pi, m; g) \ \ \forall i, T_i. \end{aligned} \tag{A.25}$$

[4] The set $\widetilde{\mathcal{P}}_k(g)$ differs from the set of *active* participants $\mathcal{P}_k(g)$ defined in (3.2) in the main text by including parties with trivial tasks: $i \in \widetilde{\mathcal{P}}_k(g)$ for $(i, i) \in \mathcal{Q}_k(g)$.

The same parameters (π, m) also pin down probabilities of successful and unsuccessful expropriation attempts for $i \in \text{Arm}(g)$:

$$\mathbb{P}_\psi(\zeta_{ij} = \theta_j) = \sum_{S:j\in S} \pi_{iS}\ m_{ij}^S, \qquad \mathbb{P}_\psi(\zeta_{ij} = -\theta_j) = \sum_{S:j\in S} \pi_{iS}\ (1 - m_{ij}^S)$$

Hence, combining the cooperative bound (A.25) with the surplus contribution of expropriation yields the following *general* upper bound on expected total surplus under owner g:

$$\begin{aligned}\mathbb{E}_\psi[TS(a; g, \theta)] \ &\leqslant \ \mathrm{W}\big(\pi, m; g\big) - \\ &- \sum_{i\in\text{Arm}(g)} \sum_{j\in J_i(g)\setminus\{i\}} \Big[(\lambda_{ij} - b_{ij}) \sum_{S:j\in S} \pi_{iS}\ m_{ij}^S + c_{ij} \sum_{S:j\in S} \pi_{iS}\ (1 - m_{ij}^S)\Big].\end{aligned} \tag{A.26}$$

A.5.2 Attainability

Fix any feasible (π, m) and an optimal coupling γ that solves (A.25). We construct an obedient decision rule that achieves the upper bound in (A.26).

Step 1: Project-Success Regions. For each $i \in \text{Arm}(g)$, choose a measurable partition of $[0,1]$ into intervals $(I_{iS})_{S\subseteq J_i(g)}$ with $|I_{iS}| = \pi_{iS}$. For each project $k \in K$, define the region

$$\mathcal{S}_{k,i} := \bigcup_{S\subseteq J_i(g)} \Big(I_{iS} \times [0, \rho_{k,i}^S]\Big) \subseteq [0,1]^2.$$

For each subset $T \subseteq K$ define the corresponding cell in the induced partition of $[0,1]^2$:

$$\Delta_{i,T} := \left(\bigcap_{k\in T} \mathcal{S}_{k,i}\right) \cap \left(\bigcap_{k\notin T} \mathcal{S}_{k,i}^c\right).$$

Then $\{\Delta_{i,T}\}_{T\subseteq K}$ is a measurable partition of $[0,1]^2$, and the Lebesgue measure of each region is $\text{Leb}(\Delta_{i,T}) = \beta_{iT}(\pi, m; g)$ by the definition of β_{iT}.

Step 2: Implement the chosen coupling γ. Draw $(T_i)_{i\in\mathcal{I}(g)}$ according to γ. Conditional on $(T_i)_i$, draw (η_i, ξ_i) independently across i with

$$(\eta_i, \xi_i) \ \sim \ \text{Unif}\big(\Delta_{i,T_i}\big).$$

Because each T_i is selected with probability equal to the area of Δ_{i,T_i}, this implies:

1. *Uniform marginals:* unconditionally, (η_i, ξ_i) is uniform on $[0,1]^2$ for each i.

2. *Type realization:* for each $k \in K$, we have $k \in T_i \ \Leftrightarrow \ (\eta_i, \xi_i) \in \mathcal{S}_{k,i}$.

Step 3: Send Recommendations. Fix a realized state θ. For each disarmed party $i \notin \mathrm{Arm}(g)$, recommend no expropriation and fully revealing matching actions for all relevant targets (so such parties match surely and never expropriate). Now fix $i \in \mathrm{Arm}(g)$. Define the realized attempt pattern from η_i:

$$S_i = S \iff \eta_i \in I_{iS}.$$

For each $j \in J_i(g)$ define the recommended sign

$$s_{ij} := \begin{cases} \theta_j, & \xi_i \leqslant m_{ij}^{S_i}, \\ -\theta_j, & \xi_i > m_{ij}^{S_i}. \end{cases}$$

Recommend cooperative actions $x_{ij} = \bar{x}_{ij} = s_{ij}$ for all $j \in J_i(g)$, and recommend expropriation

$$\zeta_{ij} = \begin{cases} s_{ij}, & j \in S_i, \\ 0, & j \notin S_i. \end{cases}$$

For targets not in $J_i(g)$, recommend uninformative (constant) cooperative actions $x_{ij} = \bar{x}_{ij} = 1$ and $\zeta_{ij} = 0$.

Posteriors. The information available to party $i \in \mathrm{Arm}(g)$ about targets $j \neq i$ before taking actions is the attempted set S_i (encoded in $\zeta_{ij} \neq 0$) and the recommended signs $s_i := (s_{ij})_{j \in J_i(g)}$ (encoded in $(x_{ij}, \bar{x}_{ij})$). The attempted set S_i is independent of θ_i. Indeed, $(T_i)_i \sim \gamma$ is drawn independently of θ, and (η_i, ξ_i) are drawn from Δ_{i,T_i} independently of θ.

Fix a realized message (S_i, s_i) and $j \in J_i(g)$. Compute the following posterior:

$$\mathbb{P}\left(\theta_j = s_{ij} \mid S_i, s_i\right) = \mathbb{E}\left[\mathbb{P}\left(\theta_j = s_j \mid S_i, s_i, \xi_i\right) \mid S_i, s_i\right]$$

Conditional on (S_i, ξ_i), we have $\theta_j = s_{ij}$ if an only if $\xi_i \leqslant m_{ij}^{S_i}$, and the map $\theta_{J_i(g)} \mapsto s_i$ is a deterministic coordinate-wise sign flip bijection on $\{-1,1\}^{J_i(g)}$. Since $\theta_{J_i(g)}$ is uniform on $\{-1,1\}^{J_i(g)}$, it follows that

$$\mathbb{P}(s_i = \bar{s} \mid S_i, \xi_i) = 2^{-|J_i(g)|} \quad \text{for every } \bar{s} \in \{-1,1\}^{J_i(g)},$$

which does not depend on ξ_i. Hence, s_i is independent of ξ_i conditional on S_i. Using this independence, we get

$$\mathbb{P}(\theta_j = s_{ij} \mid S_i, s_i) = \mathbb{P}(\xi_i \leqslant m_{ij}^{S_i} \mid S_i, s_i) = \mathbb{P}(\xi_i \leqslant m_{ij}^{S} \mid S_i) = m_{ij}^{S_i},$$

where the last equality uses $\xi_i \mid S_i \sim \text{Unif}[0,1]$. Therefore, the posterior for target j is

$$\mathbb{P}(\theta_j = s_{ij} \mid \theta_i, S_i, s_i) = m_{ij}^{S_i}, \qquad \mathbb{P}(\theta_j = -s_{ij} \mid \theta_i, S_i, s_i) = 1 - m_{ij}^{S_i}. \tag{A.27}$$

Obedience. Fix $i \in \text{Arm}(g)$. If $j \in S_i$, the recommended expropriation is ζ_{ij}. Since $m_{ij}^{S_i} \geqslant \tau_{ij}^{\text{safe}} = \tau_{ij}^{\star}$, the posterior in (A.27) implies that it is optimal for party i to follow the recommendation. If $j \notin S_i$, then $m_{ij}^{S_i} = \tau_{ij}^{\text{safe}}$, and it is optimal for i to follow the recommendation $\zeta_{ij} = 0$.

The obedience in cooperative actions follows the same logic as in the proof of Lemma 1. Since $\xi_i \mid S_i \sim \text{Unif}[0,1]$ and $m_{ij}^{S_i} \geqslant \tau_{ij}^{\text{safe}}(g) > \frac{1}{2}$, following the cooperative recommendation ensures that a project succeeds with probability higher than $\frac{1}{2}$, while deviating decreases this probability to less than $\frac{1}{2}$ in each project where i's actions are required, and i deviates.

A.5.3 The Mediator's Problem

Up to this point, we have shown that each obedient decision rule has an upper bound on the total surplus in (A.26), and this bound is attainable. Now, we formulate the mediator's problem using this upper bound.

Let $\mathcal{PM}(g)$ denote the feasible set for the expropriation attempt probabilities π and corresponding accuracies m. That is, $\pi := (\pi_1, \dots, \pi_N)$, where $(\pi_{iS})_{S \subseteq \mathcal{N}}$, $\pi_{iS} \geqslant 0$, and $\sum_{S \subseteq \mathcal{N}} \pi_{iS} = 1$. For accuracies, $m := (m_1, \dots, m_N)$, where $m_i := (m_{ij}^S)_{j \in \mathcal{N}, S \subseteq \mathcal{N}}$, $m_{ij}^S = \tau_{ij}^{\text{safe}}(g)$ for $j \notin S$, and $m_{ij}^S \in [\tau_{ij}^{\text{safe}}(g), 1]$ for $j \in S$.

The derivations in this section prove the following proposition:

Proposition 7. *Suppose that successful expropriation has arbitrary costs $\lambda_{ij} - b_{ij} > 0$. The maximum attainable total surplus under owner g is*

$$TS(g) = \max_{(\pi,m) \in \mathcal{PM}(g)} \Big\{ \text{W}(\pi, m; g) - \sum_{i \in \text{Arm}(g)} \sum_{j \in J_i(g) \setminus \{i\}} \Big[(\lambda_{ij} - b_{ij}) \sum_{S: j \in S} \pi_{iS} \, m_{ij}^S + c_{ij} \sum_{S: j \in S} \pi_{iS} \, (1 - m_{ij}^S) \Big] \Big\}, \tag{A.28}$$

where $\text{W}(\pi, m; g)$ *is a solution to the multi-marginal optimal transport problem in* (A.25).

The problem in (A.28) can be simplified under the following assumption.

Assumption 4. *For each $g \in \mathcal{N}$, the family of project task sets $\{\mathcal{Q}_k(g)\}_{k \in \mathcal{K}}$ is totally ordered by inclusion. That is, the projects can be indexed $k = 1, \dots, |\mathcal{K}|$ such that:*

$$\mathcal{Q}_1(g) \supseteq \mathcal{Q}_2(g) \supseteq \cdots \supseteq \mathcal{Q}_{|\mathcal{K}|}(g).$$

This assumption implies that projects can be ranked by difficulty: Project 1 is the "hardest" (requires the most tasks), while Project $|\mathcal{K}|$ is the "easiest."

Corollary 4. *Suppose Assumption 4 holds. The maximum attainable total surplus under owner* $g \in \mathcal{N}$ *is*

$$TS(g) = \max_{(\pi,m)\in\mathcal{PM}(g)} \Big\{ \sum_{k\in\mathcal{K}} V_k \cdot \min_{i\in\widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g) - \sum_{i\in\mathrm{Arm}(g)} \sum_{j\in J_i(g)\setminus\{i\}} \Big[(\lambda_{ij} - b_{ij}) \sum_{S:j\in S} \pi_{iS}\, m_{ij}^S + c_{ij} \sum_{S:j\in S} \pi_{iS}\, (1 - m_{ij}^S)\Big]\Big\},$$

where

$$P_{k,i}(\pi, m; g) := \sum_{S\subseteq J_i(g)} \pi_{iS}\, \rho_{k,i}^S, \qquad \rho_{k,i}^S := \min_{j\in\mathcal{Q}_{k,i}(g)} m_{ij}^S \quad (\min\varnothing := 1).$$

Proof. Under Assumption 4, for each i we have $\mathcal{Q}_{1,i}(g) \supseteq \mathcal{Q}_{2,i}(g) \supseteq \cdots$. Therefore,

$$\rho_{1,i}^S = \min_{j\in\mathcal{Q}_{1,i}(g)} m_{ij}^S \leqslant \min_{j\in\mathcal{Q}_{2,i}(g)} m_{ij}^S = \rho_{2,i}^S \leqslant \cdots \leqslant \rho_{|\mathcal{K}|,i}^S.$$

Hence $P_{k,i}(\pi, m; g) = \sum_S \pi_{iS}\rho_{k,i}^S$ is non-decreasing in k.

Step 1: Nested Regions of Success. In the construction of an optimal decision rule, this monotonicity implies that the success regions are nested:

$$\mathcal{S}_{1,i} \subseteq \mathcal{S}_{2,i} \subseteq \cdots \subseteq \mathcal{S}_{|\mathcal{K}|,i}, \qquad \mathrm{Leb}(\mathcal{S}_{k,i}) = P_{k,i}(\pi, m; g).$$

As a result, the only nonempty cells $\Delta_{i,T}$ correspond to *upper sets* of the form

$$T^{(k)} := \{k, k+1, \ldots, |\mathcal{K}|\} \quad (k = 1, \ldots, |\mathcal{K}|), \qquad T^{(|\mathcal{K}|+1)} := \varnothing,$$

with

$$\Delta_{i,T^{(k)}} = \mathcal{S}_{k,i}\setminus\mathcal{S}_{k-1,i}, \qquad \Delta_{i,\varnothing} = \mathcal{S}_{|\mathcal{K}|,i}^c,$$

where $\mathcal{S}_{0,i} := \varnothing$. Therefore the type marginals are

$$\beta_{i,T^{(k)}}(\pi, m; g) = \mathrm{Leb}(\Delta_{i,T^{(k)}}) = P_{k,i} - P_{k-1,i}, \qquad \beta_{i,\varnothing}(\pi, m; g) = 1 - P_{|\mathcal{K}|,i},$$

with the convention $P_{0,i} := 0$.

Step 2: Upper Bound on MMOT. Let γ be any feasible coupling in (A.25), i.e. with marginals $\beta_{iT}(\pi, m; g)$. For any project k, the MMOT objective involves the probability that *all required parties* succeed in project k:

$$\mathbb{P}_\gamma\big(k \in T_i \ \forall i \in \widetilde{\mathcal{P}}_k(g)\big) \ \leqslant \ \min_{i \in \widetilde{\mathcal{P}}_k(g)} \mathbb{P}_\gamma(k \in T_i).$$

Under the fixed marginal of party i,

$$\mathbb{P}_\gamma(k \in T_i) = \sum_{T:\, k \in T} \beta_{iT}(\pi, m; g) = P_{k,i}(\pi, m; g),$$

because $k \in T_i$ is the event $(\eta_i, \xi_i) \in \mathcal{S}_{k,i}$, whose area is $P_{k,i}$. Therefore,

$$\mathbb{P}_\gamma\big(k \in T_i \ \forall i \in \widetilde{\mathcal{P}}_k(g)\big) \ \leqslant \ \min_{i \in \widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g).$$

Multiplying by V_k, summing over k, and taking the maximum over feasible γ yields

$$\mathrm{W}(\pi, m; g) \ \leqslant \ \sum_{k \in \mathcal{K}} V_k \min_{i \in \widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g).$$

Step 3: Attainability of the Upper Bound on MMOT. Let $V \sim \mathrm{Unif}[0,1]$ be a *common* random variable shared across all parties. For each party i, define its type using the threshold rule determined by the cutoffs $(P_{k,i})_k$:

$$T_i(V) := \{k \in \mathcal{K} : \ V \leqslant P_{k,i}(\pi, m; g)\}.$$

Because $P_{k,i}$ is non-decreasing in k, $T_i(V)$ is always an upper set. Moreover,

$$\mathbb{P}\big(T_i(V) = T^{(k)}\big) = \mathbb{P}\big(P_{k-1,i} < V \leqslant P_{k,i}\big) = P_{k,i} - P_{k-1,i} = \beta_{i,T^{(k)}}(\pi, m; g),$$

and $\mathbb{P}(T_i(V) = \varnothing) = 1 - P_{|\mathcal{K}|,i} = \beta_{i,\varnothing}(\pi, m; g)$. Hence the induced joint law of $(T_i(V))_{i \in \mathcal{I}(g)}$ is a feasible coupling for (A.25).

Under this coupling,

$$k \in T_i(V) \ \forall \ i \in \widetilde{\mathcal{P}}_k(g) \quad \Longleftrightarrow \quad V \leqslant \min_{i \in \widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g),$$

so

$$\mathbb{P}\big(k \in T_i(V) \ \forall i \in \widetilde{\mathcal{P}}_k(g)\big) = \min_{i \in \widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g).$$

Therefore, the MMOT objective value achieved by this coupling equals

$$\sum_{k\in\mathcal{K}} V_k \min_{i\in\widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g),$$

which matches the upper bound from Step 2. Hence

$$\mathrm{W}(\pi, m; g) = \sum_{k\in\mathcal{K}} V_k \min_{i\in\widetilde{\mathcal{P}}_k(g)} P_{k,i}(\pi, m; g).$$

□