# A Theory of the Firm Based on Haggling, Coordination, and Rent-Seeking*

Sandeep Baliga
Northwestern University

Tomas Sjöström
Rutgers University

October 19, 2018

**Abstract**

Two agents want to coordinate their decisions, but may also try to extract rents from each other. Contracts are incomplete in the sense that the agents have private information at the time when decisions are made. Since the agent who owns an asset has the right to make a unilateral decision regarding this asset, ownership determines disagreement payoffs, which in turn determine the attainable surplus. We find that nonintegration (where each agent owns one asset) sometimes dominates integration (where one agent owns both assets) and vice versa. In particular, nonintegration is optimal if the gains from cooperation are large compared to the possible rents that can be extracted. Thus, integration does not always provide the most efficient way to resolve disputes resulting from incomplete contracts.

## 1 Introduction

A complete contract specifies, *ex ante*, a decision for every possible state of the world. With complete contracts, decisions will be first-best (surplus-maximizing), regardless of ownership and organizational form. However,

*We thank Robert Gibbons, Oliver Hart, Massimo Morelli, Andrew Newman and Steve Tadelis for valuable comments. We are especially grateful to the Editor (David Myatt), two anonymous referees, and Niko Matouschek for many helpful comments. We also thank Philip Marx for research assistance.

if contracts are incomplete then disputes may occur about how to fill in the missing details, and haggling over how to settle these disputes may be costly. Williamson argued that such disputes can be settled at a lower cost within an integrated firm, since the boss can make decisions unilaterally (see Williamson [29], [30]). On the other hand, if the boss engages in inefficient rent-seeking at the employee's expense, then integration may carry its own "costs of fiat". In Williamson's [29] theory of the firm, the optimal organizational form trades off the costs of fiat under integration against the higher costs of haggling under nonintegration.

If the parties can negotiate efficiently, they will agree to eliminate inefficient rent-seeking whether they are integrated or not. However, if they have private information then the agreement must satisfy incentive-compatibility (IC) and interim participation constraints. This can make it impossible to reach the first-best.[1] The resulting loss of surplus is the cost of haggling. The participation constraints depend on asset ownership (control rights): in the integrated firm, the threat point is for the boss to make a unilateral decision, while in the nonintegrated case, each agent has the right to choose a decision relating to the asset he owns. Therefore, the cost of haggling will in general depend on asset ownership. In our theory, the optimal organizational form minimizes this cost. We find that either organizational form may be optimal, depending on the parameters. This complements the analysis of Grossman and Hart [9] and Hart and Moore [12] (henceforth GHM), where the optimal organizational form provides the best incentives for ex ante investments. There are no ex ante investments in our model.

In our model, reallocating control rights can raise total surplus by allowing better decisions to be made. The logic is reminiscent of Williamson's [32] argument that "hostages" can incentivize bargaining and reduce the cost of haggling. The control rights are the hostages. Consider the relationship between Nokia and Microsoft. Even before Microsoft acquired Nokia's Devices and Services unit, Nokia produced only Windows-based products, and most Windows Phone devices sold were made by Nokia (O'Brien [19]). The fact that the nonintegrated companies each controlled assets such as patents gave them good outside options. The logic of GHM suggests that outside options give a company bargaining power, which in turn gives it an incentive to invest in the relationship. In contrast, the logic of our model suggests that the exis-

[1] In the terminology of the mechanism design literature, the equilibrium agreement is "incentive efficient" but not necessarily first-best.

tence of good outside options can make it difficult to coordinate production decisions. In fact, Warren [28] reports on such coordination failures:

> "There are real-world examples of situations where Nokia was building a phone and keeping information about it secret from us," said Joe Belfiore, a corporate vice president at Microsoft who's in charge of the company's Windows Phone project. "We would make changes in the software, or prioritize things in the software, unaware of the work that they're doing. And then late in the cycle we'd find out and say, 'If we had known that we would have done this other thing differently and it would have turned out better!' "(Warren [28]).

In our model, integration is favored if the "hostages" are very valuable, so that the potential gain from holding up the other agent – rent seeking – is large. Eventually, integration consolidated patents in the hands of Microsoft which removed Nokia's outside option (Wingfield [35]).

The situation we have in mind is one where important future production decisions cannot be specified in a long-run contract. If the agents disagree, then the agent who owns an asset has the unilateral right to make a decision relating to this asset. This determines the disagreement (reservation) payoffs. The production decision made by an agent will matter not only to himself but also to the other agent, and the reservation payoffs depend on the agents' privately known types as well as on control rights.[2] This distinguishes us from Matouschek [16], who also introduced private information into the GHM framework. In his model, a seller owns a non-tradable asset which can be used to produce, at a privately known cost to the seller, an input which is sold to a buyer with a privately known valuation. Disagreement payoffs do not depend on the agents' private information, but do depend on who owns two tradable assets. Under some conditions, the disagreement payoff

[2] Consider Grossman and Hart's [9] example of an electricity generating plant located next to a coal mine. If the former acquires the latter then the former has the right to determine the ash content of the coal. Suppose how much a lower ash content is worth to the power plant is its private information (its type). The power plant's profit will depend on the decision (the ash content) as well as on its type. Inasmuch as the manager of the coal mine is not replaced, his payoff will typically also depend on the decision and on his type (which may involve the cost of lowering the ash content). Even if the manager is replaced, the power plant's profit will still depend on both the ash content and its type. Hence, reservation payoffs are type-dependent.

should be high as bargaining is likely to break down; in other cases it should be low to facilitate bargaining. However, it seems natural that there should be some connection between payoffs from trade and no-trade. For example, if an efficient seller has a low production cost within a relationship with a particular buyer, then his disagreement payoff should be high as he (presumably) also has a relatively low cost of producing for some other buyer. An explicit trading model generating the payoffs at the outside option would be necessary to derive optimal asset ownership. This is the approach we take. In our model, disagreement payoffs are type-dependent, and our key results require knowledge of how the binding participation constraints depend on types as well as on control rights.

In our model there are two agents $A$ and $B$, and two production decisions $q_A$ and $q_B$. Each agent $i \in \{A, B\}$ has a privately known *type* denoted $\theta_i$, which represents his idiosyncratic preference over $q_i$. The decisions $q_A$ and $q_B$ are noncontractible ex ante (before $\theta_A$ and $\theta_B$ are realized). A contract specifying $q_A$ and $q_B$ can be negotiated at the interim stage, when agent $i$ knows $\theta_i$, with residual decision (or control) rights determining the outside option. As in GHM, these decision rights are derived from asset ownership. That is, there are two assets $A$ and $B$, and the owner of asset $i$ has the right to choose $q_i$. With *nonintegration*, there is separate ownership of the assets so each agent $i \in \{A, B\}$ has the right to choose $q_i$. With *integration*, one agent ("the boss") owns both assets and has the right to choose both $q_A$ and $q_B$.

The decision $q_i$ can symbolize many different types of actions. But to be specific, suppose $q_i$ is the decision to either make agent $i$'s product "exclusive", $q_i = E$, or "inclusive", $q_i = I$. An exclusive product of agent $i$ works mainly if not solely with agent $j$'s product. An inclusive product of agent $i$ is designed to work with agent $j$'s product but also works well on other agents' products. If $q_i = I$ and $q_j = E$ then agent $i$ "holds up" agent $j$ (because agent $i$ sells his product on all platforms while agent $j$ has tied his fortune to agent $i$), and rents $\rho > 0$ are transferred from agent $j$ to agent $i$. If $q_A = q_B$ then each agent enjoys a coordination benefit $\gamma > 0$.

For example, suppose agent $A$ produces a platform and agent $B$ a product such as an app or piece of hardware that works on that platform. Agent $A$ could choose to customize the platform so it works better with agent $B$'s product than with other brands ($q_A = E$). Or, agent $A$ could facilitate the creation of products that work on multiple platforms ($q_A = I$). Similarly, agent $B$'s product could be customized to work with agent $A$'s platform

($q_B = E$) or on multiple platforms ($q_B = I$). Microsoft's Windows Phone, Google's Android or Apple's iOS are examples of mobile operating systems that correspond to the standards on one side of a relationship. Nokia, Motorola and Apple smartphones are examples of the products on the other side. Google's Android is an inclusive standard while Apple's iOS is exclusive. HTC and Samsung make phones that work on Android and Windows Phone. For another example, Github, now part of Microsoft's cloud computing platform, is based on the open-source software LINUX. This allows app developers to easily create products that work on both Github and other platforms (see Greene [7]). Both the platform and the app developers have coordinated on an inclusive solution. An exclusive scenario would be subject to hold-up on both sides. The cloud service suffers if an app is sold on multiple platforms rather than just its own. An app developer that runs its product exclusively on one cloud service may worry that the cloud service will allow competitors onto the platform or even produce the app itself. This is a concern of developers on Amazon's cloud service (see Greene and Stevens [8]).

We use the revelation principle to study whether the *first-best* decision profile can be implemented under integration or nonintegration.[3] If no contract is signed then each agent gets his reservation payoff, which depends on his type and on the allocation of control rights. Each agent's *critical type* is the type whose participation constraint is most difficult to satisfy. The key result is that implementation is possible if and only if the sum of the critical types' reservation payoffs is not too large. Intuitively, if this sum is big, then any agreement must give these critical types a big expected payoff. But other types can always *pretend* to be a critical type. To prevent this preference falsification (which we refer to as "haggling"), *all* types must get a big expected payoff. Implementability requires that the first-best yields enough surplus to accomplish this. Understanding the implications of this key result requires an analysis of how the reservation payoffs depend on control rights and on the severity of the hold-up problem, which in turn depends on the gains to coordination ($\gamma$), rent seeking ($\rho$), and the realized types ($\theta$).

With nonintegration, the critical types will try to extract rents $\rho$ (by choosing $I$) if negotiations break down. Their reservation payoffs are big

[3] As long as the first-best can be implemented by either integration or nonintegration, there is no need to consider more general governance structures (as these could not improve). See Section 2.3 for further discussion.

when $\rho$ is big, making it impossible to implement the first best. Intuitively, to persuade these types to sign a contract, they must be compensated for the forgone rents. This compensation is large when $\rho$ is large. Since even a type who would choose $E$ can *pretend* to be a type that would choose $I$, incentive-compatibility requires that *all* types benefit from the "compensation". When $\rho$ is large, the first-best does not generate enough surplus to achieve this. On the other hand, if $\rho$ is small then the required compensation is small and there is enough surplus at the first-best to satisfy all constraints.

With integration, two cases must be considered separately. In the *co-operative case*, the rent $\rho$ is small compared to the coordination benefit $\gamma$; in the *rent-seeking case,* $\rho$ is instead large compared to $\gamma$. For an intuitive understanding of the two cases, consider the relationship-specificity of the technology and the benefits of customization. In the hardware/software example, a hardware manufacturer may make "exclusive" design choices that maximize the utility of a specific operating system; similarly, the operating system can be designed to exploit patents held by this particular hardware manufacturer. These choices release coordination benefits whose magnitude depends on the differentiation between products on each side of the trade. If the technology is not particularly relationship-specific, say because different hardware manufacturers use the same chips and different operating systems run on the same chips, then the rent-seeking case is more likely to occur, because an inclusive strategy can capture revenue from multiple platforms. But if different hardware manufacturers produce different chips, then designing an operating system that runs on multiple chips may not be worthwhile, and the hardware manufacturer may not find much to gain by working with different software producers who each want their own design. In this case, the technology is highly relationship-specific, and the cooperative case is more likely to occur.

Consider integration in the cooperative case ($\gamma > \rho$). Naturally, the boss has a high reservation payoff, as he gets his preferred action profile if negotiations break down. But the subordinate also has a high reservation payoff, because in the cooperative case the boss prefers to coordinate ($q_A = q_B$). Thus, the boss cannot make a credible threat to extract rents from the subordinate if negotiations break down. Since the critical types of both agents enjoy large reservation payoffs, there will be insufficient surplus at the first-best to satisfy all the constraints.

Now consider integration in the rent-seeking case ($\rho > \gamma$). With high rents $\rho$ to be extracted from hold-up, the boss has a credible threat to hold up his subordinate (forcing him to choose $E$ while the boss chooses $I$). Thus, if $\rho$ is high the subordinate has a low, and the boss a high, reservation payoff. Rent-extraction is inefficient since it involves giving up coordination benefits, so as $\rho$ increases the subordinate's reservation payoff decreases faster than the boss's reservation payoff increases. Therefore, the *sum* of the critical types' reservation payoffs is small if $\rho$ is big (compared to $\gamma$), so the integrated firm can implement the first-best in the rent-seeking case.[4]

These results formalize the Williamsonian idea that the costs of haggling depend on the organizational form. However, these costs are not necessarily minimized in an integrated firm. There are parameter regions where the (private) gains from rent-seeking are larger than coordination benefits, and the first-best can be implemented with integration but not with nonintegration. But for parameter regions where coordination benefits are larger than the potential gains from rent-seeking, the first-best can be implemented with nonintegration but not with integration.

The rest of the paper is organized as follows. Section 2 presents the basic model. Section 3 uses the methods of Myerson and Satterthwaite [18] to compute expected payoffs for incentive-compatible first-best decision rules. Section 4 studies whether interim participation constraints can be satisfied with nonintegration, while Section 5 does the same for integration. Section 6 summarizes and interprets the results. Concluding remarks can be found in Section 7.

# 2 Model

## 2.1 Decisions and Payoff Functions

Consider a relationship involving two agents, $A$ and $B$. Two decisions have to be made, $q_A$ and $q_B$. For simplicity, there are only two options, *exclusivity*, $E$, or *inclusivity*, $I$, for each decision. For $i \in \{A, B\}$, let $\theta_i$ denote agent $i$'s *type*. This type represents a private benefit from choosing $I$ (or, equivalently, cost of choosing $E$). Agent $i$'s type is his private information. The *state of the world* is denoted $\theta = (\theta_A, \theta_B)$.

[4] The surplus loss when the boss holds up the employee is reminiscent of Williamson's "cost of fiat". However, in our model this cost is incurred only *out of equilibrium*. Since the desire to avoid this cost lubricates bargaining, it is analogous to the value of a Williamsonian "hostage".

Agent $i$'s payoff function is $v_i(q, \theta_i) + t_i$, where $v_i(q, \theta_i)$ is the benefit type $\theta_i$ derives from decision profile $q = (q_A, q_B)$ and $t_i$ is a monetary transfer. An *outcome* of the game consists of a decision profile $q = (q_A, q_B)$ and a transfer profile $t = (t_A, t_B)$. Budget balance requires that the transfers always sum to zero: $t_A + t_B = 0$. There are no individual limited liability constraints.

We represent $v_i(q, \theta_i)$ in the matrix (1). The row indicates $q_i$ and the column indicates $q_j$.

$$\begin{array}{ccc} & I & E \\ I & \theta_i + \gamma & \theta_i + \rho \\ E & -\rho & \gamma \end{array} \tag{1}$$

This payoff matrix is meant to capture the basic idea that the relationship has cooperative, opportunistic (rent-seeking) and idiosyncratic elements. If the decisions are coordinated (either both $I$ or both $E$) then $\gamma > 0$ is added to each agent's payoff. If $q_i = I$ and $q_j = E$ then agent $i$ gains $\rho > 0$ at the expense of agent $j$. Finally, $\theta_i$ (which can be positive or negative) represents an idiosyncratic benefit agent $i$ gets from $q_i = I$.

There are two salient cases. In the *cooperative case* we have $\gamma > \rho$. In this case, the benefit of coordination dominates the possible gains from rent seeking, and the biggest number in payoff matrix (1) is always on the main diagonal. In the *rent-seeking case* we have $\rho > \gamma$. In this case, the gain from rent seeking dominates the benefit of coordination. The incentives to cooperate or rent-seek will determine the optimal allocation of control rights.

Types are independently drawn from a distribution with a differentiable c.d.f. $F$ that has bounded support $\left[\underline{\theta}, \overline{\theta}\right]$. We make the following two assumptions on $F$.

**Assumption 1.** $F'(\theta_i) < 1/(2\gamma)$ for all $\theta_i \in \left[\underline{\theta}, \overline{\theta}\right]$. Also, $\underline{\theta} < -\rho - \gamma$ and $\bar{\theta} > 2\gamma$.

**Assumption 2.** $\underline{\theta} = -\overline{\theta}$ and $F(\theta_i) = 1 - F(-\theta_i)$ for all $\theta_i \in \left[\underline{\theta}, \overline{\theta}\right]$.

Assumption 1 says that there is "sufficient uncertainty"; the density $F'$ is rather flat and not concentrated around one point. This will be used to show uniqueness of equilibrium when each agent $i$ chooses $q_i$ non-cooperatively. Assumption 2 says that $F$ is symmetric around 0, so the idiosyncratic element $\theta_i$ is equally likely to favor $I$ or $E$.

Note that the payoff matrix (1) is highly stylized and symmetric in order to simplify calculations. For example, one might expect that the coordination benefit would be different, depending on which decision they coordinate on. Suppose $q_A = q_B = E$ means the products are optimized to work together, which raises profits by $\gamma_E$, while $q_A = q_B = I$ means each product is optimized to work on multiple platforms, which raises profits by $\gamma_I$. To simplify, we assume $\gamma_E = \gamma_I = \gamma$. Generalizing to allow for $\gamma_E > \gamma_I$ would be straightforward and would not change the qualitative insights. Whether integration or nonintegration can implement the first-best would still depend on whether the gains from rent-seeking, $\rho$, are large or small compared with the coordination benefits. However, the exact formulation would be more complicated if $\gamma_E \neq \gamma_I$. To avoid complications that do not add any insights, we assume $\gamma_E = \gamma_I = \gamma$.

## 2.2 Time-line

If the agents cannot sign complete (state-contingent) contracts ex ante, will there necessarily be a loss in surplus (compared to the first-best)? What is the role of asset ownership? To answer these questions, we rule out *ex ante* comprehensive contracting, but allow the agents to freely negotiate a contract at the *interim* stage (after they have learned their own types). Thus, interim contracting is frictionless: incentive and participation constraints must be respected, but on top of that there are no *additional* Williamsonian haggling inefficiencies. Following the mechanism design approach, we model efficient interim negotiations by assuming an impartial mediator proposes an *incentive-compatible revelation mechanism*. A *first-best mechanism* is an incentive-compatible revelation mechanism that specifies the first-best decision profile in every state. The idea of a "mediator" is borrowed from the mechanism design literature. It is a convenient way to model negotiations, but the mediator should not be thought of as an actual person ("social planner").

At the ex ante stage, before the agents know their own types, the mediator allocates control rights (asset ownership). If he allocates one asset to each agent, we call this *nonintegration*. If he allocates both assets to the same agent, say agent $A$, we call this *integration*. The time-line is the following.

**Stage 0.** Control rights are allocated (the mediator chooses either "integration" or "nonintegration").

**Stage 1.** Each agent $i \in \{A, B\}$ privately observes his own type $\theta_i$.

**Stage 2.** The mediator proposes an incentive-compatible revelation mechanism $\Gamma$. Then each agent (simultaneously) either accepts or rejects $\Gamma$. If both accept, then move to stage 3a, otherwise move to stage 3b.

**Stage 3.** (a) If at stage 2 both agents accepted $\Gamma$, then each agent reveals his type and an outcome is implemented as specified by $\Gamma$. (b) If at least one agent rejected $\Gamma$ at stage 2, then under *integration* agent $A$ chooses both $q_A$ and $q_B$, while under *nonintegration* agent $A$ chooses $q_A$ and agent $B$ chooses $q_B$.

Decisions and transfers are contractible in the sense that the outcome at stage 3a is final and not subject to moral hazard or ex post renegotiation. This frictionless *full-commitment* benchmark for the interim negotiations will reveal whether surplus losses are a necessary consequence of the inability to sign complete contracts ex ante.

At stage 2, each agent has an outside option, namely, to refuse to participate in $\Gamma$ and move to stage 3b. For all types to participate, $\Gamma$ must satisfy interim participation, or individual rationality (IR), constraints: each type must expect to get at least his reservation payoff, i.e., what he expects to get if stage 3b is reached. This reservation payoff depends on the allocation of control rights: under integration, agent $A$ chooses both $q_A$ and $q_B$ at stage 3b, while under nonintegration, agent $A$ chooses $q_A$ and agent $B$ chooses $q_B$ simultaneously and independently. (No monetary transfers are ever made at stage 3b.) Again, the outcome is final and not subject to renegotiation. If a mechanism $\Gamma$ satisfies all interim IR constraints, then $\Gamma$ is said to be *individually rational*. It is important to note that in our model, individual rationality always refers to *interim* constraints.

The mediator's objective is to maximize the social surplus. We will characterize the conditions under which a first-best mechanism can satisfy the IR constraints with integration or nonintegration. If the IR constraints can be satisfied with integration (resp. nonintegration), then the social surplus is maximized by choosing integration (resp. nonintegration) at stage 0.

## 2.3 Discussion

### 2.3.1 Interim stage

By having a mediator propose a mechanism at stage 2, we can identify whether, in principle, the first best is implementable by *some* mechanism.

By the revelation principle, we may focus on incentive-compatible revelation mechanisms. The cost of haggling will be at the theoretical minimum, in the sense that negotiations are efficient and subject only to IC and IR constraints. Asset ownership determines the IR constraints, but assets have no intrinsic utility; all that matters is the resulting decisions. For this reason, there is nothing to be gained by allowing the revelation mechanism to reallocate assets.

An important assumption is that the contracting opportunities are the same in both the integrated and the nonintegrated case. In reality, a dispute within an integrated firm cannot be settled in court, making it harder to enforce contracts (see Williamson's [34] discussion of "forebearance"). However, real-world organizations do contain mechanisms for aggregating information and making decisions, perhaps enforced by reputational concerns. In this paper, we isolate the impact of incomplete information and interim contracting on optimal organization form. Thus, we abstract from the (interesting) possibility that different organizational forms face different contracting frictions, such as limits on enforcement.

We rule out renegotiation of the mechanism $\Gamma$. Suppose after an agent has rejected a proposed mechanism, renegotiation could occur; perhaps the agents could engage a new "mediator" to help them negotiate a new agreement. This would strengthen the IR constraints in the original mechanism, reducing the attainable surplus (under both integration and nonintegration). Gains from coordination and rents from hold-up may interact in different ways under renegotiation, which is an interesting topic for future study. However, the results would depend on the exact details of how renegotiation occurs, and on how beliefs about the state change when a mechanism is rejected. The study would also have to face a possible infinite regress: if the renegotiation under the "new mediator" breaks down, the agents may re-renegotiate under a "new new mediator", etc.

Another interesting challenge would be to eliminate the mediator(s), and model the interim contracting stage in a less abstract way. For example, if agent $A$ owns both assets then he might propose a decision-making procedure (a mechanism).[5] Such a study would face the problem of *information leakage*: agent $B$ might infer agent $A$'s type from the procedure he proposes (player $A$ would be an "informed principal"). The results would again depend on

[5] With nonintegration a more symmetric procedure could be used; perhaps each agent is equally likely to propose a mechanism.

the exact details of the game, and on how agent $B$'s beliefs are formed. The mediator approach allows us to bypass these issues, and derive a theoretical benchmark from which we can derive intuitions about key variables.

### 2.3.2 Ex ante stage

At the ex ante stage, the allocation of control rights is determined. In the spirit of the mechanism design literature, the mediator chooses integration or nonintegration. Again, the mediator could be dropped at the cost of making the game more involved. In an alternative formulation, the agents would acquire assets at the ex ante stage without the help of a mediator. This would not change the main results: the conditions under which integration or nonintegration maximizes the surplus. These results are driven by the key "incomplete contracts" assumption: the allocation of control rights is the *only* instrument available for influencing future Inclusivity/Exclusivity decisions. Thus, the two agents cannot, at the ex ante stage, contract on future decisions. Neither can they influence the payoffs they get from these decisions. Note that agent $i$'s payoff from decision profile $(q_A, q_B)$ is independent of who owns which asset. Asset ownership thus has no intrinsic utility, but conveys "pure control rights" in the parlance of Segal and Whinston [26]. Payoffs depend on decisions, but *ex ante* it is impossible to contractually influence the payoff an agent gets from a particular decision.

It is useful to keep in mind a scenario where the two agents have already acquired their private information the first time they meet. Even if their interaction will last a long time, our model applies as long as their types are persistent over time. This might be the case, for example, if the private information relates to personal skills or knowledge. Suppose instead the interaction will last a long time and the agents draw new types at some point. If their initial types are privately known, then the attainable surplus will initially depend on asset ownership as in our current (static) model. But if future production decisions can be contracted on, and if they can commit to using an AGV mechanism in the future, then the future surplus is independent of the initial ownership. However, such contracting on future production decisions would go against the spirit of the incomplete contracts literature (see Section 7 for further discussion).

Consider several ways our current game could be changed to improve the ex ante contracting abilities. First, suppose that at stage 0, the mediator first allocates control rights, and then proposes a revelation mechanism $\Gamma$

which specifies Inclusivity/Exclusivity decisions as a function of announcements made at the interim stage. The agents must accept or reject $\Gamma$ *before learning their own types.* If someone rejects $\Gamma$, then at the interim stage the agents make decisions non-cooperatively, based on their initially assigned control rights. But if both accept, they are committed to playing $\Gamma$ at the interim stage. This commitment removes the interim participation constraints: they are *forced* to play $\Gamma$ interim. Then for any initial allocation of control rights, the mediator can always implement the first-best by proposing an AGV mechanism (d'Aspremont and Gérard-Varet [3]) which satisfies *ex ante* participation constraints (see Section 3). This corresponds to a complete state-contingent contract; the outcome for each state would be specified ex ante by the AGV mechanism. The issue of integration versus nonintegration is then moot.[6]

Next, suppose at the ex ante stage the agents cannot contract on state-contingent Inclusivity/Exclusivity decisions, but they can contract on a randomized (but not state-contingent) "status quo" decision profile $\tilde{q}$. At the interim stage, they can contract on a revelation mechanism $\Gamma$ , which must satisfy participation constraints with respect to $\tilde{q}$. Segal and Whinston [25] show that if $\tilde{q}$ has the same distribution as the first-best decision profile, then an interim individually rational AGV mechanism exists.[7] Their argument is as follows: agent $i$ of type $\theta_i$ could lie and randomize his report to equal the distribution $F$. This would give him a payoff equal to the outside option when agent $j$ tells the truth. But this deviation cannot be profitable as IC is satisfied by the AGV mechanism. Hence, if IC holds then interim IR also holds. Because decisions are contractible ex ante, they are divorced from ownership of assets. Again, the first-best can always be implemented, so the issue of integration versus nonintegration is moot. We rule out such ex ante contracts. In the technology example, the parties themselves may know what a customized or standardized product might be, but it is hard to describe the features ex ante. Describing a randomization over features would be even

[6] The issue would also be moot for contracting under *complete information*, where each agent knows $\theta = (\theta_A, \theta_B)$. Consider the following mechanism. Agent $A$ makes a take-it-or-leave-it offer consisting of a proposed decision profile and a transfer to agent $B$. If agent $B$ rejects the offer, then he gets some reservation payoff (which may depend on control rights). In subgame perfect equilibrium, agent $A$ proposes the first-best decision profile and a transfer such that agent $B$ gets exactly his reservation payoff.

[7] The constants $k_i$ in the AGV mechanism can be chosen so each agent's ex ante expected payment is zero. See Segal and Whinston [25] for details.

harder. In the spirit of the incomplete contracts literature, we rule out ex ante contracting on mechanisms or on randomized decision profiles.

The allocation of control rights is the only thing to be determined ex ante. Specifically, the mediator chooses either integration or nonintegration. Restricting the mediator to these two options is without loss of generality if the parameter values are such that one of them implements the first-best. But for other parameter values, alternative allocations of control rights might do better. Indeed, one particular allocation is guaranteed to do so: the mediator keeps all the control rights for himself. He can then choose the randomized decision profile $\tilde{q}$ discussed in the previous paragraph if any agent rejects the proposed mechanism at stage 2. By the argument of the previous paragraph, the first-best is implemented. (Of course, in practice this solution would be irrelevant if for some reason it is impractical to give decision rights to a third party.) Alternatively, the mediator could allocate *stochastic* control rights: if $\Gamma$ is rejected at stage 2 then a randomizing device determines which agent has which control right. With some probability, the control rights could be reversed (agent $A$ chooses $q_B$ and agent $B$ chooses $q_A$). Finally, the mediator could lower reservation values by allocating *contingent* control rights: an agent who rejects $\Gamma$ is forced to give up (or sell) his control rights to the other agent. If neither integration nor nonintegration can implement the first-best, then stochastic or contingent control rights could improve. Of course, in practice if might be difficult to verify which agent has refused to participate, hence contingent control rights would be hard to enforce. Similarly, in practice it may be difficult to commit to a randomization.

If we were to consider which allocation of control rights is optimal for *all* parameter values, the solution is either known but hard to interpret (let the mediator hold the control rights), or we have to rule out such unappealing "solutions". However, we do not know a systematic basis for doing this, i.e., where to draw the line between feasible ("reasonable") and infeasible ("unreasonable") governance structures. If the mediator is not allowed to hold the control rights, can he allocate randomized or contingent rights? We avoid this issue by restricting attention to parameter values where either integration or non-integration implement the first-best.

# 3 Incentive Compatible First-Best Mechanisms

We begin by identifying first-best allocations. The *social surplus* is the sum of agent $A$'s and agent $B$'s payoff. We represent the social surplus in the following matrix. The row indicates $q_A$ and the column indicates $q_B$.

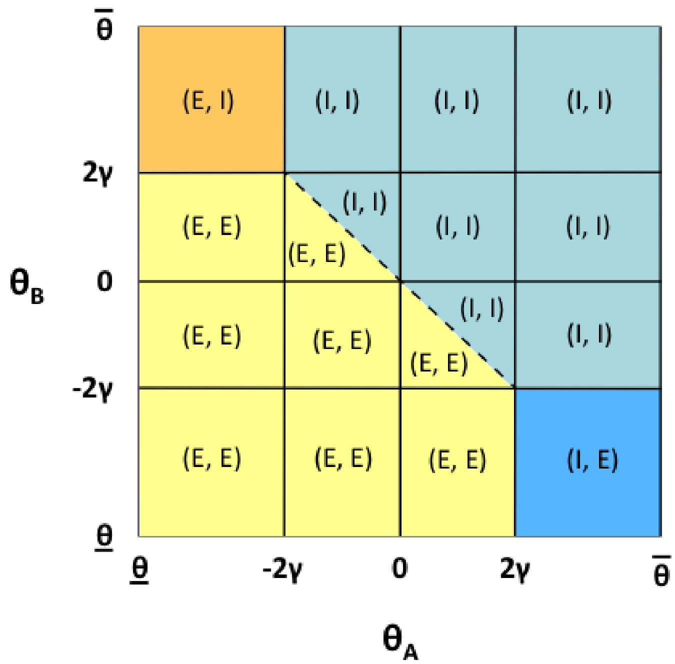
$$\begin{array}{ccc} & I & E \\ I & \theta_A + \theta_B + 2\gamma & \theta_A \\ E & \theta_B & 2\gamma \end{array} \tag{2}$$

Note that neither $\rho$ nor $(t_A, t_B)$ matter for social surplus.

For any type profile $\theta = (\theta_A, \theta_B)$, the *first-best* (socially optimal) decision profile is denoted $q^*(\theta) = (q_A^*(\theta), q_B^*(\theta))$. By definition, $q^*(\theta)$ maximizes the social surplus, i.e., it selects the biggest number in the matrix (2).[8] Figure 1 illustrates the first-best. It is easy to see that if $\theta_i \leq -2\gamma$ (**Case 1**) or $\theta_i \geq 2\gamma$ (**Case 3)** it is optimal to set $q_i^*(\theta) =E$ and $q_i^*(\theta) =I$ respectively. In the intermediate case (**Case 2**), where $-2\gamma < \theta_i < 2\gamma$, the surplus maximizing decision is always on the main diagonal of the matrix (2). Specifically, $q^*(\theta) = (E,E)$ if $\theta_A + \theta_B < 0$, and $q^*(\theta) = (I,I)$ if $\theta_A + \theta_B > 0$.

[8] If two numbers are the same, either decision profile can be chosen, but this will happen with probability zero.

Figure 1: First Best




We can sum payoffs and integrate to calculate the *ex ante* expected social surplus generated under the first-best decision rule:

$$S^* = 4\gamma F(2\gamma) - \int_{-2\gamma}^{2\gamma} (F(s))^2 ds + 2\int_{2\gamma}^{\overline{\theta}} s dF(s). \tag{3}$$

We will next consider a first-best mechanism $\Gamma$ and study the constraints imposed by incentive compatibility. We will use the first-best allocations $q^*(\theta)$ to compute any type's expected payoff from $\Gamma$, assuming all types participate (later, we will check if interim IR constraints hold). The agents are symmetric, so it suffices to consider agent $A$.

Let

$$t(\theta_A) \equiv \int_{\underline{\theta}}^{\overline{\theta}} t_A(\theta_A, \tilde{\theta}_B) dF(\tilde{\theta}_B)$$

denote agent $A$'s expected transfer when his type is $\theta_A$, and let

$$u_A(\theta_A) \equiv \int_{\underline{\theta}}^{\overline{\theta}} v_A(q^*(\theta_A, \tilde{\theta}_B), \theta_A) dF(\tilde{\theta}_B) + t(\theta_A)$$

denote type $\theta_A$'s expected payoff. The *ex ante* expected payoff for agent $A$ is

$$V_A \equiv \int_{\underline{\theta}}^{\overline{\theta}} u_A(\theta_A) dF(\theta_A).$$

There are three cases.

**Case 1:** $\theta_A \leq -2\gamma$. Then $q_A = E$. Agent $A$ gets $\gamma$ with probability $F(2\gamma)$ and $-\rho$ with probability $1 - F(2\gamma)$. Thus, agent $A$'s expected payoff is

$$u_A(\theta_A) = t(\theta_A) + \gamma F(2\gamma) - \rho(1 - F(2\gamma)) = t(\theta_A) - \rho + (\gamma + \rho) F(2\gamma). \quad (4)$$

**Case 2:** $-2\gamma < \theta_A < 2\gamma$. Then the decision is ($E$,$E$) if $\theta_B < -\theta_A$, and ($I$,$I$) otherwise, so agent $A$ expects

$$u_A(\theta_A) = t(\theta_A) + \gamma + \theta_A(1 - F(-\theta_A))) = t(\theta_A) + \gamma + \theta_A F(\theta_A). \quad (5)$$

using Assumption 2.

**Case 3:** $\theta_A \geq 2\gamma$. Then $q_A = I$. Agent $A$ gets $\gamma$ with probability $1 - F(-2\gamma) = F(2\gamma)$, and $\rho$ with probability $1 - F(2\gamma)$. Thus, agent $A$ expects

$$u_A(\theta_A) = t(\theta_A) + \theta_A + \gamma F(2\gamma) + \rho(1 - F(2\gamma)) = t(\theta_A) + \rho + \theta_A + (\gamma - \rho) F(2\gamma). \quad (6)$$

Incentive compatibility imposes restrictions on the transfer function. In Case 1, the decision profile is independent of $\theta_A$, so the expected transfer must equal some constant $t_A$ for all such types, and then from (4) the expected payoff is constant as well. Thus, for all $\theta_A < -2\gamma$ we have $t(\theta_A) = t_A$ and

$$u_A(\theta_A) = \underline{u}_A \equiv t_A - \rho + (\gamma + \rho) F(2\gamma). \quad (7)$$

In Case 2, the arguments of Myerson and Satterthwaite [18] imply that $u_A$ is differentiable almost everywhere, and $u'_A(\theta_A) = F(\theta_A)$ for $\theta_A$ such that $-2\gamma < \theta_A < 2\gamma$. This implies that

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \tag{8}$$

if $-2\gamma < \theta_A < 2\gamma$.

In Case 3, the arguments of Myerson and Satterthwaite [18] imply $u'_A(\theta_A) = 1$. Using integration by parts and Assumption 2, this implies that

$$\begin{aligned} u_A(\theta_A) &= u_A(2\gamma) + \theta_A - 2\gamma = \underline{u}_A + \int_{-2\gamma}^{2\gamma} F(s)ds + \theta_A - 2\gamma \\ &= \underline{u}_A + \theta_A \end{aligned} \tag{9}$$

if $\theta_A > 2\gamma$.

Let $V_i$ be agent $i$'s *ex ante* expected payoff. We can combine the three cases to compute $V_A$ and $V_B$ and sum to get:[9]

$$\begin{aligned} & V_A + V_B \\ = \; & \underline{u}_A + \underline{u}_B + 2\left(\int_{-2\gamma}^{2\gamma}\left[\int_{-2\gamma}^{\theta} F(s)ds\right]dF(\theta) + \int_{2\gamma}^{\overline{\theta}} \theta dF(\theta)\right) \\ = \; & \underline{u}_A + \underline{u}_B + 4\gamma F(2\gamma) - 2\int_{-2\gamma}^{2\gamma} (F(s))^2\, ds + 2\int_{2\gamma}^{\overline{\theta}} sdF(s). \end{aligned} \tag{10}$$

Budget balance requires that the sum of the expected payoffs equals the expected social surplus:

$$V_A + V_B = S^*. \tag{11}$$

Using (10) and (11), we obtain:

$$\underline{u}_A + \underline{u}_B = \int_{-2\gamma}^{2\gamma} (F(s))^2\, ds. \tag{12}$$

[9] This calculation uses the fact that

$$\begin{aligned} \int_{-2\gamma}^{2\gamma}\left[\int_{-2\gamma}^{\theta_i} F(s)ds\right]dF(\theta_i) &= F(2\gamma)\int_{-2\gamma}^{2\gamma} F(s)ds - \int_{-2\gamma}^{2\gamma} (F(\theta_i))^2 d\theta_i \\ &= F(2\gamma)2\gamma - \int_{-2\gamma}^{2\gamma} (F(\theta_i))^2 d\theta_i. \end{aligned}$$

We will use this minimum rent that low types must receive to study when the first-best is implementable.

The well-known AGV mechanism is a first-best mechanism. Given the revealed types $(\theta_A, \theta_B)$, the first-best decision $q^*(\theta_A, \theta_B)$ is chosen, and agent $A$ gets the AGV transfer

$$t_A(\theta_A, \theta_B) \equiv \int_{\underline{\theta}}^{\overline{\theta}} v_B(q^*(\theta_A, \tilde{\theta}_B), \tilde{\theta}_B) dF(\tilde{\theta}_B)$$

$$- \int_{\underline{\theta}}^{\overline{\theta}} v_A(q^*(\tilde{\theta}_A, \theta_B), \tilde{\theta}_A) dF(\tilde{\theta}_A) + k_A \tag{13}$$

where $k_A$ is a constant. Player $B$'s transfer is given by the analogous expression, with a constant $k_B$. Budget balance requires $k_A + k_B = 0$.[10]

# 4 Implementing the First-Best with Nonintegration

Whether or not a first-best mechanism satisfies IR constraints depends on asset ownership. We begin by studying nonintegration, where each agent owns his own asset. Thus, if stage 3b is reached, each agent $i \in \{A, B\}$ chooses $q_i$ knowing his own type $\theta_i$ but not the other agent's type $\theta_j$. We will identify the type whose IR constraint is the most difficult to satisfy. Recall that all IR constraints are *interim* constraints.

Consider a noncooperative (Bayesian-Nash) equilibrium of this stage 3b game.[11] Since higher types are more inclined to choose $I$, each agent must

[10] As alluded to in Section 2.3, the ex ante participation constraints can be satisfied by a judicious choice of $k_A$. Agent $i$'s ex ante reservation payoff $r_i^e$ would depend on how decisions are expected to be made if someone rejects the mechanism. However, since the AGV mechanism is first best, its total expected surplus can be no smaller than $r_A^e + r_B^e$. Setting $k_A$ appropriately ensures that each agent $i$ gets at least $r_i^e$ and so is willing, ex ante, to accept the mechanism.

[11] To simplify, we make two assumptions about the noncooperative game played at stage 3b: (i) *passive beliefs*: each agent maintains his prior beliefs about the other agent's type, and (ii) *no cheap talk*. Either assumption could be changed, and the results would be qualitatively the same. Quantitatively the results would change, because the reservation payoffs would change. But the disagreement point would still not be first best, and control rights would still matter for the attainable surplus. Passive beliefs were introduced by Rubinstein [23] and is a common assumption. The model with cheap talk would be remi-

use a cutoff strategy. Suppose agent $B$ chooses $q_B = E$ if and only if $\theta_B \leq x$, which happens with probability $F(x)$. Then, agent $A$ prefers $q_A = E$ if

$$F(x)\gamma - (1 - F(x))\rho \geq \theta_A + F(x)\rho + (1 - F(x))\gamma$$

which is equivalent to

$$\theta_A \leq (2F(x) - 1)\,\gamma - \rho.$$

Thus, if agent $B$ uses cutoff $x$, agent $A$'s best response is to use the cutoff $y$ defined by

$$y = (2F(x) - 1)\,\gamma - \rho. \tag{14}$$

By Assumption 1, the best response function has slope less than one:

$$\frac{dy}{dx} = 2F'(x)\gamma < 1.$$

Therefore, there is a unique noncooperative equilibrium. By the symmetry of the game, this equilibrium must be symmetric, and it can be found by setting $x = y$ in (14). Thus, in the unique equilibrium, each agent uses the cutoff $\theta^*$ defined by

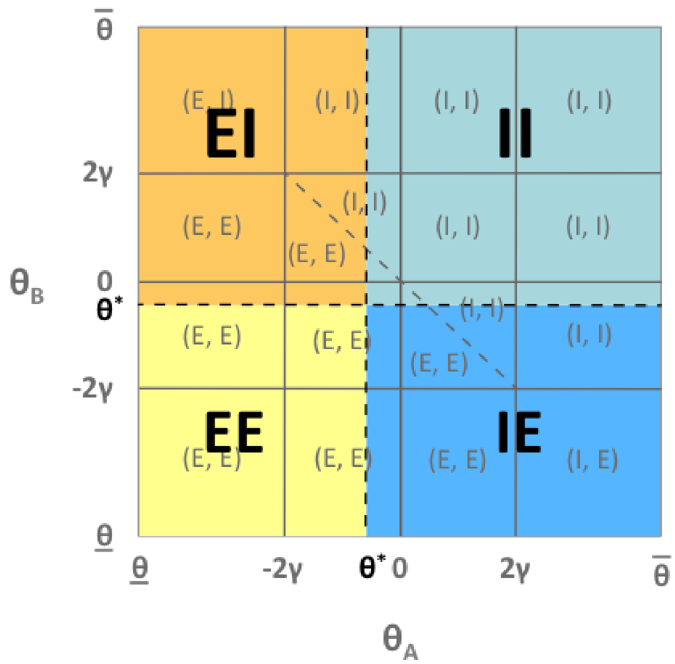
$$\theta^* + \gamma + \rho = 2F(\theta^*)\gamma. \tag{15}$$

It can be checked that $-\gamma - \rho < \theta^* < -\rho$. Figure 2 illustrates the noncooperative equilibrium under nonintegration.

---

niscent of Alonso, Dessein and Matouschek [1] and Rantakari [22], but with a comparison of integration and nonintegration, rather than centralization versus decentralization.

Figure 2: Nonintegration




Agent $i$'s noncooperative equilibrium expected payoff is

$$\gamma F(\theta^*) - \rho(1 - F(\theta^*)) = -\rho + (\gamma + \rho)F(\theta^*) \tag{16}$$

if $\theta_i < \theta^*$ and

$$\theta_i + \gamma + (\rho - \gamma)F(\theta^*) \tag{17}$$

if $\theta_i > \theta^*$.

Since the noncooperative equilibrium is played at stage 3b if $\Gamma$ is rejected at stage 2, agent $i$'s reservation payoff at stage 2 is given by (16) if $\theta_i < \theta^*$ and (17) if $\theta_i > \theta^*$.

Let $\Gamma$ be a first-best mechanism. Which type $\theta_i$ is most reluctant to accept $\Gamma$ at stage 2? If stage 3b is reached, there is no check on opportunism, and type $\theta_i > \theta^*$ benefits by playing $I$, thus gaining $\rho$ when $\theta_j < \theta^*$. Since $\rho$

is not a social gain, $\Gamma$ will tend to reduce this opportunistic benefit. Indeed, consider type $\theta_i$ such that $\theta^* < \theta_i \leq 2\gamma$. If he accepts $\Gamma$, the first-best decision rule will, with some probability, make him forego the opportunistic gain $\rho$ by implementing $q_i = E$. As the benefit from $q_i = E$ is decreasing in the type, in the range $[\theta^*, 2\gamma]$ the disadvantage of the first-best is the greatest for type $2\gamma$. On the other hand, if $\theta_i > 2\gamma$ then $q_i = I$ both in the noncooperative equilibrium and in the first-best, so the disadvantage is not any greater for types above $2\gamma$ than it is for type $2\gamma$. This intuitive argument suggests that the most difficult IR constraint to satisfy is for type $2\gamma$. That is, player $i$'s *critical type* is $\theta_i = 2\gamma$. Lemma 1 verifies that the argument is correct. (All missing results and proofs are in the Appendix.)

Next, note that under nonintegration, the noncooperative equilibrium favors type $2\gamma$ inasmuch as he can benefit from the opportunistic benefit $\rho$. Still, when $\rho$ is low, type $2\gamma$ does not need a big transfer to accept $\Gamma$. Then, other types would not gain much by pretending to be type $2\gamma$. Transfers to all types can be correspondingly low without violating IC. Then there is enough surplus at the first-best to satisfy all IC and IR constraints without violating budget balance. But when $\rho$ is high, it takes a big expected transfer to make type $2\gamma$ accept $\Gamma$. The IC constraints imply the expected transfer must be high to all types. In terms of the AGV mechanism, we need $k_A > 0$ and $k_B > 0$ which violates budget balance, so the first-best cannot be implemented. Using (12), we get the following result:

**Proposition 1** *With nonintegration, there exists an individually rational first-best mechanism if and only if*

$$\int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds \geq 2\rho F(\theta^*) + 2\gamma(1 - F(\theta^*)). \tag{18}$$

In the Appendix, we show this can be expressed in terms of rent $\rho$ and gains from coordination $\gamma$.

**Theorem 1** *With nonintegration, there exists an individually rational first-best mechanism if and only if $\rho \leq \rho^*(\gamma)$, where $0 < \rho^*(\gamma) < \gamma$.*

# 5 Implementing the First-Best with Integration

Now we study the IR constraints for a first-best mechanism $\Gamma$ under integration. Without loss of generality (since the agents are symmetric ex ante), let agent $A$ be "the boss". Thus, if stage 3b is reached, agent $A$ unilaterally chooses the decision profile $(q_A, q_B)$ which is best for himself, given $\theta_A$. That is, he picks the greatest number in the payoff matrix (1), for $i = A$. Notice that agent $A$'s decision doesn't depend on his beliefs about $\theta_B$, since $\theta_B$ doesn't influence his payoff.

Since agent $A$ can choose the decision profile unilaterally if stage 3b is reached, his reservation payoff at stage 2 is high. If agent $B$'s reservation payoff were correspondingly low, agent $B$ would be willing to pay a large transfer in a first-best mechanism. In this case, an AGV mechanism with a large positive $k_A$ would be individually rational. But agent $B$'s reservation payoff depends on agent $A$'s incentive to cooperate or rent-seek. In the cooperative case where $\gamma > \rho$, at stage 3b agent $A$ will never behave opportunistically (in the sense of transferring $\rho$ from agent $B$ to himself). In the *rent-seeking case* where $\rho > \gamma$, at stage 3b agent $A$ will sometimes choose $(q_A, q_B) = (I,E)$, a rent-seeking policy which transfers $\rho$ from agent $B$ to himself. Since the IR constraints differ in the two cases, we treat them separately. We emphasize the key fact that agent $B$'s reservation payoff at stage 2 is quite different in the two cases: high in the cooperative case (making it hard to satisfy the IR constraints) but low in the rent-seeking case (making it easy to satisfy the IR constraints).

## 5.1 The Cooperative Case

Suppose $\rho < \gamma$ and $\Gamma$ is rejected at stage 2. At stage 3b agent $A$ chooses $(E,E)$ if $\theta_A < 0$ and $(I,I)$ otherwise. Therefore, if agent $A$ rejects $\Gamma$ at stage 2 then his payoff at stage 3b will be

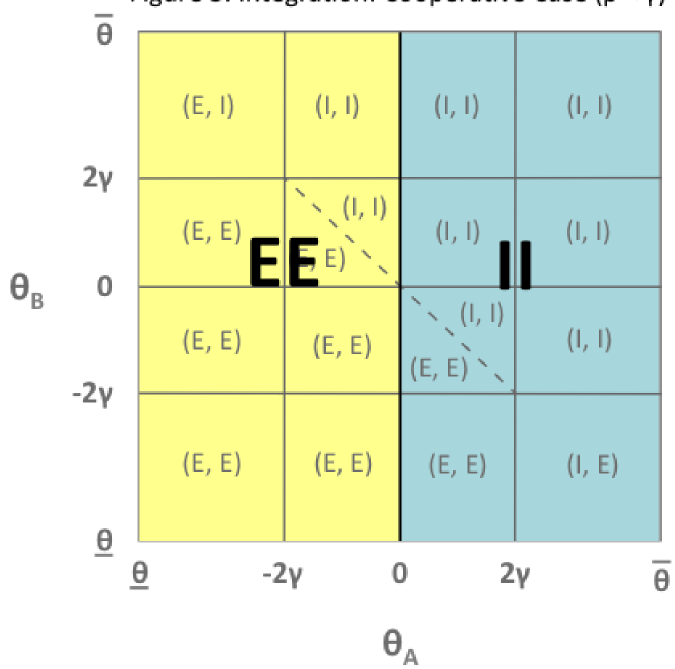
$$\gamma + \max\{0, \theta_A\}. \tag{19}$$

This is agent $A$'s reservation payoff at stage 2. Of course, since agent $A$ does not take $\theta_B$ into account, his decision at stage 3b is not necessarily first-best; the exception is when $\theta_B = 0$, since then agent $B$ is indifferent between $(E,E)$ and $(I,I)$ so whatever agent $A$ prefers will be socially efficient. At stage 3b agent $B$ gets $\gamma$ if $\theta_A < 0$ and $\theta_B + \gamma$ otherwise, so if agent $B$ rejects $\Gamma$ at stage 2 his expected payoff is

$$\gamma + (1 - F(0))\,\theta_B = \gamma + \theta_B/2. \tag{20}$$

This is agent $B$'s reservation payoff at stage 2. Figure 3 illustrates what happens at stage 3b.

Figure 3: Integration: Cooperative Case (ρ < γ)




Lemma 3 in the Appendix shows that $\theta_B = 0$ is agent $B$'s critical type. This is intuitive, because if type $\theta_B = 0$ rejects the first-best mechanism $\Gamma$ the boss would anyway choose the first-best decision. Type $\theta_B = 0$ would therefore reject any $\Gamma$ which requires him to make a payment to agent $A$. Type $\theta_B \neq 0$ would in principle be willing to pay to implement the first-best, because if he rejects $\Gamma$ then at stage 3b his preference over $(E,E)$ versus

$(I, I)$ would not be not taken into account. However, the payment that can be extracted from type $\theta_B \neq 0$ is constrained by the fact that he can haggle, pretending to be type $\theta_B = 0$ who is indifferent between $(E,E)$ and $(I, I)$ and therefore cannot be made to pay. Information rents will be too large. Intuitively, the boss will require a large "bribe" to give up his control rights and agree to $\Gamma$. But the possibility of haggling, and the fact that hold-up is not a credible threat in the cooperative case, means that only small bribes can be extracted from the subordinate. In terms of the AGV mechanism, persuading agent $A$ to give up his control rights requires that $k_A$ is large and positive. But budget balance forces $k_B = -k_A$ and type $\theta_B = 0$ would certainly not accept a large negative $k_B$. Therefore, the first-best cannot be implemented:

**Theorem 2** *If $\rho < \gamma$, then with integration no first-best mechanism is individually rational.*

Finally, combining Theorems 1 and 2 we get:

**Theorem 3** Cooperative Case. *Suppose $\rho < \gamma$. Then, with integration no first-best individually rational mechanism exists; with nonintegration such a mechanism exists if and only if $\rho \leq \rho^*(\gamma)$, where $0 < \rho^*(\gamma) < \gamma$.*

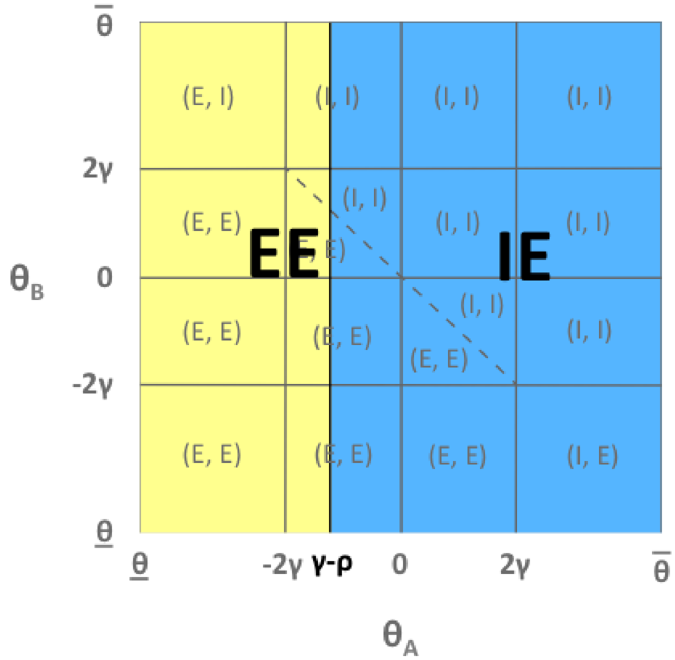
## 5.2 The Rent-Seeking Case

Suppose $\rho > \gamma$ and $\Gamma$ is rejected at stage 2. At stage 3b, agent $A$ chooses $(E,E)$ if $\theta_A < \gamma - \rho$ and $(I,E)$ otherwise. Thus, agent $A$'s reservation payoff is $\max\{\gamma, \theta_A + \rho\}$. At stage 3b agent $B$ receives $\gamma$ if $\theta_A < \gamma - \rho$ and $-\rho$ otherwise, so agent $B$'s reservation payoff is

$$F(\gamma - \rho)\gamma - \rho(1 - F(\gamma - \rho)) = -\rho + F(\gamma - \rho)(\gamma + \rho).$$

Notice that $\gamma - \rho < 0$ so $F(\gamma - \rho) < 1/2$. Figure 4 illustrates the stage 3b rent-seeking policy.

Figure 4: Integration: Rent-Seeking Case (γ < ρ)




At stage 3b, agent $A$ will always choose $q_B = E$ in the rent-seeking case. Since the high types of agent $B$ would derive a large benefit from $q_B = I$, they are eager to accept $\Gamma$. In contrast, the low types of $B$ do not mind if $q_B = E$. Indeed, if $\theta_B \leq -2\gamma$ then $q_B = E$ in the first-best for sure, i.e., $q_B$ will be the same at stages 3a and 3b when $\theta_B \leq -2\gamma$. These types all have the same IR constraint, and it is intuitively clear that this is most difficult IR constraint to satisfy for agent $B$. Lemma 4 in the Appendix shows this formally.

From Theorem 1, as $\rho^*(\gamma) < \gamma$, we already know that the first-best cannot be implemented under nonintegration. High rents imply that the benefits of haggling are high under nonintegration and this creates inefficiency. In contrast, there are circumstances where the first-best can be implemented under integration:

**Theorem 4** *If $\rho > \gamma$, then with integration a first-best individually rational mechanism exists if and only if*

$$\int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds \geq F(\gamma - \rho)(\gamma + \rho). \tag{21}$$

To interpret Theorem 4, notice that (21) holds in the following three scenarios: (i) when the idiosyncratic preference is relatively unimportant; (ii) when $\rho$ is only slightly bigger than $\gamma$; (iii) when $\rho$ is much bigger than $\gamma$.

To prove (i), define $z \equiv \rho - \gamma > 0$. We make $\theta_i$ less important, without changing the relative size of $\rho$ and $\gamma$, by simultaneously increasing $\rho$ and $\gamma$, while keeping $z \equiv \rho - \gamma$ fixed. Now (21) can be written as

$$\int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds \geq F(-z)(2\gamma + z).$$

Raising $\gamma$ increases the left hand side at the rate

$$2\left(F(2\gamma)\right)^2 + 2\left(F(-2\gamma)\right)^2 = 2\left(F(2\gamma)\right)^2 + 2\left(1 - F(2\gamma)\right)^2$$

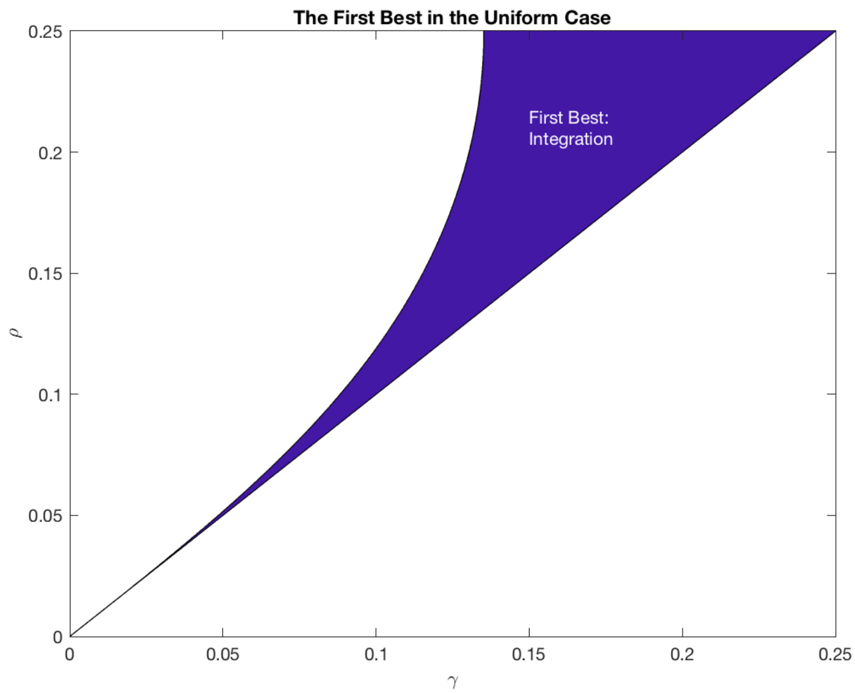
which always exceeds 1 and, in fact, will be close to 2 for $\gamma$ large. But the right hand side goes up only at the rate $2F(-z) < 1$ because $-z < 0$ and $F(0) = 1/2$. Therefore, (21) is bound to hold for $\gamma$ large enough, for any given $z$.

To prove (ii), note that (21) holds with strict inequality when $\rho = \gamma$, because the left side lies strictly between $\gamma$ and $2\gamma$ while the right side equals $\gamma$ when $\rho = \gamma$. Thus (21) also holds for $\rho$ slightly above $\gamma$.

To prove (iii), note that $F(\gamma - \rho) = 0$ whenever $\gamma - \rho < \underline{\theta}$. Therefore, (21) holds if $\rho > \gamma - \underline{\theta}$.

To visualize inequality (21), suppose $F$ is the uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. In Figure 5, the purple (or darkly shaded) area is the set of $(\gamma, \rho)$ such that $\rho > \gamma$ and the inequality (21) holds. Thus, if $(\gamma, \rho)$ belongs to the this area, the first-best can be implemented under integration. The northwest edge of the purple area is the set of $(\gamma, \rho)$ such that there is equality in (21). If both $\gamma$ and $\rho$ increase by the same amount then we move into the purple area (property (i)); the purple area includes all $(\gamma, \rho)$ slightly above the 45 degree line (property (ii)); and if $\rho$ is increased while $\gamma$ is fixed, we move into the purple area (property (iii)).

Figure 5




To understand why is it possible to implement the first-best with integration in the rent-seeking case, notice that if agent $B$ rejects $\Gamma$ then there is a chance that he will be held up at stage 3b (the outcome will be $(I,E)$ and agent $B$ loses $\rho$). This imposes an expected cost on all of player $B$'s types; in terms of the AGV mechanism, he is willing to accept $k_A = -k_B > 0$. Haggling (misrepresenting information) is not a serious problem: because all of agent $B$'s types will suffer from hold-up if negotiations break down, there is no type that has to be given a large expected payoff to participate, hence agent $B$ does not have much to gain from misrepresenting his type. Surplus is destroyed if stage 3b is reached because actions are not efficiently coordinated, and the extra surplus generated by the first-best can be sufficient to satisfy both agents' individual rationality constraints.

Combining Theorems 1 and 4, we have the following result:

**Theorem 5** Rent-Seeking Case. *Suppose $\rho > \gamma$. Then, with nonintegration no first-best individually rational mechanism exists; with integration such a mechanism exists if and only if*

$$\int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds \geq F(\gamma - \rho)(\gamma + \rho).$$

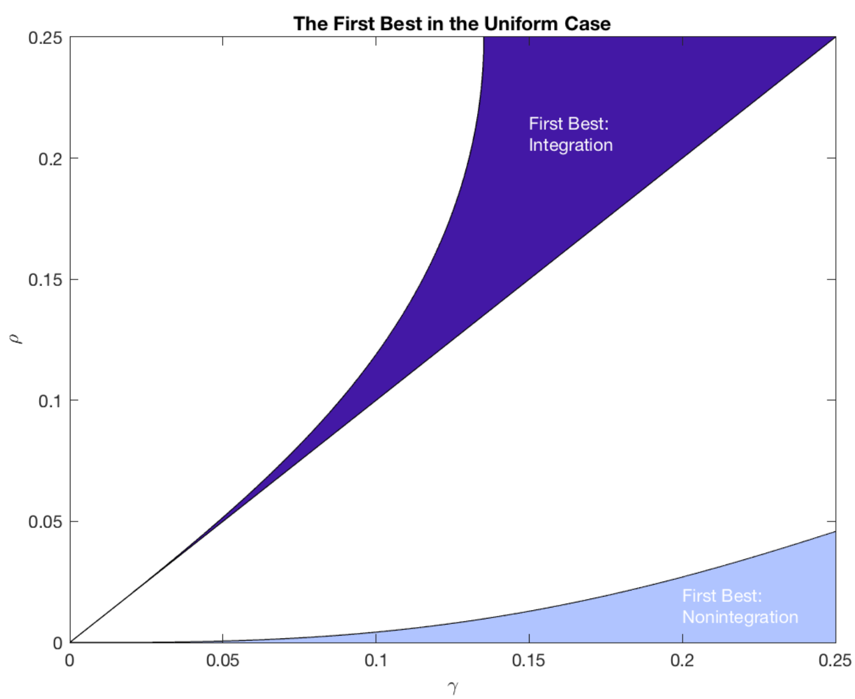
# 6 Summary of the Results

Our results show how the attainable surplus depends on the allocation of control rights. With nonintegration, each agent has a fairly good bargaining position, because he has some decision rights if negotiations fail. But the reservation payoff, which determines his bargaining strength, is type-dependent. If $\rho$ is large compared to $\gamma$ (so the rent-seeking motive is strong), then some types could gain a lot from rent-seeking if negotiations fail. Their strong bargaining position spills over to other types, who can always *pretend* to have a lot to gain from rent-seeking. The threat of such preference falsification, or "haggling", by nonintegrated agents may make the first-best unattainable. Specifically, if the rent-seeking motive is strong, then for a first-best mechanism $\Gamma$ to both attract the types who benefit the most from rent-seeking *and* prevent other types from haggling would require more surplus than is available.

Williamson [30] suggested that integration might eliminate costly haggling. But both integration and nonintegration require information transmission, and where the threat of haggling is more serious depends on the parameters. Under integration, agent $A$ has a high reservation payoff, since he has all the decision rights if negotiations break down. Agent $B$'s reservation payoff depends on agent $A$'s relative incentives to cooperate or rent-seek. In the rent-seeking case (where $\rho$ is large compared to $\gamma$), agent $B$'s reservation payoff is low because hold-up is a credible threat. Accordingly, agent $B$ is willing to make a large monetary transfer to agent $A$ in order to get the first-best decision. This large transfer persuades agent $A$ to give up his decision rights, so the first-best is implementable with integration. Intuitively, in the rent-seeking case, if negotiations break down the subordinate's losses will exceed the boss's gains, and this makes it possible to agree on the first-best. In the cooperative case, however, agent $B$ knows that he will never be held up. This means that agent $B$'s type $\theta_B = 0$, who has no idiosyncratic preference over which decision to make, has a very good bargaining position. His other types can pretend to be type $\theta_B = 0$, giving them significant bargaining power as well. Accordingly, agent $B$ will not "bribe" agent $A$ sufficiently to give up the control rights. Agent $A$ prefers to exercise his control rights unilaterally. Therefore, if $\rho$ is small compared to $\gamma$, the first-best cannot be implemented with integration.

Figure 6 summarizes the results for both integration and nonintegration. Again the calculations assume $F$ is the uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. The purple or darkly shaded area is the set of $(\gamma, \rho)$ such that with integration a first-best individually rational mechanism exists. The light blue or lightly shaded area is the set of $(\gamma, \rho)$ such that with nonintegration a first-best individually rational mechanism exists (that is, in this area we have $\rho < \rho^*(\gamma)$).

Figure 6




So far, we have focused on parameter values where either integration or nonintegration implements the first-best. For the remaining parameter values (i.e., the unshaded part of Figure 6), the qualitative principle is the same: to minimize the information rents required for incentive-compatibility, the optimal allocation of control rights minimizes the total value of outside options for the critical types (that are most likely to walk away from negotiations). Integration yields a strictly higher surplus than nonintegration within the purple (or darkly shaded) area in Figure 6. It is clear that for parameters sufficiently close to this area, integration will still yield a higher surplus than nonintegration (although decisions will be distorted since the first-best is not implemented). Conversely, nonintegration yields a strictly higher surplus than integration for parameters sufficiently close to the light blue (or lightly shaded) area. Somewhere in the unshaded part of Figure 6, between the darkly and lightly shaded areas, lies the boundary that separates the region where integration dominates nonintegration from the region where the converse is true. Recall, however, the discussion in Section 2.3: in the unshaded part of Figure 6, the first-best can be implemented by giving the mediator control rights over both assets. Thus, this ownership structure dominates both integration and nonintegration in the unshaded part of Figure 6. Since this is hard to interpret, we do not consider it further.

# 7 Concluding Comments

Milgrom and Roberts [17] contrasted the bargaining costs of nonintegration with the "influence costs" of integration (the latter being the costs incurred when a subordinate expends time and effort on rent-seeking). In contrast, we treat nonintegration and integration symmetrically, in each case deriving a cost of haggling from the IC and IR constraints.[12] This symmetry also distinguishes us from Tadelis [27]. He argued that if noncontractible changes must be made to a product ex post, then the buyer will enjoy a higher surplus under integration because he can more easily get the changes he wants; the drawback of integration is instead that the seller has less incentive to reduce costs. Each model provides a different perspective on the problem of efficient conflict resolution under incomplete contracts.

From GHM we take the idea that asset ownership confers residual rights of control which determine outside options. But we focus on Williamsonian

[12] Powell [21] considers a (symmetric) model of influence costs. In his model the two managers play a signal-jamming game where the level of manipulation of signals depends on allocation of decision rights.

costs of haggling over production decisions, rather than on the incentives to invest ex ante.[13] Control rights act as a "hostage" (Williamson [32]): an agent who gives up ownership of an asset becomes more eager to reach an agreement (his participation constraints are relaxed), which reduces the cost of haggling (lower information rents). In contrast, GHM emphasize that giving up ownership of an asset makes an agent less willing to make relationship-specific investments.[14] Thus, the models provide different perspectives on property rights.

Like GHM, we assume bargaining is efficient. However, efficient bargaining does not necessarily imply first-best production decisions because, as Williamson pointed out, private information may lead to "selective or distortive information disclosure" (Williamson [30], p. 26). In an interesting discussion of different ownership structures in the biotechnology industry,[15] Pisano [20] stressed the role of information asymmetries: "The firm in charge of R&D will accumulate asymmetric information on the technology; likewise, the partner in charge of marketing will gain asymmetric information on the technology's commercial potential. Strategic misrepresentation of new information by either party is a possibility."

Several other papers have studied the role of property rights in efficient bargaining. In Cramton, Gibbons and Klemperer's [4] model, equal division of an asset (a partnership) implements the first-best. This paper showed that ownership matters, although their result that concentrated ownership is

[13] Schmitz [24] allows a seller to invest ex ante to learn his disagreement payoff as well as to increase the value of the object being traded. The buyer has no investment opportunity so the usual GHM intuition implies the seller should be allowed to own his assets to minimize hold-up problems. But when there is asymmetric information as the seller learns his outside option, the canonical result is reversed. It can be optimal for the buyer to own the assets to reduce the seller's incentives to collect information. Again, we have no ex ante investments so our model is quite different.

[14] Hart and Holmström [11] also study ex post conflict, but they study behavioral phenomena that might destroy surplus. Our approach is not behavioral, but has instead private information at its center; control rights determine outside options which in turn determine whether agreement is possible. Coordination problems are also studied by Alonso, Dessein and Matouschek [1] and Rantakari [22], but these are models of centralized versus decentralized decision making when agents use cheap-talk, and hence quite different from our model.

[15] One firm may own a controlling share in a separate jointly controlled venture or become a large but minority shareholder in its trading partner; or two firms may have fifty-fifty ownership and control specific parts of the joint venture, say with one controlling decisions related to R&D and the other decisions related to marketing.

never optimal makes their model less suitable for studying the choice of integration versus nonintegration. Segal and Whinston [25] clarify and greatly extend Cramton, Gibbons and Klemperer's [4] insights to allow more general assumptions about assets and decisions (see also Segal and Whinston [26]). But Segal and Whinston [25] consider the case where decisions are contractible ex ante. As we discussed in Section 2.3, asset ownership becomes irrelevant in this case. In contrast, we study the importance of asset ownership when decisions are completely non-contractible ex ante.

Baker, Gibbons and Murphy [2] argue that decision rights within an organization are not contractible: the boss cannot formally delegate decision making to an employee, because the boss always has the right to overturn the subordinate's decision. More generally, they argue that formal authority can only be allocated via asset ownership, although informal authority can be allocated in a repeated game. In equilibrium the agent who owns the asset must be better off than he would be by making a unilateral decision, since the latter option cannot be contracted away. This is similar to our participation constraint. However, unlike Baker, Gibbons and Murphy [2], we assume the boss and the subordinate negotiate efficiently at the interim stage, which leads to our mechanism design approach.

The usual justification for the incomplete contracts assumption is that the possible states and actions are impossible to fully describe *ex ante*. At the interim stage, the set of possible states is known to lie in a much smaller set, and the set of feasible actions can be verified. In a well-known critique, Maskin and Tirole [15] proposed complex ex ante contracts that mimic complete contracts. However, it is not clear that a court would enforce such long-run contracts if the parties, unable to describe states and actions, could not have forecast scenarios for breach or the circumstances under which taking a certain action would be prohibitively costly. In such situations, the *penalty doctrine* in contract law specifies that courts should not (excessively) penalize an agent who walks away from the contract (see Eisenberg [5]). An interim agreement, made with better knowledge of the current conditions, would be more likely to be enforced by courts. In this case, asset ownership would play the same role as in our model.

## References


[1] Ricardo Alonso, Wouter Dessein and Niko Matouschek (2008): "When Does Coordination Require Centralization?," *American Economic Review* 98 (1), 145-179.

[2] George Baker, Robert Gibbons and Kevin Murphy (1999) "Informal Authority in Organizations" *Journal of Law, Economics and Organization* 15 (1): 56-73.

[3] Claude d'Aspremont and L. A. Gérard-Varet (1979): "Incentives and Incomplete Information," *Journal of Public Economics*, 11, 25-45.

[4] Peter Cramton, Robert Gibbons, and Paul Klemperer (1987): "Dissolving a Partnership Efficiently," *Econometrica*, 55, 615–632.

[5] Melvin Aron Eisenberg (1995): "The Limits of Cognition and the Limits of Contract," *Stanford Law Review*, 47, 211-259.

[6] Robert Gibbons (2005): "Four Formal(izable) Theories of the Firm," *Journal of Economic Behavior and Organization*, 58, 202-47.

[7] Jay Greene (2018): "Microsoft is buying Github for $7.5 Billion in Stock," *Wall Street Journal*, June 4, 2018.

[8] Jay Greene and Laura Stevens (2018): " 'You're Stupid If You Don't Get Scared': When Amazon goes from Partner to Rival," *Wall Street Journal*, June 1, 2018.

[9] Sanford J. Grossman and Oliver D. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691–719.

[10] Maija Halonen (2002): "Reputation and The Allocation of Authority," *Economic Journal*, 112, 539–558.

[11] Oliver Hart and Bengt Holmström (2010): "A Theory of Firm Scope", *Quarterly Journal of Economics*, May, 125: 483-513.

[12] Oliver D. Hart and John Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119–1158.

[13] Benjamin Klein, Robert Crawford, Armen Alchian (1978): "Vertical Integration, Appropriable Rents and the Competitive Contracting Process," *Journal of Law and Economics*, 21, 297-326.

[14] Jean-Jacques Laffont and Eric Maskin (1979): "A differentiable approach to expected utility-maximizing mechanisms", in J.-J. Laffont (ed.), *Aggregation and Revelation of Preferences*. North-Holland, Amsterdam.

[15] Eric Maskin and Jean Tirole (1999), "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, 66:83-114.

[16] Niko Matouschek (2004): "Ex Post Inefficiencies in a Property Rights Theory of the Firm," *Journal of Law, Economics, & Organization*, 20 (1), 125-147.

[17] Paul Milgrom and John Roberts (1990): "Bargaining costs, influence costs, and the organization of economic activity," in J. E. Alt and K. A. Shepsle (eds.) *Perspectives on positive political economy*, Cambridge University press, 57-89.

[18] Roger B. Myerson and Mark A. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281.

[19] Kevin O'Brien (2011): "Together, Nokia and Microsoft Renew a Push in Smartphones," *New York Times,* Feb 12.

[20] Gary Pisano (1989): "Using Equity Participation to Support Exchange: Evidence from the Biotechnology Industry," *Journal of Law, Economics and Organization*, 5, 109-126.

[21] Michael Powell (2015): "An Influence-Cost Model of Organizational Practices and Firm Boundaries," *Journal of Law Economics and Organization*, 31.

[22] Heikki Rantakari (2008): "Governing Adaptation," *Review of Economic Studies*, 75(4), 1257-1285.

[23] Ariel Rubinstein (1985): "Choices of Conjectures in a Bargaining Game with Incomplete Information," in A. Roth (ed.), *Game Theoretic Models of Bargaining*, 99-114, Cambridge University Press.

[24] Patrick W. Schmitz (2006): "Information Gathering, Transaction Costs, and the Property Rights Approach," *American Economic Review*, Vol. 96, 422-434.

[25] Ilya Segal and Michael Whinston (2011): "A Simple Status Quo that Ensures Participation (with Application to Efficient Bargaining)," *Theoretical Economics*, 6(1), 109-125.

[26] Ilya Segal and Michael Whinston (2011): "Property Rights," *Handbook of Organizational Economics*, ed. by Robert Gibbons and John Roberts, Princeton University Press, 2012

[27] Steven Tadelis (2002): "Complexity, Flexibility and the Make or Buy Decision," *American Economic Review: Papers and Proceedings*, 92(2), 433-437.

[28] Tom Warren (2013): "Nokia kept secrets from Microsoft over its Windows Phone plans, " *CNET*, September 7.

[29] Oliver E. Williamson (1971): "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 61, 112-123.

[30] Oliver E. Williamson (1975): *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York, N.Y.

[31] Oliver E. Williamson (1979) "Transaction-Cost Economics: The Governance of Contractual Relations" *Journal of Law and Economics* Vol. 22, No. 2, 233-261

[32] Oliver E. Williamson (1983): "Credible Commitments: Using Hostages To Support Exchange," *American Economic Review*, 73(4), 519-540.

[33] Oliver E. Williamson (1985): *The Economic Institutions of Capitalism*, Free Press, New York, N.Y.

[34] Oliver E. Williamson (2002): "The Theory of the Firm as Governance Structure: From Choice to Contract", *Journal of Economic Perspectives* Vol. 16, No.3, 171-195

[35] Nick Wingfield (2013): "Microsoft Gets Nokia Units, and Leader," *New York Times*, Sept 3.

# 8 Appendix

## 8.1 Missing Results and Proofs from Section 4

**Lemma 1** *Let $\Gamma$ be a first-best mechanism. With nonintegration, agent $i$'s interim IR constraints are satisfied for all his types if the interim IR constraint is satisfied for any type $\theta_i \geq 2\gamma$, which is true if and only if*

$$\underline{u}_i \geq \gamma + (\rho - \gamma)F(\theta^*). \tag{22}$$

**Proof of Lemma 1.** By symmetry, it suffices to consider agent $A$. The IR constraints require that agent $A$'s payoff under $\Gamma$, which is either (7), (8) or (9) depending on $\theta_A$, exceeds the reservation payoff, which is either (16) or (17), again depending on $\theta_A$. Thus, we consider the possible cases that can occur.

If $\theta_A \leq \min\{\theta^*, -2\gamma\}$, then the IR constraint is that (7) should be no less than (16), that is,

$$u_A(\theta_A) = \underline{u}_A \geq -\rho + (\gamma + \rho)F(\theta^*).$$

If $\min\{\theta^*, -2\gamma\} \leq \theta_A \leq \max\{\theta^*, -2\gamma\}$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \geq -\rho + (\gamma + \rho)F(\theta^*).$$

If $\max\{\theta^*, -2\gamma\} \leq \theta_A \leq 2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \geq \theta_A + \gamma + (\rho - \gamma)F(\theta^*).$$

If $\theta_A \geq 2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \theta_A \geq \theta_A + \gamma + (\rho - \gamma)F(\theta^*). \tag{23}$$

Notice that

$$\gamma + (\rho - \gamma)F(\theta^*) > -\rho + (\gamma + \rho)F(\theta^*)$$

using (15) and the fact that $\theta^* < 0$. Therefore, the individual rationality constraint for $\theta_A \geq 2\gamma$, which is $\underline{u}_A \geq \gamma + (\rho - \gamma)F(\theta^*)$, implies all the others. ■

**Proof of Proposition 1.** The necessity of (18) follows from (12) and Lemma 1.

To prove sufficiency of (18), consider the AGV mechanism with transfers given by (13). Since the mechanism is incentive compatible and the outcome is first-best, the results of Section 3 apply. If $k_A = k_B = 0$ then the mechanism is symmetric, so $\underline{u}_A = \underline{u}_B$. From (12) we obtain

$$\underline{u}_A = \frac{1}{2} \int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds. \tag{24}$$

Inequality (18) says that (24) exceeds $\gamma + (\rho - \gamma) F(\theta^*)$. Therefore, by Lemma 22, all IR constraints are satisfied. ■

The following result helps us express Proposition 1 in terms of $\gamma$ and $\rho$.

**Lemma 2** *For any $\gamma$, there is $\rho^*(\gamma) \in (0, \gamma)$ such that (18) holds if and only if $\rho \leq \rho^*(\gamma)$.*

**Proof of Lemma 2.** The left side of (18) can be written as

$$\begin{aligned} \int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds &= \int_{-2\gamma}^{0} (1 - F(-s))^2 \, ds + \int_{0}^{2\gamma} (F(s))^2 \, ds \\ &= \int_{0}^{2\gamma} \left[ (F(s))^2 + (1 - F(s))^2 \right] ds. \end{aligned}$$

Since $1/2 < (F(s))^2 + (1 - F(s))^2 < 1$, this expression lies strictly between $\gamma$ and $2\gamma$. Therefore, if $\rho \geq \gamma$, it is impossible to satisfy (18). If $\rho$ is close to 0, then $\theta^*$ is close to 0 and the right hand side of (18) is close to $\gamma$. If $\rho$ is close to $\gamma$, then the right hand side of (18) is close to $2\gamma$.

We now claim that the right hand side of (18) strictly increases from $\gamma$ to $2\gamma$ as $\rho$ increases from 0 to $\gamma$. As the left hand side lies strictly between $\gamma$ and $2\gamma$ and is independent of $\rho$, this claim will prove the lemma.

From (15), we get

$$\frac{d\theta^*}{d\rho} = \frac{1}{2F'(\theta^*)\gamma - 1} < 0.$$

The derivative of the right side of (18) with respect to $\rho$ is

$$2F(\theta^*) + 2(\rho - \gamma) F'(\theta^*) \frac{d\theta^*}{d\rho} = 2F(\theta^*) + 2(\rho - \gamma) F'(\theta^*) \frac{1}{2F'(\theta^*)\gamma - 1}.$$

We claim this is positive. This is equivalent to showing

$$F(\theta^*) > \frac{(\rho - \gamma) F'(\theta^*)}{1 - 2F'(\theta^*)\gamma}$$

which is the same as

$$F(\theta^*) > (\rho - \gamma + 2\gamma F(\theta^*)) F'(\theta^*) = (2\rho + \theta^*) F'(\theta^*) \tag{25}$$

where the equality uses (15). Because $2F'(\theta^*)\gamma < 1$,

$$(2\rho + \theta^*) F'(\theta^*) < (2\rho + \theta^*) / 2\gamma.$$

Therefore, to show (25) holds we need to show that $2F(\theta^*)\gamma > 2\rho + \theta^*$. This is true because of (15) and $\rho < \gamma$. ■

Proposition 1 and Lemma 2 together prove Theorem 1.

## 8.2 Missing Results and Proofs from Section 5.1

**Lemma 3** *Let $\Gamma$ be a first-best mechanism and suppose $\rho < \gamma$. With integration the following is true. (a) Agent A's interim IR constraints are satisfied for all his types if the interim IR constraint is satisfied for any type $\theta_A \geq 2\gamma$, which is true if and only if*

$$\underline{u}_A \geq \gamma. \tag{26}$$

*(b) Agent B's interim IR constraints are satisfied for all his types if the interim IR constraint is satisfied for type $\theta_B = 0$, which is true if and only if*

$$\underline{u}_B \geq \gamma - \int_{-2\gamma}^{0} F(s)ds. \tag{27}$$

**Proof of Lemma 3.** The proof is analogous to the proof of Lemma 1. The IR constraints require that the payoff under the mechanism, which is either (7), (8) or (9) depending on the agent's type, exceeds the reservation payoff, which is (19) for agent $A$ and (20) for agent $B$. Thus, we consider the possible cases that can occur. First, we consider agent $A$.

If $\theta_A \leq -2\gamma$ then the IR constraint is that (7) should be no less than $\gamma$, that is, $u_A(\theta_A) = \underline{u}_A \geq \gamma$. If $-2\gamma \leq \theta_A \leq 0$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \geq \gamma.$$

If $0 \leq \theta_A < 2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \geq \gamma + \theta_A.$$

If $\theta_A \geq 2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \theta_A \geq \gamma + \theta_A.$$

It is easy to see that all these IR constraints are satisfied if the IR constraint holds for $\theta_A \geq 2\gamma$, which is (26).

For agent $B$, the individual rationality constraints are as follows. If $\theta_B \leq -2\gamma$, then the IR constraint is

$$u_B(\theta_B) = \underline{u}_B \geq \gamma + \frac{1}{2}\theta_B.$$

If $-2\gamma \leq \theta_B \leq 2\gamma$, then the IR constraint is

$$u_B(\theta_B) = \underline{u}_B + \int_{-2\gamma}^{\theta_B} F(s)ds \geq \gamma + \frac{1}{2}\theta_B.$$

If $\theta_B \geq 2\gamma$, then the IR constraint is

$$u_B(\theta_B) = \underline{u}_B + \theta_B \geq \gamma + \frac{1}{2}\theta_B.$$

It is easy to check that all of agent $B$'s IR constraints are satisfied if the IR constraint holds for type $\theta_B = 0$, which is (27). ■

Then, from Lemma 3 and (12), if the first-best is implementable under integration then

$$\underline{u}_A + \underline{u}_B = \int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds \geq 2\gamma - \int_{-2\gamma}^{0} F(s)ds. \tag{28}$$

But this never holds, because the middle term and the third term are equal at $\gamma = 0$ and the derivative of the middle term with respect to $\gamma$ is always strictly less than the derivative of the third term. Therefore, the inequality in (28) is violated for all $\gamma > 0$. This proves Theorem 2.

## 8.3 Missing Results and Proofs from Section 5.2

**Lemma 4** *Let* $\Gamma$ *be a first-best mechanism and suppose* $\rho > \gamma$. *With integration the following is true. (a) Agent A's interim IR constraints are satisfied for all his types if the interim IR constraint is satisfied for any type* $\theta_A \geq 2\gamma$, *which is true if and only if*

$$\underline{u}_A \geq \rho. \tag{29}$$

*(b) Agent B's interim IR constraints are satisfied for all his types if the interim IR constraint is satisfied for any* $\theta_B \leq -2\gamma$, *which is true if and only if*

$$\underline{u}_B \geq -\rho + F(\gamma - \rho)(\gamma + \rho). \tag{30}$$

**Proof of Lemma 4.** The proof is analogous to the proofs of Lemmas 1 and 3. Suppose first that $\gamma - \rho < -2\gamma$. Then, the individual rationality constraints for agent $A$ are as follows.

If $\theta_A \leq \gamma - \rho$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A \geq \gamma. \tag{31}$$

If $\gamma - \rho \leq \theta_A \leq -2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A \geq \theta_A + \rho. \tag{32}$$

If $-2\gamma \leq \theta_A \leq 2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \geq \theta_A + \rho. \tag{33}$$

If $\theta_A \geq 2\gamma$, then the IR constraint is

$$u_A(\theta_A) = \underline{u}_A + \theta_A \geq \theta_A + \rho.$$

If instead $\gamma - \rho > -2\gamma$ then (31) applies when $\theta_A \leq -2\gamma$, (33) applies when $\gamma - \rho \leq \theta_A \leq 2\gamma$, and (32) is replaced by the following IR constraint for the case $-2\gamma \leq \theta_A \leq \gamma - \rho$:

$$u_A(\theta_A) = \underline{u}_A + \int_{-2\gamma}^{\theta_A} F(s)ds \geq \gamma.$$

In either case, it can be checked that all of these IR constraints hold if and only if the IR constraint holds for $\theta_A \geq 2\gamma$, which is (29).

For agent $B$, the individual rationality constraints are as follows. If $\theta_B \leq -2\gamma$ then the IR constraint is

$$\underline{u}_B \geq -\rho + F(\gamma - \rho)(\gamma + \rho).$$

If $-2\gamma \leq \theta_B \leq 2\gamma$, then the IR constraint is

$$u_B(\theta_B) = \underline{u}_B + \int_{-2\gamma}^{\theta_B} F(s)ds \geq -\rho + F(\gamma - \rho)(\gamma + \rho).$$

If $\theta_B \geq 2\gamma$, then the IR constraint is

$$u_B(\theta_B) = \underline{u}_B + \theta_B \geq -\rho + F(\gamma - \rho)(\gamma + \rho).$$

It can be checked that all of agent $B$'s IR constraints hold if and only if the IR constraint holds for $\theta_B \leq -2\gamma$, which is (30). ∎

**Proof of Theorem 4.** The necessity of (21) follows from Lemma 4 and (12). To prove sufficiency of (21), consider the AGV mechanism with transfers given by (13). Since the mechanism is incentive compatible and the outcome is first-best, the results of Section 3 apply. Choose $k_A$ such that $\underline{u}_A = \rho$. By Lemma 4, all of agent $A$'s IR constraints are satisfied. From (12) and (21),

$$\underline{u}_B = \int_{-2\gamma}^{2\gamma} (F(s))^2 \, ds - \rho \geq F(\gamma - \rho)(\gamma + \rho) - \rho.$$

Again by Lemma 4, all of agent $B$'s IR constraints are satisfied as well. ∎