Contracting with Third Parties

By Sandeep Baliga and Tomas Sjöström*

In bilateral holdup and moral hazard in teams models, introducing a third party allows implementation of the first best, even if renegotiation is possible. Fines paid to the third party provide incentives for truth-telling and investment. This result holds even if the third party is corruptible, as long as the grand coalition has access to the same contracting technology as any colluding subcoalition. (*JEL* D86, D82)

Models of bilateral contracting, such as the canonical hold-up model, typically assume that third parties are not available. Given this assumption, the equilibrium outcome is not first-best efficient if contracts can be renegotiated. To be more specific, suppose an architect and a builder must cooperate to build a building. The quality of the building will depend on three things: the quality of the architect's design, the builder's skill, and a stochastic shock. We will refer to these three variables as the *state of the world*. The architect and the builder know the true state but no one else does. Thus, the state is "observable but not verifiable." (An outsider may be able to judge the quality of the building after it is built, but he cannot disentangle the various contributions to it.) Suppose the contract specifies ex post transfers as a function of announcements made by the architect and the builder (a "message game"). If both report the state truthfully, the transfers will reflect the contributions of each party and provide correct incentives to invest in the transaction. In order to support an equilibrium where both tell the truth, however, it may be necessary to punish both of them if they disagree. In the absence of a third party, this punishment typically involves an ex post inefficiency (say, the destruction of the building). If such outcomes can be renegotiated, then it may be impossible to implement the first best. This hold-up problem underlies recent work on the "foundations of incomplete contracts" (see Yeon-Koo Che and Donald B. Hausch 1999 and Ilya Segal 1999).

If we introduce a third party (who may not know the true state), ex post renegotiation becomes less of a problem. The punishment can now consist of fines paid to the third party who acts as a "budget breaker" (as suggested by Bengt R. Holmström 1982). Since a fine is simply a transfer from one person to another, renegotiation is not an issue. This suggests that models that rely on renegotiation to generate hold-up

* Baliga: Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 (e-mail: baliga@kellogg.northwestern.edu); Sjöström: Department of Economics, Rutgers University, New Brunswick, NJ 08901 (e-mail: tsjostrom@economics.rutgers.edu). We thank two anonymous referees of the *American Economic Review*, Joel Watson and seminar audiences at Haas School of Business, University of California, Berkeley, New York University, Ohio State University and University of Chicago Graduate School of Business for helpful comments.

problems may not be robust to the introduction of third parties. In the literature, the "third-party solution" has been dismissed by arguing that it requires the third party to be completely honest and incorruptible (see Oliver D. Hart and John H. Moore 1988 and Hart 1995, 79–80). A corrupt third party might side contract with the builder to extract a fine from the architect. However, in this paper we show that in two important contracting models—the buyer-seller model and the moral hazard in teams model—it is possible to design the original contract so that no side contracting occurs in equilibrium and the first-best outcome is implemented.

The original contract regulates the relationship within the grand coalition (consisting of the two original agents plus the third party). A subcoalition can collude by signing a side contract. In the previous literature, two polar cases can be distinguished. Much of classical contract theory and mechanism design considered the polar case where side contracting is impossible. More recently, the opposite polar case of "perfect side contracting" has been considered. For example, Hart and Moore (1988, footnote 20) put no restrictions on the ability to side contract. The colluding parties can in effect "merge" and thus collude perfectly. But the two original agents, the buyer and the seller, cannot solve their hold-up problem by an unrestricted merger.

In this article, we consider a case that is intermediate between the two polar cases of no side contracting and perfect side contracting. Our starting point is that all coalitions should have access to the same contracting technology. This will be referred to as the symmetry assumption.[1] Symmetry requires that side contracting is subject to the same constraints as the original contract. A recent literature develops this approach, including Jean-Jacques Laffont and David Martimort (1997, 2000), Baliga and Sjöström (1998), Dilip Mookherjee and Masatoshi Tsumagari (2004), Gorkem Celik (2007), and Gregory Pavlov (2006). All of these papers build on the seminal work of Jean Tirole (1986).

The original contract is supervised by an original judge who collects messages from the agents and orders both "real" actions and transfers of money. The "judge" is impartial and incorruptible and may not be an actual person but a computer program that collects inputs and selects an output. If it is a person, it can be a private arbitrator rather than a public official. Similarly, a side contract will be supervised by a supplementary judge who is similar to the original judge (and so is incorruptible and may not be a physical person).[2] It is important to specify precisely what a judge can observe, since this defines the verifiable information for the contract he supervises. Following Laffont and Martimort (1997, 2000), we assume any contract may use a message game to elicit unverifiable information from the contracting parties.

[1] In reality, side contracts may be illegal and therefore harder to enforce than the original contract. We will show that the grand coalition can implement the first best in the symmetric case. A fortiori, the first best can also be implemented if side contracts are harder to enforce, since the harder it is to side contract, the easier it is to implement the first best.

[2] Symmetry requires that all judges be identical copies of each other. We do not want to assume, for example, that a subcoalition can use an honest judge, while the grand coalition can use only a corrupt judge. A reasonable starting point is that all judges are incorruptible. If anything, intuition suggests that finding an honest judge might be more difficult for a colluding subcoalition than for the grand coalition.

All judges are limited by the informational constraints imposed by the physical and contractual environment. Real actions, such as a trade of commodities, are assumed to be observed publicly. Therefore, if the original contract implements a real action, then this real action is verifiable by any supplementary judge. This facilitates side contracting (and makes it harder to implement the first best). There are two possible assumptions about collusive messages and side transfers: they might be public and hence verifiable by the original judge, or secret and unverifiable by the original judge. The first case is trivial, because it is easy for the original judge to punish collusion if he can verify what the colluding parties are doing. In effect, this case is equivalent to the polar case of no side contracting studied in classical mechanism design. To make the problem nontrivial, side contracting must take place behind closed doors, unobserved by the original judge. Since the grand coalition should be able to use the same contracting technology as any subcoalition, the grand coalition also can use secret messages and side payments. The supplementary judge (or arbitrator) cannot use cameras or microphones to monitor the original proceedings. (Alternatively, except for the real action, the inputs and outputs to one computer program cannot be direct inputs into the other.) The important implication is that messages and cash transfers exchanged under the original contract are not verifiable by a supplementary judge and hence not contractible for a subcoalition.[3]

Even though a supplementary judge cannot directly monitor the original judge's proceedings, there are two ways for information to leak from the original contract to the side contract. First, the supplementary judge may try to infer the original messages or cash transfers by observing which real action the original judge ordered. Second, although the original messages and cash transfers are not verifiable by the supplementary judge, they are observed by the colluding parties (who, after all, participate in the original proceedings). The supplementary judge may use a message game to extract information about things that are observed by the colluding parties but not by the judge.

The first type of information leakage can be prevented by the original contract. It suffices to make sure the real action specified by the original mechanism does not reveal messages or transfers. Our paper will show that the second type of information leakage is not a problem either. To see this intuitively, suppose the builder and the third party conspire to make the architect pay a fine to the third party, which the third party is then supposed to share with the builder. The side contract has to provide incentives for the builder to send a message in the original mechanism which triggers the fine, and for the third party to share the fine with the builder. By assumption, the messages and transfers exchanged under the original contract are observed by the builder and the third party, but they are not verifiable to the supplementary judge. So, suppose the side contract specifies a message game, where the builder and the third party agree to tell the supplementary judge what happened when the

[3] Hart and Moore (1988) argue that it might be impossible to outlaw collusion even if the side contracting is publicly observed, because the side contract might be so complicated that an outsider cannot understand its true (collusive) meaning. In this case, a symmetric treatment of all coalitions requires that the grand coalition also has access to this kind of contracting technology. It must therefore be possible for the original judge's court hearings to be incomprehensible to outsiders. This is equivalent to the case we are considering, where all contracting (original or collusive) takes place behind closed doors, unobserved by outsiders.

original mechanism was played out. The supplementary judge then decides whether or not the third party should transfer a sum of money to the builder. However, the third party can always claim that the builder never sent the message, which would have triggered the fine, so no fine was paid. The supplementary judge has no direct knowledge of whether or not this is true because the original court proceedings took place behind closed doors. Moreover, since past messages and cash transfers are not "payoff relevant" (assuming quasi-linear utility functions), there is no way for the supplementary judge to extract the truth of the matter. The fact that the original messages and cash transfers are "observable but not verifiable" puts severe constraints on the side contract, and indeed makes it impossible to enforce an agreement whereby the builder and the third party share any fine paid by the architect to the third party.

In the terminology of the implementation literature, "full implementation" requires that there are no undesirable equilibrium outcomes. If, on the other hand, there are both desirable and undesirable equilibrium outcomes, then the mechanism "weakly implements" the desired outcome. When the original contract is designed, the main difficulty is to ensure the existence of an equilibrium which produces a first-best, efficient outcome. Consequently, we begin by studying weak implementation. Once weak implementation is accomplished, full implementation is easy to accomplish as well. Indeed, there are many ways to destroy undesirable equilibria. For example, suppose we want to rule out an undesirable equilibrium where the builder and the third party collude against the architect. In such an equilibrium, the architect knows that he is being conspired against and that his payoff is likely to be low. (Each player has "equilibrium knowledge," i.e., knows which equilibrium is played.) But it is easy to include a message ("blowing a whistle") that the architect prefers to send if (and only if) he knows that the others conspired against him, thereby destroying the bad equilibrium.

We model side contracting in a fully noncooperative way, and we do not allow "coalitional deviations" or "mergers" outside the formal moves of the game. The agents propose and accept side contracts as part of an extensive form game, and we investigate the set of perfect Bayesian equilibria (PBE) of this game. In order to achieve weak implementation of the first-best outcome, we design an original mechanism which has an optimal PBE. No side contracting occurs along the equilibrium path. To support this PBE, we must ensure that, for any coalition and any side contract, there exists a continuation equilibrium[4] that would be bad for at least one member of the coalition. Suppose the builder and the third party can write a side contract that induces multiple continuation equilibria, one of which is bad for the builder. In the optimal PBE, if the third party were to propose this side contract, the builder would reject it, anticipating the outcome that is bad for him. Of course, the third party can propose any side contract, including complex contracts with integer games, etc. If he proposes a side contract that induces a unique continuation equilibrium, which is good for him and the builder, then the builder must accept it, and

[4] Since side contracts are secret, outsiders do not know that a coalition has formed. By a continuation equilibrium, we mean the actions the agents plan to take, conditional on what they know.

the optimal PBE would collapse. More generally, a side contract can threaten the optimal PBE if and only if all continuation equilibrium outcomes are good for all members of the coalition. We show that no such side contract exists if the original contract is well designed.

Formally, we will show that any side contract induces a continuation equilibrium where the messages sent to the supplementary judge about the original court proceedings are uninformative. This continuation equilibrium supports the optimal PBE. A collusive agreement could threaten the original contract only if it could be "fully implemented" in the sense that every continuation equilibrium is good for both colluding parties, but this is impossible. We stress that we are not arbitrarily introducing some kind of contracting asymmetry by requiring weak implementation for the grand coalition but full implementation for subcoalitions. It is an implication of the definition of PBE that a collusion-free PBE exists if every side contract induces at least one continuation equilibrium, which is bad for some colluding party. Also, although for convenience we begin by studying weak implementation for the grand coalition, we will show that it is easy to adjust the original mechanism to support full implementation.

In our buyer-seller model, the first best can be implemented by an original mechanism that always recommends the same real action (i.e., the same trade), regardless of the messages. It is impossible for a subcoalition to implement fully a collusive agreement of the form: "if the third party receives a fine, he will share it with his coalition partner." For any side contract, there will exist a continuation equilibrium where the third party makes the same transfer to his coalition partner regardless of what happened in the original mechanism. But this only adds or subtracts a constant from the payoffs and does not affect marginal incentives within the original mechanism. It does not threaten the optimal PBE. Since the first best is implemented, there is no inefficiency caused by holdup.

In the moral hazard in teams model, the total output is a "real" amount of goods produced. This is verifiable for any judge. Unlike Holmström (1982), we assume side contracting is possible.[5] The third party supplies no effort but acts as a budget breaker. A side contract can require the third party to make a transfer to a colluding team member as a function of the (verifiable) output of the team.[6] Such a side contract does affect the team member's incentives in the original mechanism and may potentially upset an optimal PBE. Indeed, we show that if the original contract does not specify a message game, then the budget breaker is redundant. This captures Eswaran and Kotwal's (1984) and Hart and Moore's (1988) insight—the budget breaker and one of the original agents will collude, so it is useless to include a budget breaker in the original contract. But this insight only applies to original

[5] Mukesh Eswaran and Ashok Y. Kotwal (1984) introduce side contracting into Holmström's team model. Sandro Brusco (1997) looks at side contracting in a model where the team members can observe each others' effort levels.

[6] Since individual effort is unobserved, unlike in the buyer-seller model, there is nothing special about the number two. Implementation of the first best is difficult even when the team has three or more members. As Holmström (1982) showed, big teams can also benefit from a third party (i.e., an outside budget breaker who is not a team member). To maintain symmetry with the buyer-seller model, we assume the team has only two members, but our results apply for bigger teams as well.

contracts without message games. We will show that if a message game is included in the original contract, then a third party is valuable. We construct an original mechanism with a "whistle-blowing" clause, where the third party reveals any collusive activity he knows about. The original contract will not reveal whether the third party blew the whistle or not, i.e., whistle-blowing will not be verifiable information for a supplementary judge. We show that for any collusive coalition which includes the third party, there is always a continuation equilibrium where the third party blows the whistle in front of the original judge. The original judge then punishes the other coalition member, who therefore does not want to collude in the first place.

Even though the team members cannot observe each other's effort, a message game is necessary for weak implementation of the first best. This seems to contradict the "revelation principle"— if effort is unobserved, what is there to send messages about? In the presence of side contracting, however, the revelation principle cannot be interpreted to mean that the agents should reveal what they know about effort levels. They should also reveal what they know about collusion (by blowing the whistle). Leonardo Felli (1996) makes a similar point with respect to Tirole's (1986) principal-supervisor-agent model.

Our buyer-seller model is similar to Segal and Michael D. Whinston (2002), and encompasses models that have used Eric S. Maskin and Moore's (1999) implementation with renegotiation paradigm to provide "foundations for incomplete contracts." For example, in Che and Hausch (1999), traders make cooperative investments and decide what quantity to trade. The first best cannot always be achieved, but the second best can be implemented without any explicit contract. In Segal (1999), traders make selfish investments and n possible goods can be traded (see also Maskin and Tirole 1999 and Hart and Moore 1999). Under some assumptions, as n becomes large, the first best cannot be achieved, and the second best can be implemented without any explicit contract. These results are important theoretical foundations for incomplete contracts. We show that implementation of the first best can be achieved by including a third party, even if agents can side contract and renegotiate inefficient outcomes.

In some models of bilateral holdup with renegotiation, the first best can be implemented even without a third party. Che and József Sákovics (2004) consider a dynamic model where the timing of investment is endogenous. They show that the hold-up problem can be resolved, even with no contract, if the buyer and seller are sufficiently patient. Philippe Aghion, Mathias Dewatripont, and Patrick Rey (1994); Georg Nöldeke and Klaus M. Schmidt (1995); and Aaron S. Edlin and Stefan Reichelstein (1996) also found positive results for the bilateral case. Joel Watson (2007) has shown that the extent of the hold-up problem depends on the technological details of renegotiation. He argues that if the decision to take a verifiable action lies in the hands of individuals rather than an outside enforcer, it may be possible to implement a larger set of state-contingent payoffs. All these results complement ours by clarifying the extent to which the hold-up problem can be resolved even in the absence of a third party.

We now turn to the results. Section I contains the buyer-seller model with a third party. Section II studies the moral hazard in teams model. Section III concludes.

I. The Buyer-Seller Model

A. *The Buyer-Seller Relationship*

There is a buyer B and a seller S. Let $b \geq 0$ denote a relationship-specific investment made by the buyer, and let $s \geq 0$ denote a relationship-specific investment made by the seller. Let $\omega \in \Omega$ denote a random variable which is realized after investments are made. The buyer's realized cost of making investment b is $\varphi_B(b,\omega)$, and the seller's realized cost of making investment s is $\varphi_S(s,\omega)$. The vector $\boldsymbol{\theta} = (b,s,\omega)$ is the *state of the world*. We make the standard assumption that $\boldsymbol{\theta}$ is observed by B and S but by no one else.

Trade between the buyer and seller is represented by a set of possible real actions denoted by X. Specifically, a real action, $x \in X$, may specify what kind of good (and how much of it) the seller delivers to the buyer. The buyer's gross value from the trade is denoted by $v(x,b,s,\omega)$. The seller's cost from the trade is $c(x,b,s,\omega)$. This formulation is quite general. For example, it allows the possibility of cooperative investments (where one agent's investment directly influences the other's payoff). There is a "null outcome," $x_\emptyset \in X$, which we interpret as "no trade." In addition to the "real" actions in X, monetary transfers can be made. Utility functions are quasi-linear in money. Thus, for example, if the buyer receives a monetary transfer, t_B, then his final payoff is $v(x,b,s,\omega) - \varphi_B(b,\omega) + t_B$. The ex post surplus is $v(x,\boldsymbol{\theta}) - c(x,\boldsymbol{\theta})$. Let $x^*(\boldsymbol{\theta}) \in X$ be the real action (assumed unique), which maximizes the ex post surplus in state $\boldsymbol{\theta}$. That is,

$$x^*(\boldsymbol{\theta}) \equiv \arg\max_{x \in X} \{v(x,\boldsymbol{\theta}) - c(x,\boldsymbol{\theta})\}.$$

Define

$$v^*(\boldsymbol{\theta}) \equiv v(x^*(\boldsymbol{\theta}),\boldsymbol{\theta})$$

and

$$c^*(\boldsymbol{\theta}) \equiv c(x^*(\boldsymbol{\theta}),\boldsymbol{\theta}).$$

The maximized ex post surplus is

$$\Sigma^*(\boldsymbol{\theta}) \equiv \max_{x \in X} \{v(x,\boldsymbol{\theta}) - c(x,\boldsymbol{\theta})\} = v^*(\boldsymbol{\theta}) - c^*(\boldsymbol{\theta}).$$

The first-best investment levels (b^*,s^*) maximize the expected value of $\Sigma^*(b,s,\omega) - \varphi_B(b,\omega) - \varphi_S(s,\omega)$, where the expectation is with respect to the random variable ω. That is,

$$(b^*,s^*) \equiv \arg\max_{(b,s) \geq 0} E_\omega\{\Sigma^*(b,s,\omega) - \varphi_B(b,\omega) - \varphi_S(s,\omega)\}.$$

We assume, for simplicity, that there is a unique first-best pair (b^*, s^*). The first-best solution to the contracting problem is for the buyer and seller to make investments (b^*, s^*), and for every $\omega \in \Omega$ to take the real decision $x^*(b^*, s^*, \omega)$. If the first-best solution is implemented with monetary transfers t_B and t_S, then the buyer's expected payoff is

$$B(t_B) \equiv E_\omega\{v^*(b^*, s^*, \omega) - \varphi_B(b^*, \omega)\} + t_B.$$

The seller's expected payoff is

$$S(t_S) \equiv E_\omega\{-c^*(b^*, s^*, \omega) - \varphi_S(s^*, \omega)\} + t_S.$$

B. *The Third Party*

A third party T may be invited to play the role of budget breaker in the buyer-seller relationship. Thus, there are three players: B, S, and T. The third party cares only about money (not about x or $\boldsymbol{\theta}$), and his payoff is linear in wealth. He does not observe $\boldsymbol{\theta}$.

C. *Time Line*

The relationship between B, S, and T is governed by an original contract that specifies an original mechanism Γ_0. A mechanism is a message game, which specifies message spaces, and an outcome function, which maps messages into outcomes.[7] We do not model the bargaining process, which produces Γ_0. We simply assume that at the beginning of the game that Γ_0 has already been determined. The extensive form game $G(\Gamma_0)$, induced by Γ_0, is described by the following time line.

At time 0, there is a coalition-formation game with two stages. In the first stage, T can propose a side contract to B or S (but not to both).[8] A proposal to player $j \in \{B, S\}$ is an invitation to form a two-player coalition, $C = \{j, T\}$. The proposal specifies a set of "side-contracting mechanisms," $\{\Gamma_C(x)\}_{x \in X}$.[9] As the notation suggests, we allow the side-contracting mechanism to depend directly on x, the action implemented by the original mechanism, because x is verifiable. At time 0, the coalition does not know what real action will be implemented by Γ_0, so they write a contingent side contract of the form "if Γ_0 produces the outcome x, then the coalition will play message game $\Gamma_C(x)$ at time 4."

[7] We restrict attention to normal form mechanisms. Allowing the agents to send messages sequentially would not change any results.

[8] It is evident that side contracting between B and S will not be a problem. Therefore, allowing B and S to make proposals would not change anything.

[9] The only restriction we put on the set of possible side contracts is that they be regular enough that a continuation equilibrium exists. If the strategy space in a side-contracting mechanism is an open set, for example, no continuation equilibrium may exist. Presumably, the supplementary judge would not tolerate it. Formally, we assume T can only propose side contracts with the best-response property. Each player will aways have a best response to any strategy chosen by his opponents.

If T does not make any proposal in the first stage, then we bypass the second stage and proceed to time 1. If T makes a proposal to player $j \in \{B,S\}$ in the first stage, then in the second stage player j responds to T by either accepting or rejecting the proposal. If player j accepts, then coalition $\{j,T\}$ has formed, and the side contract is in force. If j rejects (or no proposal was made in stage one), then no coalition is formed.

Side contracting is done secretly. Thus, if T proposes a coalition $\{B,T\}$, then S is not informed about this (neither is S informed about B's decision to accept or reject the proposal). Similarly, B is never told about any side contract between T and S. Consequently, if a two-person coalition forms, then the party who is left out is not informed about this and cannot react to it in any way.

At time 1, the buyer and the seller make investments $b \geq 0$ and $s \geq 0$, respectively. The buyer observes the seller's investment s, and the seller observes the buyer's investment b, but no one else observes b or s.

At time 2, the random variable $\omega \in \Omega$ is realized. The realization of ω is observed by the buyer and the seller but not by anyone else.

At time 3, the original mechanism Γ_0 is played out among the original parties. The mechanism specifies a message space M_i for each player $i \in \{B,S,T\}$. For each message profile $m \in M_B \times M_S \times M_T$, the mechanism produces an outcome of the form $(x(m),t(m))$. Here, $x(m) \in X$ is a real action ordered by the original judge, and $t(m) = (t_B(m),t_S(m),t_T(m)) \in \mathbb{R}^3$ is a vector of monetary payments, where $t_i(m)$ is a transfer to player i. We require $t_B(m) + t_S(m) + t_T(m) = 0$ for all m. Thus, the budget must always balance. The original judge cannot destroy wealth, only reallocate it among the three parties.

At time 4, nothing happens if no coalition was formed at time 0. However, if a coalition C was formed, then the side-contracting mechanism $\Gamma_C(x)$ is played out among the coalition (where $x = x(m)$ is the real decision determined by the messages m at time 3). The side-contracting mechanism $\Gamma_C(x)$ specifies a message space $M_i^C(x)$ for each player $i \in C$. For each message profile $m^C \in \times_{i\in C} M_i^C(x)$, $\Gamma_C(x)$ produces an outcome $(t_i^C(m^C))_{i\in C} \in \mathbb{R}^{|C|}$. Here, $t_i^C = t_i^C(m^C)$ is a monetary payment to agent $i \in C$. We require $\sum_{i\in C} t_i^C \leq 0$. If there is strict inequality, then the supplementary judge in effect destroys wealth (see the discussion below). The side-contracting mechanism $\Gamma_C(x)$ cannot specify a different real action than the original mechanism. Any attempt to overrule the original contract by choosing a different x would constitute a violation of the original contract, which we assume can be ruled out by the original judge. Indeed, since the physical transaction x directly involves B and S (but not T), a coalition that excludes either B or S clearly cannot have any right to choose x.[10] However, the side-contracting mechanism can specify cash transfers among the coalition because these are not publicly observable.

[10] If the real action ordered by the supplementary judge is publicly observable, then by definition is it verifiable and can be ruled out by the original contract. On the other hand, to assume a supplementary judge could "secretly order" a real action would have absurd consequences. To be specific, suppose the judge supervising a secret, collusive agreement between B and T orders a "secret trade" between B and S. That means S will receive an order to secretly deliver some goods to B signed by a judge he never heard of. We assume that, since S is not a part of the collusive agreement—indeed it is directed against him—he has no obligation to obey this order. That

At time 5, B and S may renegotiate the decision $x = x(m)$ produced by Γ_0 at time 3. The renegotiation takes place in secret and cannot be observed by anyone except B and S.[11] With probability λ_B, the buyer makes a take-it-or leave-it proposal consisting of a new decision, $x^R \in X$, and a pair of transfers (t_B^R, t_S^R) (which are added to any previous transfers the players have received). We require $t_B^R + t_S^R = 0$. If the seller accepts, the proposal is implemented. If the seller rejects, there is no renegotiation and the game ends. With probability $\lambda_S = 1 - \lambda_B$, it is the seller who makes a take-it-or leave-it proposal. If the buyer accepts, the proposal is implemented. If the buyer rejects, there is no renegotiation and the game ends. Our results are not sensitive to the precise specification of the renegotiation game. Any reasonable specification (e.g., alternating-offer bargaining) would lead to the same results.

D. *The Renegotiated Outcome*

The game $G(\Gamma_0)$ is solved backwards. Suppose time 5 has been reached and consider the continuation equilibrium of the renegotiation stage. The true state of the world $\boldsymbol{\theta} = (b, s, \omega)$ is known to B and S. The mechanism Γ_0 has recommended the real decision $x = x(m)$ at time 3. Player i receives a transfer $t_i = t_i(m)$ at time 3. If a coalition C formed at time 0, then player $i \in C$ also receives a transfer t_i^C at stage 4. Notationally, if $i \notin C$ then set $t_i^C \equiv 0$. Let $\hat{t}_i$ denote the sum of the transfers,

$$\hat{t}_i = t_i + t_i^C, \tag{1}$$

for $i \in \{B, S, T\}$.

In the continuation equilibrium, the renegotiated outcome x^R will maximize ex post surplus, i.e., $x^R = x^*(\boldsymbol{\theta})$. Whichever party makes the take-it-or-leave-it offer will appropriate all the surplus. If B makes the proposal, he will make sure that S is indifferent between accepting and rejecting the proposal. In order to convince S to switch from x to $x^*(\boldsymbol{\theta})$, S will be compensated by the amount $t_S^R = c^*(\boldsymbol{\theta}) - c(x, \boldsymbol{\theta})$. Conversely, if S makes the proposal, then B will be compensated by the amount $t_B^R = v(x, \boldsymbol{\theta}) - v^*(\boldsymbol{\theta})$.

Given state $\boldsymbol{\theta}$, real decision x as recommended by Γ_0, and the pair of transfers $(\hat{t}_B, \hat{t}_S)$, we can now calculate the buyer's expected payoff, taking renegotiation into account, but not including the cost of the investment. It is

$$\begin{aligned} u_B(x, \hat{t}_B, \boldsymbol{\theta}) &= v^*(\boldsymbol{\theta}) + \hat{t}_B - \lambda_B[c^*(\boldsymbol{\theta}) - c(x, \boldsymbol{\theta})] + \lambda_S[v(x, \boldsymbol{\theta}) - v^*(\boldsymbol{\theta})] \\ &= \lambda_B \Sigma^*(\boldsymbol{\theta}) + \hat{t}_B + \lambda_B c(x, \boldsymbol{\theta}) + \lambda_S v(x, \boldsymbol{\theta}). \end{aligned} \tag{2}$$

is, a judge has no jurisdiction over agents other than those who signed the agreement he supervises. Otherwise, B and T could sign an absurd agreement that would force S to hand over everything he owns.

[11] If renegotiation is public, then the original contract can forbid it and use inefficient punishments to support a truthful equilibrium. Then, a third party would not be necessary to achieve first best.

Similarly, the seller's expected payoff is

$$u_S(x,\hat{t}_S,\boldsymbol{\theta}) = \lambda_S \Sigma^*(\boldsymbol{\theta}) + \hat{t}_S - \lambda_B c(x,\boldsymbol{\theta}) - \lambda_S v(x,\boldsymbol{\theta}). \tag{3}$$

Since we know what will happen at time 5, we will suppress the renegotiation stage in what follows. Thus, if there is no side contract in force and the message-game form Γ_0 produces the outcome $(x(m),t_B(m),t_S(m),t_T(m))$ at time 4, then the buyer's final payoff will be $u_B(x(m),t_B(m),\boldsymbol{\theta})$, as defined by (2). The seller's final payoff will be $u_S(x(m),t_S(m),\boldsymbol{\theta})$, as defined by (3). The third party's payoff will be $t_T(m)$. If there is a side contract in force, then the transfers $(t_i^C)_{i\in C}$ are added to the payoffs in the obvious way, according to (1).

E. *Participation Constraints*

The buyer and seller may have the option of not trading. To formalize this, let Γ_0^* denote a "null" mechanism that simply recommends the outcome $x_\emptyset$ and no transfers (there are no messages). Of course, the outcome $x_\emptyset$ will be renegotiated at time 5. The payoffs in state (b,s,ω) will be

$$u_B(x_\emptyset,0,(b,s,\omega)) - \varphi_B(b,\omega) \tag{4}$$

for the buyer and

$$u_S(x_\emptyset,0,(b,s,\omega)) - \varphi_S(s,\omega) \tag{5}$$

for the seller, using the definitions in (2) and (3). Under the null contract, the buyer will set b to maximize the expectation of (4) with respect to ω, taking s as given. The seller will set s to maximize the expectation of (5) with respect to ω, taking b as given. Let $\hat{b}$ and $\hat{s}$ denote the equilibrium investments. In general, these investments will not be at the efficient level, $(\hat{b},\hat{s}) \neq (b^*,s^*)$. The expected payoffs from the equilibrium induced by the null mechanism Γ_0^* are

$$B_\emptyset \equiv E_\omega\{u_B(x_\emptyset,0,(\hat{b},\hat{s},\omega)) - \varphi_B(\hat{b},\omega)\} \tag{6}$$

and

$$S_\emptyset \equiv E_\omega\{u_S(x_\emptyset,0,(\hat{b},\hat{s},\omega)) - \varphi_S(\hat{s},\omega)\}. \tag{7}$$

Of course, with the null contract, the third party plays no role and gets zero payoff.

The participation constraints will ensure that both B and S are better off than under the null contract. It is useful to define a pair of transfers (t_B^*,t_S^*), where $t_B^* = -t_S^*$, such that

$$B^*(t_B^*) \equiv E_\omega\{u_B(x_\emptyset,t_B^*,(b^*,s^*,\omega)) - \varphi_B(b^*,\omega)\} \geq B_\emptyset \tag{8}$$

and

$$S^*(t_S^*) \equiv E_\omega\{u_S(x_\emptyset, t_S^*, (b^*, s^*, \omega)) - \varphi_S(s^*, \omega)\} \geq S_\emptyset. \tag{9}$$

Suppose the buyer and seller make the first-best investments b^* and s^*, respectively, and in every state $x_\emptyset$ is implemented and transfers (t_B^*, t_S^*) are made. Renegotiation will take no trade to the first-best decision $x^*(\boldsymbol{\theta})$ in every state, and (8) and (9) guarantee that the buyer's and the seller's participation constraints are satisfied. That is, they are better off than they would be under the null contract.

F. *Discussion*

As mentioned above, a reasonable starting point for the analysis of side contracting is the symmetry assumption that the grand coalition and all subcoalitions have access to the same contracting technology. In our model, however, we have made two assumptions that give two-player subcoalitions a slight contracting advantage: (a) subcoalitions can destroy wealth, but the grand coalition cannot; (b) there is only one round of side contracting, so a subcoalition does not have to worry about making the side contract robust against further deviations. We will show that the first best can be implemented, even though two-player subcoalitions have this contracting advantage over the grand coalition. Of course, the first-best can also be implemented if this advantage is removed. Thus, the first best can be implemented if the symmetry assumption holds. Indeed, by giving subcoalitions a slight contracting edge, we strengthen our result that the first best can be implemented.

The assumption that subcoalitions can destroy wealth implies that a side-contracting mechanism can have a "truthful" equilibrium, where the colluding parties reveal what happened during the original proceedings, in addition to many untruthful equilibria. Truth-telling is enforced by the threat that the supplementary judge will punish both colluding parties (by destroying wealth) if they disagree with each other. As is well known, such truthful equilibria typically exist when both parties have the same (unverifiable) information and ex post inefficient punishments are feasible. If we had ruled out ex post inefficient punishments in side contracts, then the truthful equilibrium would not exist, and the difficulties of side contracting would have been revealed even more starkly. The model would have become less tractable, however, because if a two-agent coalition could not destroy wealth then it might benefit from an outside budget breaker (for the same reason that a third party is used in the original contract). Enlisting such a "fourth party" would open up further possibilities for side contracting, and the second assumption (that there is only one round of subcontracting) would have to be dropped. The analysis would become quite complicated. Our assumption that subcoalitions can use ex post inefficient punishments means they have no need for a budget breaker. But in the original contract, we do not allow any kind of ex post inefficiency such as the destruction of wealth. Despite this handicap, the original contract will implement the first best.

We now further discuss our assumptions about observability and verifiability.

First, we make the standard assumption that the state is observed by the buyer and the seller, but unobservable (and hence unverifiable) to outsiders, including the third party and all judges.

Second, we assume coalitions can form secretly. Thus, the original contract cannot simply outlaw coalition formation. Maskin and Tirole (1999) suggest that the original contract might reward any agent who produces evidence of a side contract (and the original contract will punish the other members of the coalition). We assume hard evidence about side contracts is impossible to produce, otherwise it would be too easy to implement the first best (by simply outlawing side contracting). Any member of a coalition will have "soft" information about the side contract, however, and can be asked to reveal it ("whistle-blowing"). Thus, at time 3 a "revelation mechanism" should not just collect reports about the state of the world, but also about side contracts. (It turns out that whistle-blowing is useful in the team model, but not in the buyer-seller model).

Third, what goes on behind the closed doors of a judge (messages and cash payments) cannot be observed by outsiders. This assumption is needed to make side contracting possible. Indeed, if the original judge could observe the supplementary judge's court proceedings, then side contracts would be verifiable information, and so the original judge could outlaw them. Since the original judge should be able to use the same contracting technology as any supplementary judge, the original mechanism can also be played out behind closed doors. In other words, since the original judge does not have any microphone installed in the supplementary judge's court room, then the supplementary judge cannot have a microphone installed in the original judge's court room.[12] In general, the outcome of a contract cannot depend directly on messages or cash transfers exchanged under another contract (but perhaps indirectly, via message games).

Fourth, we assume that the real decision x produced by Γ_0 at time 3 is publicly observable, hence verifiable by any judge. The real decision x has a physical manifestation outside the court room, so unlike messages and monetary transfers, it is impossible to keep x secret. For example, the original judge may order the seller to deliver a certain quantity of goods to the buyer. This physical action cannot take place in secret. This helps the agents side contract, because any coalition can make its agreement conditional on x.

Fifth, we make the standard assumption that renegotiation at time 5 is unverifiable. If renegotiation is verifiable, then the original mechanism can prescribe that large payments be made by B and S to T, should the final decision x' differ from the decision x prescribed by the original contract. Maskin and Tirole (1999) suggest that the original contract might reward any agent who produces evidence of renegotiation (and the contract will punish the others who participated in the renegotiation).[13] With such a scheme, even if renegotiation occurs, some of the surplus generated

[12] If agent i gives cash to the judge, who transfers it to agent j, neither agent will have any proof that the transaction took place. The judge does not give out any receipts (he is incorruptible so no receipts are necessary). A bank statement showing agent i has withdrawn cash from his account does not reveal what happened to the money. Even if the mechanism is a computer program, it automatically can open a new bank account in agent j's name and deposit money in it. It would be impossible for agent j to prove he did not receive money in this way. Notice that this scheme makes it possible to secretly reward whistle-blowers.

[13] A difficult situation occurs if a new, renegotiated contract surfaces that contradicts the original contract. Suppose the new contract is signed by B, T, and S and contains a clause invalidating Γ_0. It is not clear which contract would in fact be enforced. If the new contract has precedence, then renegotiation cannot be eliminated by Γ_0 in the way Maskin and Tirole suggest.

by it might go to T, so renegotiation might be costly for B and S. By adding more and more parties to the contract, the share of the surplus going to B and S might be lowered even further. We assume evidence of renegotiation is impossible to produce, however, so renegotiation is impossible to rule out in the original contract.[14] This assumption makes it more difficult to design Γ_0 to implement efficient outcomes.

Given our assumptions, a side contract for coalition C can specify transfers as a function of the decision x ordered by the original judge and the messages sent in the side-contracting mechanism. This might indirectly influence which decision x is implemented by the original contract. For example, a badly designed original contract might specify that if B says the quality of the good is low then the outcome is $x_\emptyset$ and S pays a fine to T, but otherwise they trade a positive amount. Now B and T can agree secretly that B should always report that the quality is low and split the fine with T. The side contract can enforce this by specifying that, if the outcome produced by the original mechanism is anything else than $x_\emptyset$, then B pays a large fine to T. Of course, a well designed original contract would not reveal the messages in this blatant way. But the colluding coalition might try to use a message game to elicit information about the messages sent in the original mechanism. That is, the supplementary judge could ask B and T about what B told the original judge.

G. *Implementation*

We will design an original mechanism Γ_0 and construct an equilibrium of $G(\Gamma_0)$, where no coalition forms along the equilibrium path and produces the first-best solution. In this section, we will not worry about the possible existence of other, nonoptimal, equilibria. Thus, this section deals with "weak" implementation of the first best.

In order to support an equilibrium with no side contract in force, T should not have any incentive to try to form a coalition at time 0. To ensure this, we need to show that for any coalition, there exists a continuation equilibrium that is bad for at least one of the members of the coalition. That is, for any possible side-contract proposal T could make to player $j \in \{B, S\}$, either the proposal makes player j worse off, so he will reject it, or T is made worse off, so he does not want to make the proposal in the first place.

We now define the original mechanism Γ_0 that is played out at time 3. This particular mechanism will be called the secret message mechanism. In the secret message mechanism, the buyer and the seller announce the state simultaneously. Formally, player $i \in \{B, S\}$ sends a message $\boldsymbol{\theta}^i$ from the message space $M_i \equiv \Theta$. The third party sends no message. To ensure that the participation constraints are satisfied, the pair of transfers (t_B^*, t_S^*) satisfy (8) and (9), with $t_B^* = -t_S^*$. The outcome function is defined as follows.

[14] Another reason to allow renegotiation is that B and S may want to trade in the future. Ruling out future transactions might be inefficient if not impossible (B and S may use intermediaries to trade).

Rule 1. If $\boldsymbol{\theta}^B = \boldsymbol{\theta}^S = \boldsymbol{\theta}$, then the real decision is $x(m) = x_\emptyset$, and transfers are determined as follows. If $\boldsymbol{\theta} = (b^*, s^*, \omega)$, then the buyer pays t_S^* to the seller. If $\boldsymbol{\theta} = (b, s^*, \omega)$ with $b \neq b^*$, then the buyer pays F^1 to the seller. If $\boldsymbol{\theta} = (b^*, s, \omega)$ with $s \neq s^*$, then the seller pays F^1 to the buyer. If $\boldsymbol{\theta} = (b, s, \omega)$ with $b \neq b^*$ and $s \neq s^*$, then no transfers are made.

Rule 2. If $\boldsymbol{\theta}^B \neq \boldsymbol{\theta}^S$, then $x(m) = x_\emptyset$. The buyer and seller each pay F^2 to the third party.

Although the messages are observed by B, S, and T, they are not verifiable by outsiders. The no-trade outcome is always implemented to avoid signaling the message profile indirectly. We will show that a side contract cannot elicit information about the original messages, so there is no way to collude profitably. The equilibrium is first best, because the no-trade outcome is renegotiated to the efficient decision in every state, and transfers are designed to give efficient incentives.

THEOREM 1: *We can choose F^1 and F^2, so the game $G(\Gamma_0)$ has a perfect Bayesian equilibrium which produces the first-best outcome. Transfers (t_B^*, t_S^*) are implemented by the mechanism in every state, so the participation constraints are satisfied.*

To prove the theorem, we construct a PBE of $G(\Gamma_0)$ as follows. At time 0, no side contract is proposed. If the players have not joined any coalition at time 0, then they play as follows from time 1 on. The buyer and seller invest at the first-best level at time 1, and at time 3 they tell the truth (in all states of the world). If, at time 3, either the buyer or the seller deviates and lies about the state, then they incur the fine F^2 by Rule 2. If F^2 is large enough, neither the buyer nor the seller has an incentive to deviate from truth-telling. Furthermore, if F^1 is sufficiently big, then Rule 1 implies that both agents prefer to choose the first-best investment levels, anticipating that the truth is revealed at time 3. From this it follows that, if there is no side contract at time 0, then the proposed strategies are sequentially rational from time 1 on. The outcome is first best by construction, and the equilibrium payoffs are $B^*(t_B^*)$ for the buyer, $S^*(t_S^*)$ for the seller, and 0 for the third party.

It remains to specify behavior after a time 0 deviation, where T proposes a side contract. To support the equilibrium such a deviation should not be profitable. To be specific, suppose T proposes a side contract to B (the argument concerning a proposal to S will be exactly the same). We construct the strategies so that if B accepts and coalition $C = \{B, T\}$ forms, then either B or T will get no more than their equilibrium payoff. If T gets no more than zero, it is certainly not profitable for him to propose the coalition. If B gets no more than $B^*(t_B^*)$ by joining the coalition, then we stipulate that his equilibrium strategy is to reject the proposal, and this behavior is certainly sequentially rational. Knowing that B will reject, it is again not profitable for T to make the proposal.

So suppose B accepts. The coalition $C = \{B, T\}$ forms and a side contract $\{\Gamma_C(x)\}_{x \in X}$ is in force. S is unaware of the side contract and will play as described

above. The equilibrium strategies must specify how B and T will behave in $G(\Gamma_0)$ after they have formed a coalition. Consider a pair of messages $(m_B^C, m_T^C) \in M_B^C(x_\emptyset) \times M_T^C(x_\emptyset)$ in the side contract such that

$$t_B^C(m_B^C, m_T^C) \geq t_B^C(m_B, m_T^C)$$

for all $m_B \in M_B^C(x_\emptyset)$, and

$$t_T^C(m_B^C, m_T^C) \geq t_T^C(m_B^C, m_T)$$

for all $m_T \in M_T^C(x_\emptyset)$. That is, (m_B^C, m_T^C) would be a Nash equilibrium of a game where only the transfers in the side contract matter. Some such pair must exist (we can allow mixed strategies), because T is not allowed to propose a badly behaved side contract that causes an existence problem. Now, we let the equilibrium strategies for $G(\Gamma_0)$ specify that when C has formed, B and T choose this particular pair (m_B^C, m_T^C) regardless of what else has happened in the game before time 4. This can be done because nothing that happens before time 4 can change the strategic incentives in $\Gamma_C(x_\emptyset)$, which are always just to maximize one's side payment. So, (m_B^C, m_T^C) will be part of a continuation equilibrium, following any history. If (m_B^C, m_T^C) are always sent, then B's investment and the message he sends at time 3 will not affect his side payment, so B maximizes his payoff by making the first-best investment and telling the truth at time 3, just as if no side contract had been signed. This is then, what B's equilibrium strategy tells him to do. Rule 1 will apply, and S will get a payoff $S^*(t_S^*)$. But then, either B gets no more than $B^*(t_B^*)$ or T gets no more than zero. As argued above, this implies that T does not gain by making the proposal. This completes the proof of Theorem 1.

The buyer and the third party would jointly benefit if they could enforce the following side contract: "The third party pays the buyer $2F_2 - \varepsilon$ at time 4 if and only if the buyer contradicted the seller at time 3." But this is not an enforceable side contract, because messages in Γ_0 are not verifiable by the supplementary judge, and the real action is always $x_\emptyset$. Moreover, a message game in the side contract cannot elicit information about messages sent in Γ_0. With quasi-linear utilities, previous transfers are payoff irrelevant, so there is always an uninformative continuation equilibrium where the time 4 messages are independent of what happened at time 3. But then, B may as well tell the truth at time 3, to avoid paying an extra fine to T.

We offer three further comments on Theorem 1.

First, one cannot eliminate the "uninformative" continuation equilibrium by appealing to some refinement of Nash equilibrium, such as undominated Nash equilibrium (Thomas R. Palfrey and Sanjay Srivastava 1991), or even by relying on virtual implementation (Dilip Abreu and Arunava Sen 1991 and Hitoshi Matsushima 1988). The reason is that a "preference reversal" condition would still be necessary. If the continuation equilibria are to differ across two histories of play, there must be

some agent whose preferences over two outcomes reverse (see Maskin and Sjöström 2003, Sections 4.1 and 4.2 for details). But, as we have stressed, the time 3 transfers are payoff irrelevant at time 4. Hence, there is no preference reversal. In our model, coalition formation is fully noncooperative, so the only way for a coalition to ensure a profitable deviation is to design a side contract which fully implements it. Theorem 1 shows that this is impossible.

Second, like most of the literature, we assume B and S are risk neutral. However, suppose that they have strictly concave von Neumann-Morgenstern utility functions. If the degree of risk aversion varies with wealth, and if a coalition can implement lotteries in the side contract played at time 4, then previous transfers become payoff relevant. At time 4, a lottery mechanism might extract indirect information about previous transfers, and enforce nontrivial side payments. However, if such lottery mechanisms can be part of a side contract, then they can also be part of the original contract, and the buyer and seller can implement the first best even without the help of a third party (see Maskin and Moore 1999). On the other hand, Hart and Moore (1999) argued that lottery mechanisms are impractical. In this case, the first best can be implemented with the help of a third party, by using the secret message mechanism described above, because a coalition cannot use a lottery mechanism in a side contract to extract information about what happened at time 3.

Third, we have assumed that the set of possible real actions X is describable ex ante. If parts of it are indescribable, it is impossible to write an original contract that fully identifies which action to implement. Indeed, the typical assumption in the incomplete contracts literature is that some actions, such as asset ownership, are always describable. Others, such as which object to trade, are indescribable ex ante but describable ex post (e.g., Sanford J. Grossman and Hart 1986, Hart and Moore 1990, and Hart 1995). The indescribability may be pertinent as it may be optimal to trade different objects in different states of the world. However, as Maskin and Tirole (1999) argued, there is a tension between the assumption that certain actions are indescribable ex ante and the assumption that agents are able to perform dynamic programming. Maskin and Tirole's (1999) Theorem 4 shows that a contract that is implementable (with renegotiation) when actions are describable is also implementable (with renegotiation) when they are indescribable. In this sense, indescribability is irrelevant and the only binding constraints are those imposed by renegotiation. Thus, even if actions are indescribable, Maskin and Tirole's (1999) irrelevance theorem, together with the results of our paper, implies that a third party contract can implement the first best.

H. *Full Implementation*

Although, for convenience, we began by studying weak implementation, there are various ways to make sure *all* PBE produce the first-best outcome (see Palfrey and Srivastava 1991 for a guide to the literature on full implementation). One simple way is to amend the time line in Section C by allowing B and S to send messages at the very beginning of the game, just before time 0. We call this time -1. The announcements at time -1 will decide whether the mechanism played at time 3 will be the secret message mechanism described in Section G or the null mechanism Γ_0^*

described in Section E. (Recall that Γ_0^* has no messages, and always recommends $x_\emptyset$ and no transfers.) The other parts of the time line, such as the secret side contracting at time 0, are unchanged.

Specifically, the augmented secret message mechanism works as follows. At time -1, B and S announce nonnegative integers simultaneously. These announcements are also secret and hence unverifiable. There are two cases.

Case 1. Suppose someone announces a strictly positive integer at time -1. Then there are no messages and no transfers at time 3, and the outcome produced in period 3 is $x_\emptyset$. That is, they play the null mechanism at time 3. However, the player who announces the highest integer receives a cash payment from his trading partner at time -1.[15] If the player with the highest integer is B, then S must pay B the amount

$$\tau_B = B^*(t_B^*) - B_\emptyset. \tag{10}$$

That is, the transfer equal to the difference between B's first-best payoff and the payoff from playing the null mechanism. Similarly, if the player with the highest integer is S, then B must pay S the amount

$$\tau_S = S^*(t_S^*) - S_\emptyset. \tag{11}$$

Thus, in case 1, the payoffs will be the same as under the null mechanism, with the time -1 transfers added on. Notice that in this case, side contracting is moot.

Case 2. Suppose both B and S say 0 at time -1. Then the game unfolds just as described in Section G. Thus, at time 3 the secret message mechanism described in Section G is operated. That is, at time 3, the buyer and the seller make simultaneous announcements of the state, and the outcome is determined according to rules 1 and 2 described in Section G. In this case, the side contracting possibilities at time 0 are nontrivial.

THEOREM 2: *The game induced by the augmented secret message mechanism fully implements the first-best outcome in perfect Bayesian equilibrium.*

We prove this theorem by proving two claims.

Claim 1. There exists a PBE of the game induced by the augmented secret message mechanism, where the outcome is first best.

PROOF OF CLAIM 1:

The equilibrium strategies specify that both B and S say 0 at time -1. After both have said 0, the equilibrium strategies are isomorphic to those described in Section

[15] Ties are broken arbitrarily, say in favor of B.

G. Thus, by the arguments in that section, there exists a continuation equilibrium that produces the first-best outcome.

If some player should say anything else than 0 at time -1, then they play a continuation equilibrium induced by the null mechanism. In this continuation equilibrium, B's expected payoff is $B_\emptyset$ plus whatever transfer is received at time -1, and S's expected payoff is $S_\emptyset$ plus whatever transfer is received at time -1. But at time -1, either S pays τ_B to B or B pays τ_S to S. If S pays τ_B to B, then B's expected payoff is $B_\emptyset + \tau_B = B^*(t_B^*)$, from (10). Thus, B gets exactly $B^*(t_B^*)$, which is what he would get if both had said 0 (and S gets less than $S^*(t_S^*)$ because the total surplus will be less than first best). Similarly, if B pays τ_S to S then neither agent is better off than he would be if both had said 0. It follows that neither B nor S has any incentive to deviate and say anything else than 0 at time -1. This proves claim 1.

Claim 2. All PBE produce the first-best outcome.

PROOF OF CLAIM 2:

The buyer and seller can guarantee themselves the payoffs $B^*(t_B^*)$ and $S^*(t_S^*)$, respectively, by announcing a high integer at time -1. Therefore, in any PBE, the payoffs must be at least this high. They cannot be strictly greater, since the third party never pays. Thus, in all PBE, the payoffs must be at the first-best level. This proves Claim 2.[16]

II. Moral Hazard in Teams

A team consists of two agents, B and S. At time 1, B and S choose effort levels $b \geq 0$ and $s \geq 0$, respectively. Neither agent observes the other agent's effort. However, the team's total output $x \in \mathbb{R}$ is publicly observable. Output is a deterministic function of effort, $x = x(b,s)$. For simplicity there is no stochastic shock (it could easily be added). Assume $x(b,s)$ is increasing, concave, and differentiable. The two main differences, compared to Section I, is the unobservability of b and s, and the fact that the verifiable outcome x is a function of b and s (in Section I, x was a real action implemented by the original judge).

Each agent is risk neutral. B's cost of effort is $\varphi_B(b)$, and S's cost of effort is $\varphi_S(s)$. Each φ_i is increasing, differentiable, and strictly convex, and $\varphi_i(0) = 0$. The first-best action profile is

$$(b^*, s^*) \equiv \arg\max_{(b,s)\geq 0} \{x(b,s) - \varphi_B(b) - \varphi_S(s)\}.$$

[16] Our result would go through if the integer game is played out between time 0 and time 1. Indeed, Claim 2 goes through because B and S can refuse to sign all side contracts and announce a high integer between time 0 and time 1, and guarantee themselves $B^*(t_B^*)$ and $S^*(t_S^*)$, respectively. To prove Claim 1, we construct an optimal PBE as follows. At time 0, no side contract is proposed. If the players have not joined any coalition at time 0, they both announce 0 in the integer game and then play as described in Section G. There is no incentive to trigger the integer game. If a side contract is offered, say to B, for it to be accepted it must give B more than $B(t_B^*)$ and, for it to be offered, it must give T more than 0. If the integer game is not triggered, this is impossible (see the proof of Theorem 1). But if the integer game is triggered, the coalition of B and T can generate at most a payoff of $B(t_B^*)$ and this cannot make them both strictly better off. Hence, Claim 1 also goes through.

The first-best solution specifies effort levels (b^*, s^*) and transfers t_B^* and t_S^*, such that $t_B^* + t_S^* = x(b^*, s^*)$. For the problem to be nontrivial, we assume $b^* > 0$ and $s^* > 0$. Individual rationality requires

$$B^* \equiv t_B^* - \varphi_B(b^*) \geq 0 \tag{12}$$

and

$$S^* \equiv t_S^* - \varphi_S(s^*) \geq 0. \tag{13}$$

If there is no message game, then, as in Holmström's (1982) pioneering article, the original contract simply specifies transfers as a function of output x. Since ex post inefficient outcomes are renegotiated, it suffices to consider contracts that satisfy budget balance, that is $t_B(x) + t_S(x) = x$ for all x. The budget-balance condition implies that it is impossible to implement the first-best without a third party (Holmström 1982). However, suppose there is a third party T who does not exert effort, does not observe any agent's effort, and whose transfer is $t_T(x)$. The budget-balance condition becomes $t_B(x) + t_S(x) + t_T(x) = x$. If there is no side contracting then the first-best can be implemented by the following contract. For $i \in \{B, S\}$, let

$$t_i(x) = \begin{cases} t_i^* \text{ if } x \geq x(b^*, s^*) \\ 0 \text{ if } x < x(b^*, s^*) \end{cases}. \tag{14}$$

The third party's transfer is $t_T(x) \equiv x - t_B(x) - t_S(x)$. This contract (weakly) implements (b^*, s^*) when side contracting is not possible (Holmström 1982). However, side contracting compromises this particular contract (Eswaran and Kotwal 1984). We now consider how side contracting impacts other kinds of contracts, including those that ask for messages.

The time line is similar to the one described in Section C. Thus, at time 0, there is a coalition-formation game where T can make a proposal to some player $i \in \{B, S\}$ to form coalition $C = \{i, T\}$. The proposal specifies a set of side-contracting mechanisms $\{\Gamma_C(x)\}_{x \in \mathbb{R}}$. At time 1, agents B and S choose effort levels b and s, and joint output $x = x(b, s)$ is realized. Agent i's effort is not observed by anyone except agent i, but the output x is observed publicly and verifiable by outsiders. At time 2, nothing happens (there is no stochastic shock).

At time 3, the original mechanism Γ_0 is played out among the original parties. The mechanism specifies a message space M_i for each player $i \in \{B, S, T\}$. For each message profile $m \in M_B \times M_S \times M_T$ and output $x \in \mathbb{R}$, the transfers are $(t_B(m, x), t_S(m, x), t_T(m, x))$. We require $t_B(m, x) + t_S(m, x) + t_T(m, x) = x$ for all x.

At time 4, nothing happens if no coalition was formed at time 0. However, if a coalition C was formed, then the side contract $\Gamma_C(x)$ is played out among the coalition (where x is the output realized at stage 1). The side contract $\Gamma_C(x)$ specifies a message space $M_i^C(x)$ for each player $i \in C$. For each message profile $m^C \in \times_{i \in C} M_i^C(x)$ and output x, $\Gamma_C(x)$ specifies transfers $(t_i^C(m^C, x))_{i \in C}$. Here, $t_i^C = t_i^C(m^C, x)$ is a monetary

payment that agent $i \in C$ receives when messages m^C are sent at time 4, and x is the team output. Finally, at time 5, there is no scope for renegotiation because the budget is balanced and there is no real decision to be made.

An effort profile (b, s) is (weakly) implementable if there is an original mechanism Γ_0 such that the induced game $G(\Gamma_0)$ has a PBE where the effort levels are (b, s). Since effort is unobserved, intuition suggests that the message game at time 3 is redundant. This intuition is incorrect, however. To see this, we first consider implementation without message games ($M_B = M_S = M_T = \emptyset$). We will show that in this case the third party plays no useful role. This generalizes Eswaran and Kotwal's (1984) result in a way reminiscent of footnote 20 in Hart and Moore (1988).

THEOREM 3: *If we restrict attention to original contracts without messages, then introducing a third party does not expand the set of implementable effort profiles.*

To prove the theorem, suppose there is a third party T, and the effort profile $(\hat{b}, \hat{s})$ is implemented by Γ_0 without messages. We show that $(\hat{b}, \hat{s})$ can also be implemented without a third party. There are two cases depending on whether or not a side contract is in force in equilibrium.

Case 1: The PBE of $G(\Gamma_0)$ which implements $(\hat{b}, \hat{s})$ is such that, in equilibrium, a side contract is in force. To be specific, suppose coalition $C = \{B, T\}$ forms with side contract $\{\Gamma_C(x)\}_{x \in \mathbb{R}}$.

Notice that S will maximize his payoff only if, for all $s' \geq 0$,

$$t_S(x(\hat{b}, \hat{s})) - \varphi_S(\hat{s}) \geq t_S(x(\hat{b}, s')) - \varphi_S(s'). \tag{15}$$

If B rejects T's proposal, his payoff is sure to be

$$\mu \equiv \max_{b \geq 0} \{t_B(x(b, \hat{s})) - \varphi_B(b)\}.$$

Indeed, S does not observe the side contract, hence his choice of $\hat{s}$ is independent of it. Since B will accept any proposal that gives him more than μ, in fact, B's equilibrium payoff must be exactly μ. The third party's equilibrium payoff must then be

$$t_B(x(\hat{b}, \hat{s})) + t_T(x(\hat{b}, \hat{s})) - \varphi_B(\hat{b}) - \mu. \tag{16}$$

Claim 1. For all $b' \geq 0$,

$$t_B(x(\hat{b}, \hat{s})) + t_T(x(\hat{b}, \hat{s})) - \varphi_B(\hat{b}) \geq t_B(x(b', \hat{s})) + t_T(x(b', \hat{s})) - \varphi_B(b'). \tag{17}$$

PROOF OF CLAIM 1:

Suppose there is $b' \geq 0$ such that (17) is violated. Suppose T deviates from the equilibrium by offering B the following side contract. If $x = x(b', \hat{s})$, then T pays B a side transfer $\hat{t}_B$ such that

$$t_B(x(b',\hat{s})) + \hat{t}_B - \varphi_B(b') = \mu + \varepsilon, \tag{18}$$

where $\varepsilon > 0$. If $x \neq x(b',\hat{s})$, then B pays a big fine to T. (There are no messages in the side contract.) Since B only gets μ by rejecting, (18) implies that the unique, sequentially rational response is to accept and choose b' so that the output is $x(b',\hat{s})$. Then T's payoff will be

$$t_T(x(b',\hat{s})) - \hat{t}_B = t_B(x(b',\hat{s})) + t_T(x(b',\hat{s})) - \varphi_B(b') - (\mu + \varepsilon), \tag{19}$$

where the equality uses (18). But, for small enough $\varepsilon > 0$, the violation of (17) implies that (19) is strictly greater than (16). Therefore, T is strictly better off by proposing the new side contract, contradicting the definition of PBE. This proves the claim.

Suppose we get rid of the third party, and any transfer that T would have received is added to B's transfer, so B's transfer is $t_B(x) + t_T(x)$ for any x. Now (15) and (17) imply that this new mechanism implements $(\hat{b},\hat{s})$, so, if case 1 applies, then the third party is useless.

Case 2: The PBE of $G(\Gamma_0)$ which implements $(\hat{b},\hat{s})$ is such that, in equilibrium, no side contract is in force.

In this case, the proof of Claim 1 goes through, so for any $b' \geq 0$, (17) must hold. But then, just as in case 1, we can get rid of the third party. This completes the proof of Theorem 3.

Theorem 3 implies that, if the original contract does not include a message game, then the first best is unattainable, even if a third party is available. We now show that an original mechanism with a message game can implement the first best, as long as a third party is available. This particular mechanism, Γ_0, will be called the whistle-blowing mechanism. In this mechanism, only T sends a message at time 3, with message space $M_T = \{\emptyset, \beta, \sigma\}$. Message $\emptyset$ is interpreted as "stay quiet," β is interpreted as "blow the whistle on agent B," and σ is interpreted as "blow the whistle on agent S." The message is observed by B and S but not by outsiders.

Recall that (t_B^*, t_S^*) are transfers which satisfy (12) and (13). The outcome function is as follows.

Rule 1. If $x = x(b^*,s^*)$, pay $t_i = t_i^*$ to each $i \in \{B,S\}$, and set $t_T = 0$.

Rule 2. If $x \neq x(b^*,s^*)$ and T reports β, then B is paid $t_B = -2F$, T is paid $t_T = x + F$, and S is paid $t_S = F$.

Rule 3. If $x \neq x(b^*,s^*)$ and T reports σ, then S is paid $t_S = -2F$, T is paid $t_T = x + F$, and B is paid $t_B = F$.

Rule 4. If $x \neq x(b^*,s^*)$ and T reports $\emptyset$, then pay $t_T = x$ to T, and set $t_B = t_S = 0$.

The key idea is that if the output is not first best and the third party "blows the whistle" on someone, then that person is punished (by Rule 2 or Rule 3). If B and S

expect that the third party will blow the whistle on them if they try to side contract with him, side contracting will be deterred. They will only want to accept the third party's proposal if the side contract can deter whistle blowing. Conversely, side contracting is prevented if every side contract induces a continuation equilibrium with whistle-blowing.

THEOREM 4: *We can choose $F > 0$, so that the game $G(\Gamma_0)$ has a perfect Bayesian equilibrium which implements the first-best outcome.*

To prove the theorem, a PBE which implements the first-best outcome is constructed as follows. At time 0, T's strategy specifies that no side contract is proposed. At time 1, any agent who has not signed a side contract sets effort at the first-best level. At time 3, T stays quiet if no side contract is in force.

Notice that the third party cannot effect the outcome by blowing the whistle on either agent as long as $x = x(b^*, s^*)$ (by Rule 1). At time 1, Rules 1 and 4 imply that both agents want to choose the first-best actions, anticipating that a deviation will lead to the entire output being given to the third party. From this it follows that, if there is no side contract in force at time 0, then the proposed strategies are sequentially rational from time 1 on. The outcome is first-best by construction.

It remains to specify behavior after a time 0 deviation, where T proposes a side contract. To support the equilibrium, such a deviation should not be profitable. To be specific, suppose T makes a proposal to B (the argument is similar for a proposal to S). We construct the strategies so that if B accepts and coalition $C = \{B, T\}$ forms, then either B or T will get no more than their equilibrium payoff. If T gets no more than zero, it is certainly not profitable for him to propose the coalition. If B gets no more than B^* from joining the coalition, then we stipulate that his equilibrium strategy is to reject the proposal, and this behavior is certainly sequentially rational. Knowing that B will reject, it is again not profitable for T to make the proposal.

So suppose B accepts the proposal. The coalition $C = \{B, T\}$ forms and a set of side-contracting mechanisms $\{\Gamma_C(x)\}_{x \in x}$ is in force. S is unaware of a deviation and will play as described above, i.e., his effort is s^*. The equilibrium strategies need to specify how the coalition of B and T will behave in $G(\Gamma_0)$ following the deviation. Moreover, strategies should be such that either B gets less than B^*, or T gets less than 0. Also, we specify, if possible, that T and B believe that S chose the effort $s = s^*$ at time 1. In addition, if possible, T infers B's effort from the joint output, assuming $s = s^*$. In other cases, i.e., if x is inconsistent with $s = s^*$, then we may leave the beliefs unspecified.

For a given x, the message game in the side contract cannot be used to (uniquely) extract truthful information about whether or not T blew the whistle in Γ_0. This argument is the same as in Section G. Since T's message in Γ_0 only changes the transfers, the strategic incentives in the side-contracting mechanism $\Gamma_C(x)$ do not depend on it. That is, whistle-blowing does not induce any "preference reversal" at time 4. Hence, the side contract $\Gamma_C(x)$ cannot be designed to (uniquely) extract information about whether or not T blew the whistle at time 3. By assumption, the side contract must induce *some* continuation equilibrium (T is not allowed to propose a badly behaved

contract that causes an existence problem). By the payoff-irrelevance argument, we may assume the continuation equilibrium strategies are such that the messages sent at time 4 by B and T only depend on x, not on whether or not T blew the whistle at time 3.

At time 3, we specify that T blows the whistle on B. As $F > 0$, this is sequentially rational for T.

Under this specification, if the coalition $C = \{B, T\}$ forms, either Rule 1 or Rule 2 will apply at time 3. In either case, if F is large enough, S will not get less than his equilibrium payoff S^*. But then, at least one of the members of the coalition is not made strictly better off. As argued above, this implies that T does not gain by making the proposal. This completes the proof of Theorem 4.

The key to the equilibrium construction is the third party's behavior at time 3. He stays quiet at time 3 if no side contract is in force. But if he belongs to a coalition, then he blows the whistle on the other member of the coalition. In order to be assured of a profitable deviation, the agents should design a side contract where whistle-blowing is punished, and therefore not attractive to T, in all continuation equilibria. However, this is impossible because blowing the whistle only triggers a monetary reward which is not "payoff relevant" at time 4. Thus, all side contracts *must* have a continuation equilibrium where whistle-blowing is *not* punished, and in such a continuation equilibrium, the third party may as well blow the whistle. We support the optimal PBE of $G(\Gamma_0)$ by selecting the "whistle-blowing continuation equilibrium" whenever a coalition is formed. As before, full implementation of the first-best can be achieved by augmenting the mechanism.

III. Concluding Comments

The theory of mechanism design studies the question: given some desirable outcome, does there exist a mechanism (contract) which implements it? If information is complete, the answer is often yes, although the optimal contract is typically quite complex (see Maskin and Sjöström 2003). However, there is one case where even implementation with complete information is quite difficult. That is, when there are only two agents and they can renegotiate inefficient outcomes (Maskin and Moore 1999). This case is emphasized in contract theory. If a third party is added, then the problem of renegotiation is less severe, but collusion may be a problem. The key issue becomes, what variables are contractible for a colluding subcoalition? To address this issue, a natural starting point is a symmetry assumption—the grand coalition and all subcoalitions have access to the same contracting technology. The symmetric case is intermediate between the two polar cases of no side contracting and perfect side contracting that have been prominent in the literature.

For collusion to be a problem, the collusive activity (including messages and side payments) must take place behind closed doors. It cannot be verifiable information for the judge who supervises the original contract, otherwise, he could punish the colluders. By symmetry, the original judge should be able to conduct his proceedings behind closed doors. Therefore, messages and cash transfers exchanged under the original contract are not verifiable information for the judge who supervises

the side contract. But this puts severe constraints on the ability to collude. We have shown that, as a consequence, the first best can be implemented in two well-known models, the buyer-seller model and the moral hazard in teams model.

Our results seem to suggest that collusion may not matter too much, and there may be no inefficiencies caused by "hold up." Notice, however, that we make some strong (but common) assumptions. We assume the original contract can impose large fines on the buyer and seller. In reality, fines may be restricted due to wealth constraints and limited liability. Adding limited liability constraints to the original contracting problem would make implementation harder. Of course, symmetry recommends imposing the same constraints on side contracting, as in Baliga and Sjöström (1998). We also assume the seller and the buyer have complete information about the true state of the world. If this assumption is dropped, the original contract is subject to incentive-compatibility constraints. Again, symmetry recommends imposing incentive-compatibility constraints also on side contracts, as in Laffont and Martimort (1997, 2000). Future research might reveal the exact circumstances where collusion matters, where contracts are optimally "incomplete," and where the hold-up problem cannot be fully resolved.

REFERENCES

Abreu, Dilip, and Arunava Sen. 1991. "Virtual Implementation in Nash Equilibrium." *Econometrica*, 59(4): 997–1021.

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey. 1994. "Renegotiation Design with Unverifiable Information." *Econometrica*, 62(2): 257–82.

Baliga, Sandeep, and Tomas Sjöström. 1998. "Decentralization and Collusion." *Journal of Economic Theory*, 83(2): 196–232.

Brusco, Sandro. 1997. "Implementing Action Profiles When Agents Collude." *Journal of Economic Theory*, 73(2): 395–424.

Celik, Gorkem. 2007. "Mechanism Design with Collusive Supervision." Unpublished.

Che, Yeon-Koo, and Donald B. Hausch. 1999. "Cooperative Investments and the Value of Contracting." American Economic Review, 89(1): 125–47.

Che, Yeon-Koo, and József Sákovics. 2004. "A Dynamic Theory of Holdup." *Econometrica*, 72(4): 1063–1103.

Edlin, Aaron S., and Stefan Reichelstein. 1996. "Holdups, Standard Breach Remedies, and Optimal Investment." *American Economic Review*, 86(3): 478–501.

Eswaran, Mukesh, and Ashok Y. Kotwal. 1984. "The Moral Hazard of Budget-Breaking." *RAND Journal of Economics*, 15(4): 578–81.

Felli, Leonardo. 1996. "Preventing Collusion through Discretion." Unpublished.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691–719.

Hart, Oliver D. 1995. *Firms, Contracts and Financial Structure*. Oxford: Oxford University Press.

Hart, Oliver D., and John H. Moore. 1988. "Incomplete Contracts and Renegotiation." *Econometrica*, 56: 755–85.

Hart, Oliver D., and John H. Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6): 1119–58.

Hart, Oliver D., and John H. Moore. 1999. "Foundations of Incomplete Contracts." *Review of Economic Studies*, 66(1): 115–38.

Holmström, Bengt R. 1982. "Moral Hazard in Teams." *Bell Journal of Economics*, 13(2): 324–40.

Laffont, Jean-Jacques, and David Martimort. 1997. "Collusion under Asymmetric Information." *Econometrica*, 65(4): 875–911.

Laffont, Jean-Jacques, and David Martimort. 2000. "Mechanism Design with Collusion and Correlation." *Econometrica*, 68(2): 309–42.

Maskin, Eric, and John H. Moore. 1999. "Implementation and Renegotiation." *Review of Economic Studies*, 66(1): 39–56.

Maskin, Eric S., and Tomas Sjöström. 2003. "Implementation Theory." In *Handbook of Social Choice and Welfare*, ed. Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura, 237–88. Amsterdam: North-Holland.

Maskin, Eric S., and Jean Tirole. 1999. "Unforeseen Contingencies and Incomplete Contracts." *Review of Economic Studies*, 66(1): 83–114.

Matsushima, Hitoshi. 1988. "A New Approach to the Implementation Problem." Journal of Economic Theory, 45(1): 128–44.

Mookherjee, Dilip, and Masatoshi Tsumagari. 2004. "The Organization of Supplier Networks: Effects of Delegation and Intermediation." *Econometrica*, 72(4): 1179–1219.

Nöldeke, Georg, and Klaus M. Schmidt. 1995. "Option Contracts and Renegotiation: A Solution to the Hold-up Problem." *RAND Journal of Economics*, 26(2): 163–79.

Palfrey, Thomas R., and Sanjay Srivastava. 1991. "Nash Implementation Using Undominated Strategies." *Econometrica*, 59(2): 479–501.

Pavlov, Gregory. 2006. "Colluding on Participation Decisions." Unpublished.

Segal, Ilya. 1999. "Complexity and Renegotiation: A Foundation for Incomplete Contracts." *Review of Economic Studies*, 66(1): 57–82.

Segal, Ilya, and Michael D. Whinston. 2002. "The Mirrlees Approach to Mechanism Design with Renegotiation (with Applications to Hold-Up and Risk-Sharing)," *Econometrica*, 70: 1–46.

Tirole, Jean. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics, and Organization*, 2(2): 181–214.

Watson, Joel. 2007. "Contract, Mechanism Design, and Technological Detail." *Econometrica*, 75(1): 55–81.