
1 Web Appendix: Equilibrium outcome under collusion (multiple types-multiple contracts)

We extend our setup by allowing more than two types of agent. The agent's type is now $\beta \in \{\beta_1, \beta_2, \dots, \beta_N\}$, where $\beta_i > \beta_j$ for $i > j$, and ν_i denotes the probability that the agent has type β_i , with $\sum_i \nu_i = 1$. In addition, define $\Delta\beta_{i,j} \equiv \beta_i - \beta_j$. Similarly, the auditor receives one signal which may be informative of the agent type. Denote by the superscript $w \in \{I, U\}$ the states of the world where the signal is informative (I) or uninformative (U). The signal is informative with probability ξ , and uninformative with probability $1 - \xi$. The auditor's signal belongs to the set $\{\sigma_1, \sigma_2, \dots, \sigma_N, \emptyset\}$, where $\sigma_i^I = \beta_i$ for all i , and $\sigma^U = \emptyset$ irrespective of the agent's realized type. Having this schedule in place, we derive the equilibrium outcome by following the same solution strategy adopted in the proof of proposition 3: We establish an upper bound on expected profit by deriving a couple of necessary conditions, and we show that this upper bound can be reached by an incentive scheme that is collusion proof and that satisfies the properties discussed in section 3.2. The discussion introducing the previous appendix largely applies;¹ here, we point out where they differ. First modify the conditions on *final allocations*, equations (A.1)-(A.7) as follows. For all $i = \{1, \dots, N\}$ and $w = \{I, U\}$:

$$\hat{s}_i^w = s_i^w + \tilde{s}_i^w \quad (\text{N.1})$$

$$\hat{t}_i^w = t_i^w - (1 + \lambda)\tilde{s}_i^w \quad (\text{N.2})$$

$$\tilde{s}_i^w \geq 0 \quad (\text{N.3})$$

$$\hat{U}_i^w = \hat{t}_i^w - \psi(e_i^w) \quad (\text{N.4})$$

$$\hat{V}_i^w = \hat{s}_i^w. \quad (\text{N.5})$$

In addition, the usual ex-ante and ex-post participation constraints must be met in equilibrium. For all $i = \{1, \dots, N\}$ and $w = \{I, U\}$:

$$\hat{V}_i^w \geq 0 \quad (\text{N.6})$$

$$\hat{U}_i^w \geq 0. \quad (\text{N.7})$$

$$E(\hat{U}|\beta_i) = \xi\hat{U}_i^I + (1 - \xi)\hat{U}_i^U \geq 0 \quad (\text{N.8})$$

$$E(\hat{V}) = \sum_i \nu_i \left[\xi\hat{V}_i^I + (1 - \xi)\hat{V}_i^U \right] \geq 0. \quad (\text{N.10})$$

¹As before, the auditor's participation decision cannot be used as a screening device by the principal. In any equilibrium the auditor must accept the mechanism.

As before, if (N.6) and (N.7) are satisfied, so will (N.8) and (N.10). Finally, the incentive compatibility constraints that exist whenever the signal is uninformative are:

$$\hat{U}_i^U \geq \max_{j \neq i} [\hat{U}_j^U + \Phi_i(e_j^U)], \quad (\text{N.11})$$

where $\Phi_i(e_j) = \psi(e_j) - \psi(e_j - \Delta\beta_{j,i})$ is the information rent gained by agent i mimicking type j . Next, as usual the following necessary condition must be satisfied for all i :

$$(1 + \lambda)(\hat{s}_i^I - \hat{s}_i^U) \geq \hat{U}_i^U - \hat{U}_i^I. \quad (\text{N.12})$$

As for proposition 3, the *equilibrium allocation* also includes the equilibrium contract selected by the agent at the participation stage, i.e., $\{\hat{t}_i^w, \hat{s}_i^w, \hat{U}_i^w, \hat{V}_i^w, \hat{\Gamma}_i^w\}_{i \leq N}^{w=I,U}$, where $\hat{\Gamma}_i^w$ represents the equilibrium contract selected by the agent in state (i, w) . Clearly, $\hat{\Gamma}_i^w$ need not be the same in all states. However, since the agent cannot distinguish between states U and I , the equilibrium contract selected by the agent must be the same in these two states. Hence, we simply denote the equilibrium contract by $\hat{\Gamma}_i$. The necessary conditions (N.6) and (N.7) must still be satisfied by the final allocation in any equilibrium. Condition (N.12) is still a necessary condition because, after the participation decisions are made, the agent and the auditor are both informed about the agent type, and could pretend being uninformed. However, as in appendix B, condition (N.11) need not bind: the principal can design a mechanism such that, in equilibrium, the agent selects a different contract in different states. To see the point formally, note that the agent makes his participation decision without knowing the auditor's signal. The i -agent expects to obtain $\xi \hat{U}_i^I + (1 - \xi) \hat{U}_i^U$ by selecting $\hat{\Gamma}_i$. On the other hand, he expects to obtain $(1 - \xi) [\hat{U}_j^U + \Phi_i(e_j^U)] + \xi \check{U}^I$ by mimicking type $j \neq i$ and selecting $\hat{\Gamma}_{j \neq i}$, where \check{U}^I denotes the payoff in the case where the agent selects $\hat{\Gamma}_{j \neq i}$ and the auditor reports that the type is not j . Clearly, the principal wants to set \check{U}^I as low as possible (compatibly with [N.7]) to discourage the agent from misreporting. Because of the option to exit, the lowest payoff is $\check{U} = 0$.² Thus, (N.13) can be written as

$$\xi \hat{U}_i^I + (1 - \xi) \hat{U}_i^U \geq (1 - \xi) [\hat{U}_j^U + \Phi_i(e_j^U)] \quad \forall j \neq i,$$

which is the set of incentive compatibility constraints that must hold on the equilibrium

²In the case where the agent selects $\hat{\Gamma}_j$ and the auditor reports that the type is different from j , the auditor should also be suitably incentivized to report her signal. Because this occurs out-of-equilibrium, incentivizing the auditor is costless for the principal.

path. Out of all possible $N - 1$ constraints, at most one will be binding:

$$\xi \hat{U}_i^I + (1 - \xi) \hat{U}_i^U \geq (1 - \xi) \left\{ \max_{j \neq i} [\hat{U}_j^U + \Phi_i(e_j^U)] \right\} \quad \forall j \neq i, \quad (\text{N.13b})$$

where this condition must be satisfied in order to prevent the agent from selecting another type's contract.³ We will show later in the proof that the relevant necessary condition is (N.13b) and not (N.11). For the time being, we remain agnostic.

We are now ready to determine the optimal allocation. The expected profit function is

$$W = G - \sum_{i=1}^N v_i \{ \xi (t_i^I + s_i^I + \beta_i - e_i^I) + (1 - \xi) (t_i^U + s_i^U + \beta_i - e_i^U) \}. \quad (\text{N.22})$$

Use (N.1), (N.2), (N.4), and rearrange the profit function to obtain,

$$\begin{aligned} W = G - \sum_{i=1}^N v_i \{ & \xi (\hat{s}_i^I + \lambda \tilde{s}_i^I + \beta_i - e_i^I + \hat{U}_i^I + \psi(e_i^I)) + \\ & (1 - \xi) (\hat{s}_i^U + \lambda \tilde{s}_i^U + \beta_i - e_i^U + \hat{U}_i^U + \psi(e_i^U)) \}. \end{aligned} \quad (\text{N.23})$$

subject to constraints (N.3), (N.6), (N.7), and either (N.11) or (N.13b) if we look at the collusion-free benchmark. Otherwise, (N.12) must also be satisfied if collusion is allowed. Irrespective of whether collusion is allowed or not, it is readily observed that

$$\hat{s}_i^U = 0 \quad \forall i, \quad (\text{N.24})$$

$$\tilde{s}_i^w = 0 \quad \forall i, w \quad (\text{N.25})$$

Thus, the optimal allocation under the threat of collusion is collusion-proof. Next, rewrite (N.23) as follows:

$$\begin{aligned} W = G - \sum_{i=1}^N v_i \{ & \xi (\hat{s}_i^I + \beta_i - e_i^I + \psi(e_i^I)) + (1 - \xi) (\beta_i - e_i^U + \psi(e_i^U)) \} \\ & - \sum_{i=1}^N v_i \{ \xi \hat{U}_i^I + (1 - \xi) \hat{U}_i^U \} \end{aligned}$$

We want now to verify that our necessary condition is indeed (N.13) and that in the rest of this proof we can ignore (N.11). Note that when $\hat{U}_i^I = 0$ (N.13b) reduces to

³The binding incentive compatibility constraint need not be the upward adjacent one. One can easily verify this by taking $N = 3$ and noting that the most efficient type might prefer mimicking contract $\hat{\Gamma}_3$ rather than $\hat{\Gamma}_2$.

(N.11). Otherwise for any $\hat{U}_i^I > 0$, the information rent implied by (N.13b) is smaller compared with (N.11), something that can be easily verified by observing that (N.13b) implies $\hat{U}_i^U \geq \max_{j \neq i} [\hat{U}_j^U + \Phi_i(e_j^U)] - \frac{\xi}{(1-\xi)} \hat{U}_i^I \forall i$. Hence, given that (N.7) requires $\hat{U}_i^I \geq 0$, the principal needs to leave the agent with a (weakly) lower information rent in the uninformative stage ($w = U$) by forcing him to reveal his type at the participation stage, rather than eliciting it once the contract has been selected. However, this comes at the cost of leaving an additional information rent in the state where the auditor's signal is informative ($w = I$). This cost is \hat{U}_i^I . But \hat{U}_i^I is paid out with probability ξv_i , whereas $\frac{\xi}{(1-\xi)} \hat{U}_i^I$ is accrued to the principal with probability $(1 - \xi)(v_i + \sum_z v_z)$ where z represents the (potentially empty) set of relevant types.⁴ Clearly the benefits (weakly) outweigh the costs. It follows that the principal prefers to screen the agent at the participation stage, and (N.13b) is our necessary condition. Observe that (N.13b) should always be binding and can be substituted directly into the profit function presented above to obtain

$$W = G - \sum_{i=1}^N v_i \{ \xi (\hat{s}_i^I + \beta_i - e_i^I + \psi(e_i^I)) + (1 - \xi) (\beta_i - e_i^U + \psi(e_i^U)) \} - (1 - \xi) \sum_{i=1}^N v_i \left\{ \max_{j \neq i} [\hat{U}_j^U + \Phi_i(e_j^U)] \right\} \quad (\text{N.23b})$$

From this point onward our proof with and without the threat of collusion needs be further specialized. Let us first consider the collusion-free benchmark.

1.1 Collusion-free Benchmark

In the collusion-free benchmark the principal receives σ directly (or alternatively the principal can perfectly control the auditor's actions). Thus, $\hat{s}_i^I = 0 \forall i$. To show that $\hat{U}_j^U = 0$ for all j , it suffices to note that maximizing W in (N.23b) with respect to \hat{U}_j^U is equivalent to maximizing $-(1 - \xi)v_z \hat{U}_j^U$ for all the relevant types z .⁵ Clearly, given the constraint $\hat{U}_j^U \geq 0$,

⁴From (N.13b) follows that the set includes (a) z' -types whose incentive constraint (N.11) binds for type i , i.e., $\arg \max_x [\hat{U}_x^U + \Phi_{z'}(e_x^U)] = i$, (b) z'' -types whose incentive constraint binds for type z' -types, and (c) so on and so forth recursively.

⁵A quick look at the last term of (N.23b) reveals that the set includes all z' -types whose incentive constraint (N.11) binds for type j , all z'' -types whose incentive constraint (N.11) binds for z' -types, and so on and so forth recursively.

optimality requires $\hat{U}_j^U = 0$ for all j . From the binding (N.23b) follows that

$$\begin{aligned}\xi \hat{U}_j^I &= (1 - \xi) \max_{z \neq j} \left[\hat{U}_z^U + \Phi_i(e_z^U) \right] \\ &= \max_{z \neq j} \left[0 + \Phi_i(e_z^U) \right] \\ &= (1 - \xi) \Phi_j(e_N^U) \quad \forall j.\end{aligned}$$

Before proceeding, it is important to point out that there is a problematic aspect to the implementation of such a mechanism. Namely, the principal can withhold the hard evidence in the informative state of the world (i, w) and gain $\hat{U}_i^I - \hat{U}_i^U = \hat{U}_i^I$. The agent knows the principal's signal but cannot access its hard evidence, making him vulnerable to this kind of manipulation. If the principal is unable to commit to disclose the hard evidence, the mechanism unravels. In order to avoid this problem the principal is forced to set $\hat{U}_i^U \geq \hat{U}_i^I \quad \forall i$, which is motivated by the observation that the principal can always report that the signal is uninformative but not viceversa. Given that we showed that the profit decreases monotonically with \hat{U}_i^U , the condition needs to be binding, i.e., $\hat{U}_i^U = \hat{U}_i^I$, which restores the insurance result from the 2-types case. Then, (N.13b) requires $\hat{U}_i^U = \hat{U}_i^I = (1 - \xi) \max_{j \neq i} \left[\hat{U}_j^U + \Phi_i(e_j^U) \right]$. As before, effort is first best in the informative states and $\{e_i^U\}_{i=1, \dots, N}$ solves (N.23b) subject to the monotonicity constraint.

1.2 Collusion-Proof Benchmark

Under the threat of collusion, condition (N.12) binds. Substituting this in (N.23b),

$$\begin{aligned}W = G - \sum_{i=1}^N v_i \left\{ \xi \left[\frac{\hat{U}_i^U - \hat{U}_i^I}{1 + \lambda} + \beta_i - e_i^I + \psi(e_i^I) \right] + (1 - \xi)(\beta_i - e_i^U + \psi(e_i^U)) \right\} \\ - (1 - \xi) \sum_{i=1}^N v_i \left\{ \max_{j \neq i} \left[\hat{U}_j^U + \Phi_i(e_j^U) \right] \right\},\end{aligned}$$

it is clear that maximizing W with respect to \hat{U}_j^U is equivalent to maximizing $-(1 - \xi)v_z \hat{U}_j^U - \xi v_j \frac{\hat{U}_j^U}{(1 + \lambda)}$, where the first term is the same as the one in the Collusion-Free Benchmark (see previous section). Thus, W decreases monotonically with \hat{U}_j^U . On the other hand, maximizing W with respect to \hat{U}_j^I is equivalent to maximizing $\xi v_j \frac{\hat{U}_j^I}{(1 + \lambda)}$. Thus, W increases monotonically with \hat{U}_j^I . However, the same issue that afflicted the implementation of this collusion-free allocation, also afflicts that of the collusion-proof one: there is scope for principal-auditor coalition formation, wherein the principal bribes the auditor to withhold the hard evidence in the informative state of the world. The mechanism is unraveled as a result. As in the

collusion-free benchmark, the principal is forced to set $\hat{U}_i^U \geq \hat{U}_i^I \forall i$ in order to rule this problem out. But profit decreases monotonically with \hat{U}_i^U , implying again that the condition needs to be binding, i.e., $\hat{U}_i^U = \hat{U}_i^I$. Having this schedule in place, (N.12) is satisfied by setting $\hat{s}_i^I = 0$; collusion is eliminated at zero cost. Thus, the optimal effort levels and the equilibrium allocation are the same as in the collusion-free benchmark. This proves that the upper bound on expected profit is equal to the collusion-free profit (second-best.)

1.3 The second-best contract

Having established the collusion-proof benchmark, we next show that there exists a contract that achieves the upper bound. The contract is as follows: Take a direct revelation mechanism where the auditor reports her signal by sending the message m_{s2} , the agent chooses contract $\hat{\Gamma}_i$, and then, once the contract is selected, reports his type by sending message m_{a2} . Let (i, w) be the state in which $m_{s2} = \sigma_i^w$, and $m_{a2} = \beta_i$. Next consider the following transfers and cost targets under contract $\hat{\Gamma}_i$

$$s_i^w = \hat{s}_i^w \quad (\text{A.19})$$

$$t_i^w = \hat{U}_i^w + \psi(e_i^w) \quad (\text{A.20})$$

$$C_i^w = m_{a2} - e_i^w \quad (\text{A.21})$$

where $\{e_i^w, \hat{s}_i^w, \hat{U}_i^w\}$ are the solutions to the maximization of the collusion-free benchmark for each pair (i, w) . Note that in the above the utility of the agent is insured, $\hat{U}_i^I = \hat{U}_i^U$, but not his effort since it is possible that $e_i^I \neq e_i^U$. We have left to specify the out of equilibrium payoffs, both when the auditor's signal is informative and uninformative.

Informative signal. Let us denote by $\check{U}_{i,j}^I = 0$ and $\check{s}_{i,j}^I = L$ respectively the i -agent's and the auditor's payoff when the agent selects contract $\hat{\Gamma}_{j \neq i}$, and the auditor reports $m_{s2} = \sigma_i^I = \beta_i$, where $L = (U_{i,j}^U - U_{i,j}^I)/(1 + \lambda)$. Note that $\check{U}_{i,j}^I = 0$ can be achieved by setting the payout to the agent $t_{i,j}^I = \psi(e_i^I)$. Moreover, the auditor obtains a payoff equal to zero if she reports $m_{s2} = \emptyset$, irrespective of the agent's signal. Thus, L ensures that the auditor reports truthfully when her signal is informative, because her out-of-equilibrium payoff satisfies (N.12) with equality.⁶

Uninformative signal. Before proceeding, let us highlight one feature of the FO from the simple 2-type model: In the (out-of-equilibrium) event that the efficient agent selects the

⁶An argument for reducing the out-of-equilibrium payoff to the minimum is presented in the next paragraph.

insourcing contract and the auditor’s signal is uninformative, the agent is still incentivized to truthfully reveal his type and to exert optimal effort.⁷ This might be useful if one is concerned with the possibility of random realization of out of equilibrium events. This could occur if the agent makes a mistake in selecting the contract or in evaluating the probability of a successful inspection by the auditor.⁸ It is interesting to investigate whether the principal can implement such a feature in the N-type case as well. In other words, can the principal costlessly ensure that in the (out of equilibrium) event that type $i < j$ selects $\hat{\Gamma}_j$, the latter reports his type truthfully and exerts optimal effort? It is easy to show that the answer is negative. A simple proof by contradiction suffices. Suppose not. Take contract $\hat{\Gamma}_N$ and, within that contract, let the principal offer a menu of transfers t_i^U and corresponding efforts e_i^U as defined in (A.20) for all $i < N$. Clearly, this set of transfers and efforts would be incentive compatible if it were to satisfy (N.11). But this cannot be the case: the transfers t_i^U in (A.20) are constructed to satisfy (N.13b) but to violate (N.11). Moreover, altering t_i^U and e_i^U to satisfy (N.11) would prevent the principal from implementing the second best. (Recall that the implementation of the collusion-free allocation hinges precisely on ignoring [N.11] as discussed above.)

A Comment on Robustness The principal can implement a mechanism that (partially) retains truthful revelation off the equilibrium path even when the signal is uninformative. The mechanism allows types that are more efficient than i to report their true type under contract $\hat{\Gamma}_i$ and (with a probability) earn an information rent. It does not allow agents that are less efficient than i to do the same.

Consider the mechanism that provides, with probability τ , a fixed transfer t_i^U , fixed effort e_i^U , and fixed cost reimbursement C_i^U when contract $\hat{\Gamma}_i$ is selected. With probability $(1 - \tau)$, contract $\hat{\Gamma}_i$ specifies a menu of transfers $\tilde{t}_{j,i}^U$ and corresponding efforts $\tilde{e}_{j,i}^U$ where j indexes the report given by the agent. We construct these transfers in such a way that $\tilde{t}_{i,i}^U = t_i^U$ and $\tilde{e}_{i,i}^U = e_i^U$. That is, the transfer provided to type i agent who (correctly) chose contract $\hat{\Gamma}_i$ is the same as the one in (A.20). We next construct the remaining transfers $\tilde{t}_{j,i}^U$ such that agent $j < i$ will report being type j and will receive a cost reimbursement $C_{j,i}^U$. These transfers must satisfy [N.11]. By the standard treatment, the agent’s lowest utility levels compatible with a weakly decreasing effort profile (i.e., $\tilde{e}_i^U \geq \tilde{e}_j^U \forall i < j$) are derived using the binding upward adjacent incentive compatibility constraints for all types, expect

⁷In this case, the cost to the principal stems from the fact that the auditor receives a positive payoff when the signal is informative—something that is avoided if the efficient agent selects the outsourcing contract.

⁸This concern does not affect the LT mechanism. In LT the agent makes his report at the stage where the auditor’s signal is already realized and common knowledge. Being a problem inherent to our FO, it is potentially important for our mechanism to address this problematic aspect.

type i . When [N.11] is satisfied, type i has no incentive to misreport as type $j < i$, and no extra rent is required to prevent this misreport. In addition, the principal can set τ high enough to reduce the ex ante information rents going to agent $j < i$ under contract $\hat{\Gamma}_i$, such that this agent will select contract $\hat{\Gamma}_j$.⁹

We finally conclude by showing that it is always costly for the principal to obtain truthful revelation of types $j > i$, and so the principal will never do so. The fundamental problem is that truthful revelation of types $j > i$ would require an increase in the information rent available to agent i : that is, t_i^U would need to be larger. This is a problem because t_i^U is paid out on the equilibrium path, and increasing rents for agent i in equilibrium reduces the principal's expected profits.¹⁰ Thus, it is necessary to impose a cap in the effort schedule available in each contract. In particular, the effort in contract $\hat{\Gamma}_i$ must be capped below e_i^U , and costs are reimbursed by the principal up to C_i .

⁹Trivially, this is always the case when τ is arbitrarily close to zero.

¹⁰To see this, note that the second-best transfer t_i^U in (A.20) and corresponding efforts e_j^U satisfy [N.13] (by construction), but also strictly violate [N.11] (again, by construction). To construct transfers $\tilde{t}_{j,i}^U$ and corresponding efforts $\tilde{e}_{j,i}^U$ for $j \geq i$ requires that [N.11] is satisfied. The resultant transfer for type i would be $\tilde{t}_{i,i}^U > t_i^U$.