ARTICLE IN PRESS
JID:YJETH AID:105614 /FLA [m1+; v1.361] P.1 (1-33)
Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory ••• (••••) ••••••
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Regulatory mechanism design with extortionary collusion ☆

Dilip Mookherjee [a], Masatoshi Tsumagari [b]

[a] *Department of Economics, Boston University, 270 Bay State Road Boston MA 02215, United States of America*
[b] *Department of Economics, Keio University, 2-15-45 Mita Minato-ku Tokyo 108-8345, Japan*

Received 21 August 2021; final version received 13 January 2023; accepted 22 January 2023

Abstract

We study regulatory mechanism design with collusion between a privately informed agent and a less well-informed supervisor, incorporating 'extortion' which permits redistribution of rents within the coalition. We show the Collusion Proof Principle holds, and that the allocation of bargaining power between the supervisor and agent matters. Specifically, the Principal does not benefit from hiring the supervisor if the latter has less bargaining power vis-a-vis the agent. We provide an example where hiring the supervisor is valuable if she has greater bargaining power. These results indicate the importance of anti-collusion strategies that augment bargaining power of supervisors vis-a-vis agents.

JEL classification: ???

Keywords: Mechanism design; Supervision; Collusion; Extortion; Bargaining power

1. Introduction

The design of mechanisms to limit the harmful effects of collusion between supervisors and agents in adverse selection settings has been studied by many authors following Tirole (1986)

☆ We thank an Associate Editor and referees of this journal, Alessandro Bonatti, Robert Gibbons, Alberto Motta, Juan Ortner, Silvia Vanutelli and participants in seminars at Boston University, MIT and New York University for useful discussions and comments.

E-mail addresses: dilipm@bu.edu (D. Mookherjee), tsuma@econ.keio.ac.jp (M. Tsumagari).

https://doi.org/10.1016/j.jet.2023.105614
0022-0531/

Please cite this article as: D. Mookherjee and M. Tsumagari, Regulatory mechanism design with extortionary collusion, Journal of Economic Theory, https://doi.org/10.1016/j.jet.2023.105614

and Laffont and Tirole (1993), with many applications to design of procurement, regulation and internal organization of firms. Subsequent literature has explored the implications of enlarging the severity of the collusion problem, such as soft information which provides wider scope for manipulation of reports (Faure-Grimaud et al. (2003), Celik (2009)), and collusion over both reporting and participation decisions (Mookherjee et al. (2020)).

All these papers however assume that collusion is 'weak' in the sense that the supervisor and agent play non-cooperatively if either vetoes the offered side-contract. This ensures collusion occurs only to realize joint gains for the colluding parties. The allocation of bargaining power within the coalition then does not matter for the Principal's capacity to control corruption. This implication of weak collusion follows from the Collusion Proof Principle (CPP) which applies quite generally in this class of models (Tirole (1992)).[1] It is important to note that the irrelevance of bargaining power (among parties engaging in weak collusion) applies more generally than conditions usually needed for the Coase Theorem to apply: for instance, bargaining takes place with asymmetric information between the supervisor and agent. These models are therefore incapable of explaining the observed success of anti-corruption policies which lowered bargaining power of the agent vis-a-vis the supervisor by preventing agents from selecting their own auditor in India (Duflo et al. (2013)) and Italy (Vanutelli (2020)).[2]

In this paper we argue that a strengthening of the nature of collusion to allow for 'extortion' can explain why relative bargaining power of the supervisor could matter. Extortion permits the colluding partner with greater bargaining power to extract rents from the other party, by threatening to send reports to the Principal which would hurt the latter if (s)he did not agree to the offered side-contract. Colluding parties can commit to such threats at the time of bargaining. This alters the sub-game following refusal of an offered side-contract by one party, from a simultaneous move noncooperative game, to one where the report of the accepting party is stipulated by the side-contract (with the other party choosing a best response). Earlier work by Dequiedt (2007) and Che and Kim (2009) has studied an analogous notion in an auction design setting, using the term 'strong collusion'. As some other authors (e.g., Quesada (2003)) have used the term 'strong collusion' to mean something different, we refer to the combination of extortion and collusion as 'extortionary collusion'.

While we do not endogenize the source of commitment power, we have in mind settings where S and A interact with one another in many other transactions with many principals, either at the same time or at later dates, and reputational concerns provide the required enforcement.[3] It should be added that this is consistent with the rest of the mechanism design literature which examines the consequence of alternative assumptions regarding commitment among players, rather than providing an explicit microfoundation for these assumptions. The existing literature which

[1] See Faure-Grimaud et al. (2003) or Mookherjee et al. (2020). Here is an outline of the argument. CPP implies that attention can be confined to revelation mechanisms which do not allow scope for any deviating (feasible) side contract for the coalition to increase the welfare-weighted sum of their payoffs. Hence there cannot exist a deviating feasible side contract which makes both strictly better off. As this property is independent of the relative welfare weights, it follows that the allocation of bargaining power within the coalition does not matter.

[2] See Section 2 for more details of these policies and resulting outcomes.

[3] Consider a setting with many customers, each of whom wants to contract with an agent to commission a required production task, along with a supervisor who is better informed than customers about the cost of completing the task. There are many ex ante identical potential agents and brokers/supervisors in the industry which constitutes a close-knit community, in which members know one another, enter into side-contracts whose outcomes are observed by the rest of the community. If they are far-sighted, any agent and supervisor appointed by a given customer have an incentive to develop and maintain a reputation for following through on promises and threats they make during side contracting.

has studied weak collusion is based on a specific assumption of lack of commitment among colluding players. Here we are interested in understanding the consequence of the opposite assumption where they can commit to reporting threats when they fail to agree on a side-contract.

In a law review article, Ayres (1997) refers to 'bribery' and 'extortion' as the 'twin faces of corruption'. In the language of mechanism design theory, bribery corresponds to weak collusion, while the combination of bribery and extortion corresponds to extortionary collusion. The importance of extortion has been stressed by numerous authors in descriptive accounts of corruption, from medieval England (Cam (1930)), to more contemporary accounts of corruption in Burma (Furnivall (1956)) or other developing countries (Klitgaard (1988)). Ayres (1997) and Andrianova and Melissas (2008) discuss extortion from a legal standpoint. Some papers studying tax evasion, regulations or intra-firm organization in a moral hazard setting (Mookherjee (1997), Hindricks et al. (1999), Khalil et al. (2010)) have shown that anti-corruption policy design is significantly altered by the presence of extortion. By contrast our focus is on an adverse selection context, where the modeling issues as well as results are quite different. Section 2 provides a more detailed discussion of the relation to existing literature.

Section 3 describes the model of extortionary collusion in which S obtains a noisy signal of A's cost, and A also observes the realization of this signal. A and S enter into a collusive side contract at the ex ante stage, when neither have received their respective signals. Moreover, the allocation of bargaining power between the supervisor (S) and agent (A) is given and known by the Principal (P). The primary question studied in this paper is how relative bargaining power matters in the presence of extortion. In some contexts P may have no control over the allocation of bargaining power between S and A. Our results will show that the consequences of extortion on P's welfare and the value of hiring S vary with bargaining power allocation. In other contexts P may be able to influence bargaining power: in these settings our results imply that raising relative bargaining power of S can be an important policy instrument.

For instance, P may be able to influence the process by which S and A are appointed, which affects their relative bargaining power. Consider a setting where A is allowed to appoint her own supervisor, from a set of potential supervisors. This would give A the opportunity to exercise monopsony power over S in the choice of side contract, resulting in a higher welfare weight on A's payoff. An alternative institutional rule (as in the policy reforms in India and Italy studied by Duflo et al. (2013) and Vanutelli (2020)) is one where the supervisor is selected by P instead, and assigned to a given agent. This would alter the side contract negotiation to a bilateral monopoly where bargaining power is more equally divided between S and A. In some settings (e.g., in construction, or procurement of a particular service) there could be many potential agents available to carry out the project for P, where P could appoint a supervisor S and delegate the choice of A to S. This would confer monopoly power to S over A.

Given a certain allocation of bargaining power within the coalition, Section 4 verifies that the Collusion Proof Principle continues to apply in a extortionary collusion setting, once message spaces are augmented to include some non-type messages. This result is of some independent interest, as it contrasts with the moral hazard setting studied by Khalil et al. (2010). Other differences in our results from the moral hazard setting are described in Section 2. The result is used to characterize the class of feasible allocations in terms of a set of coalition incentive compatibility constraints.

This characterization is used in Section 5 to provide results for alternative ranges of bargaining power allocation. First, if A has a greater bargaining power than S,[4] appointing a supervisor is worthless for P (Proposition 2). To explain the underlying intuition, consider the case where A has all the bargaining power, and the side-contract is designed to maximize A's ex ante payoff subject to S's participation constraint. Since A knows the realization of S's signal as well as the latter's outside option, there is no asymmetric information within the coalition. The resulting absence of bargaining frictions allows A to extort all of S's rents without creating any distortion, implying that the coalition behaves in a manner that maximizes their joint payoff. In effect, A becomes the residual claimant to the coalition's joint surplus and the presence of S becomes redundant (by having no choice other than to 'rubber-stamp' whatever A wants to report). This logic extends to the case where A has greater (but not all) the bargaining power.

Turning next to contexts where S has greater bargaining power, we provide an example to illustrate that hiring S can be valuable to P. Intuitively, the main difference from contexts where A has higher bargaining power is that the 'direction' of extortion is reversed: the side-contract is now designed to transfer rents from A to S and is subject to asymmetric information constraints (owing to A's private information regarding the realization of production cost). The associated bargaining frictions resemble 'sand in the wheels of corruption' which benefits P. The example features procurement of an indivisible good by P, where S's signal has two possible realizations and satisfies a monotone likelihood property. We construct a specific mechanism where hiring S enables P to achieve a higher payoff. An earlier working paper version (Mookherjee and Tsumagari (2022)) explores the corresponding optimal contracting problem with extortionary collusion, and shows that (conditional on S being valuable) the value of hiring S is globally decreasing in S's bargaining power over the range where S has greater bargaining power.[5]

The concluding Section discusses consequences of extending the model in different directions, while the Appendix provides details of proofs omitted from the text.

2. Related literature

In the mechanism design literature, implications of a similar notion of extortionary collusion have been studied in the context of auction design featuring collusion among bidders, by Dequiedt (2007) and Che and Kim (2009). Quesada (2003) uses the term 'strong collusion' to refer to a different concept, which requires every Bayesian equilibrium of the game induced by P's contract to be weakly collusion proof. We are not aware of any previous study of the implications of extortionary collusion in settings of supervision with a privately informed agent.[6] In such a setting, our analysis shows how standard characterizations of feasible and optimal allocations in the existing literature on weak collusion (Faure-Grimaud et al. (2003), Celik (2009), Mookherjee et al. (2020)) need to be modified, besides providing results concerning the costs imposed by extortion and how they depend on intra-coalition bargaining power allocation. Most importantly, manipulating the bargaining power of the supervisor becomes an important tool of

[4] By this we mean A is assigned a higher welfare weight in the selection from the set of feasible side-contract payoff pairs.

[5] However the working paper restricts attention to optimal deterministic contracts, which may entail some loss of generality in the setting of an indivisible good.

[6] However there have been a number of studies of extortionary collusion in a moral hazard setting, described below.

the Principal in curbing the costs of corruption, by increasing asymmetric information frictions within the supervisor-agent coalition.[7]

The practical relevance of these results is highlighted by recent policy experience. Duflo et al. (2013) study a controlled experiment in India during 2009-10 where a treatment group of firms were no longer allowed to appoint their own pollution auditors, but were randomly assigned auditors instead. They found significant increases in pollution reports by the assigned auditors, and corresponding decline in actual pollution levels (verified from special backchecks conducted by the research team). Vanutelli (2020) studies a related policy reform in Italy introduced in 2011, where auditors of municipal budgets were randomly assigned instead of being appointed by local mayors. This resulted in increased property tax collections, larger budget surpluses and debt repayments. These effects were significantly larger in places with higher 'risk of corruption' (measured by prior investigations of corruption-related crimes).[8]

The consequences of extortion for agent incentives in a supervision setting with moral hazard have been studied by a class of models (Mookherjee (1997), Hindricks et al. (1999), Khalil et al. (2010)). The most closely related paper in this group is Khalil et al. (2010), which also finds that extortion imposes a larger cost to the Principal's welfare when the agent has larger bargaining power. However, many features and results of their model differ from ours. The information structure is different (the supervisor is either perfectly informed or perfectly uninformed; the supervisor's information is hard unless the agent agrees to collude), and the Collusion Proof Principle does not hold in their setting. Most of their attention is subsequently devoted to the question whether bribery (i.e., resulting in mutual gain to both supervisor and agent) or extortion is the 'greater evil'. This question is not meaningful in our setting as the Collusion Proof Principle applies in our setting and thereby implies the optimality of eliminating both bribery and extortion. Moreover, in contrast to our results, hiring the supervisor is still valuable in their model even if the agent has all the bargaining power within the coalition (when the supervisor's signal is informative enough).

3. Model

3.1. Preferences, technology and information

An appointed agent A delivers an output q to the Principal P at a personal cost of θq. Let $Q \subseteq \Re_+$ denote the set of feasible outputs. We do not impose any restriction on Q except for $0 \in Q$. In Section 5.3 we consider the specific case where $Q = \{0, 1\}$. P's return from q is denoted by $V(q)$ which is increasing on Q, with $V(0) = 0$. The realization of θ is privately observed by A. Θ, which denotes the support of θ, constitutes an interval $[\underline{\theta}, \bar{\theta}] \subset (0, \infty)$. It is common knowledge that everybody shares a common distribution function $F(\theta)$ over Θ. It has a density function $f(\theta)$ which is continuous and everywhere positive on Θ.

[7] This is broadly similar though different in details from the strategy studied by Ortner and Chassang (2018) and von Negenborn and Pollrich (2020) in which P deliberately creates asymmetric information between agent and monitor by randomizing the latter's incentive contract and not letting the agent observe the monitor's contract with P. This particular tool is presumed unavailable in the settings we examine, e.g., the incentive contract for both parties is required to be in the public domain, as is commonly the case for public sector procurement or regulatory settings.

[8] Using suitable extensions of our model, an earlier version of this paper illustrated how random assignment of a supervisor would raise bargaining power of the supervisor relative to a context where the agent appoints the supervisor. The idea is simple: if there are many competing potential agents and supervisors on each side of the market, giving a chosen agent or supervisor the power to appoint the other party would tilt relative bargaining power in its favor.

An appointed supervisor S costlessly acquires an informative signal $\eta \in \Pi \equiv \{\eta_1, \eta_2, ..., \eta_m\}$ about A's cost θ with $m \geq 2$.[9] The realization of S's signal is observed by A. $a(\eta \mid \theta) \in [0, 1]$, which denotes the likelihood function of η conditional on θ, is continuous and positive-valued on Θ.[10] We assume that for any $\eta \in \Pi$, $a(\eta \mid \theta)$ is not a constant function on Θ, and there are some subsets of θ with positive measure satisfying $a(\eta \mid \theta) \neq a(\eta' \mid \theta)$ for every $\eta, \eta' \in \Pi$. In this sense each possible signal realization conveys information about the agent's cost. The information conveyed is partial, since Π is finite. The distribution function over θ conditional on η is denoted $F(\theta|\eta)$. Conditional on η, the density function and distribution function are respectively denoted by $f(\theta \mid \eta) \equiv f(\theta)a(\eta \mid \theta)/p(\eta)$ and $F(\theta \mid \eta) \equiv \int_{\underline{\theta}}^{\theta} f(\tilde{\theta} \mid \eta)d\tilde{\theta}$, where $p(\eta) \equiv \int_{\underline{\theta}}^{\bar{\theta}} f(\theta)a(\eta \mid \theta)d\theta$. Let $K \equiv \Theta \times \Pi$ denote the set of possible states.

All players are risk neutral. S's payoff is $u_S = X_S + t_S$ where X_S is the transfer from P to S, and t_S is a transfer received by S within the coalition. A's payoff is $u_A = X_A + t_A - \theta q$ where X_A is the transfer from P to A and t_A is a transfer received by A within the coalition. Transfers within the coalition are subject to a budget balance condition $t_A + t_S \leq 0$. P's objective is to maximize the expected value of profit $V(q) - X_A - X_S$.[11]

3.2. *Mechanism, collusion game and equilibrium concept*

P designs a grand contract (GC) played by A and S, describing transfers and production decisions (X_A, X_S, q) made by P in response to message sent by A and S. It is possible that selection of (X_A, X_S, q) is randomized conditional on messages. Owing to the risk neutrality of A and S and linearity of their payoff functions, grand contracts generating the same expected value of (X_A, X_S, q) for any given message generate the same payoffs and incentives for A and S. Hence they induce the same expected allocation (i.e., expected value of (X_A, X_S, q) in each state of the world) in any equilibrium of the collusion game (to be specified below). Moreover, if some expected allocation is achieved in an equilibrium, any randomized allocation with the same expected value is also achievable with the use of a stochastic grand contract. Define $\bar{Q} \equiv [0, \sup Q]$, which is the set of feasible values of expected output. This argument implies that without loss of generality, P can restrict attention to grand contracts with deterministic transfers, which assigns $(X_A, X_S, q) \in \Re^2 \times \bar{Q}$ for any possible message combination[12]:

$$GC = (X_A(m_A, m_S), X_S(m_A, m_S), q(m_A, m_S); M_A, M_S).$$

M_A (resp. M_S) denotes a message set for A (resp. S). Message spaces include exit options for A and S respectively ($e_A \in M_A, e_S \in M_S$), where $X_A = q = 0$ whenever $m_A = e_A$, and $X_S = 0$

[9] Even though our proof of CPP does not depend on the discrete property of Π, this assumption simplifies the exposition. If S incurs a fixed cost c to acquire the signal, transfers received by S must be replaced by transfers net of this fixed cost while measuring S's payoff. Increases in c will of course lower the value of appointing the supervisor, but it is easy to see how the results will be modified.

[10] This assumes that the support of θ given η is Θ for all η, in which sense θ has full support. We adopt this assumption purely to simplify the exposition; our results extend to the case of non-full support. See Concluding Section 6.

[11] Our analysis can be extended to the case that P maximizes a weighted average of profit ($V(q) - X_A - X_S$) and welfare of A and S ($u_A + u_S$), with a lower relative weight on the latter.

[12] Note that the allocations with the same expected value may not always perceived to be identical by P if $V(q)$ is not linear. Moreover, outputs may need to be randomized. If Q is an interval of the real line and V is concave, P would not benefit from randomizing the output, in which case attention can be restricted to deterministic contracts. In the indivisible good case $Q = \{0, 1\}$ examined in Section 5.3, P may conceivably benefit from randomizing the output. However, in Section 5.3 we restrict attention to deterministic contracts.

whenever $m_S = e_S$. The set of grand contracts satisfying these restrictions is denoted by $\mathcal{GC}$. As a special case, P has the option to not hire S, which we denote by *No Supervision* (NS), where M_S is null and $X_S \equiv 0$. We assume that GC is observable publicly, which makes it infeasible to combat collusion by creating asymmetric information (about the mechanism) between A and S, in contrast to Ortner and Chassang (2018) and von Negenborn and Pollrich (2020).

For technical convenience, we assume that M_A and M_S are compact subsets of finite dimensional Euclidean spaces, and allow message choices to be randomized. Let $\Delta(M_A)$, $\Delta(M_S)$ and $\Delta(M)$ denote the set of probability measures on M_A, M_S and $M \equiv M_A \times M_S$ respectively. For $(\mu_A, \mu_S) \in \Delta(M_A) \times \Delta(M_S)$ and $\mu \in \Delta(M)$, the expected values of allocations assigned in grand contracts are as follows:

$$\begin{aligned}\bar{GC} &\equiv (\bar{X}_A(\mu_A, \mu_S), \bar{X}_S(\mu_A, \mu_S), \bar{q}(\mu_A, \mu_S))\\ &= \int_{M_A}\int_{M_S} (X_A(m_A, m_S), X_S(m_A, m_S), q(m_A, m_S)) d\mu_A(m_A) d\mu_S(m_S)\end{aligned}$$

and

$$\hat{GC} \equiv (\hat{X}_A(\mu), \hat{X}_S(\mu), \hat{q}(\mu)) = \int_M (X_A(m), X_S(m), q(m)) d\mu(m).$$

$\bar{GC}$ denotes the expected value of the allocation (assigned for any given message combination) when A and S do not collude, and thus select their messages independently. $\hat{GC}$ denotes the corresponding expected value of the allocation when A and S collude, by coordinating on their respective messages according to the joint distribution μ. We restrict attention to mechanisms where $(\bar{X}_A(\mu_A, \mu_S), \bar{X}_S(\mu_A, \mu_S), \bar{q}(\mu_A, \mu_S))$ and $(\hat{X}_A(\mu), \hat{X}_S(\mu), \hat{q}(\mu))$ are continuous functions. These assumptions imply the standard minimax theorem (Nikaido (1954)) can be applied to S's payoffs[13]: in any GC, there exist minmax strategies $(\underline{\mu}_A, \bar{\mu}_S)$ for A and S respectively which satisfy

$$\begin{aligned}\underline{w}_S(GC) \equiv \bar{X}_S(\underline{\mu}_A, \bar{\mu}_S) &= \min_{\mu_A \in \Delta(M_A)} \max_{\mu_S \in \Delta(M_S)} \bar{X}_S(\mu_A, \mu_S)\\ &= \max_{\mu_S \in \Delta(M_S)} \min_{\mu_A \in \Delta(M_A)} \bar{X}_S(\mu_A, \mu_S).\end{aligned}$$

$\underline{w}_S(GC)$ denotes the minmax value of the S's payoff. Since S always has the option to exit from the grand contract, $\underline{w}_S(GC) \geq 0$ for any GC.

Collusion between A and S takes the form of a side contract (SC) which is unobserved by P. As explained in the Introduction, we treat the allocation of bargaining power as a parameter, and represent it by relative welfare weights on the ex ante payoffs of S and A at the time that the side contract is chosen by the coalition in response to the contract GC offered by P. Formally, the side contract is selected at the ex ante stage by a fictional (uninformed) third party acting as a mediator, who maximizes ex ante expected value of $\alpha u_A + (1 - \alpha) u_S$ ($\alpha \in [0, 1]$) in response to choice of GC made by P. The third party does not play any budget breaking role, hence the coalition needs to satisfy a budget balance condition $t_A + t_S \leq 0$.[14] No side payments can be

[13] Owing to the linearity of expected payoffs in mixed strategy probability measures, the required curvature assumptions for the minimax theorem are satisfied.

[14] It is possible that the third party receives a positive payment, whence $t_A + t_S < 0$. But we argue later that it cannot happen on the equilibrium path.

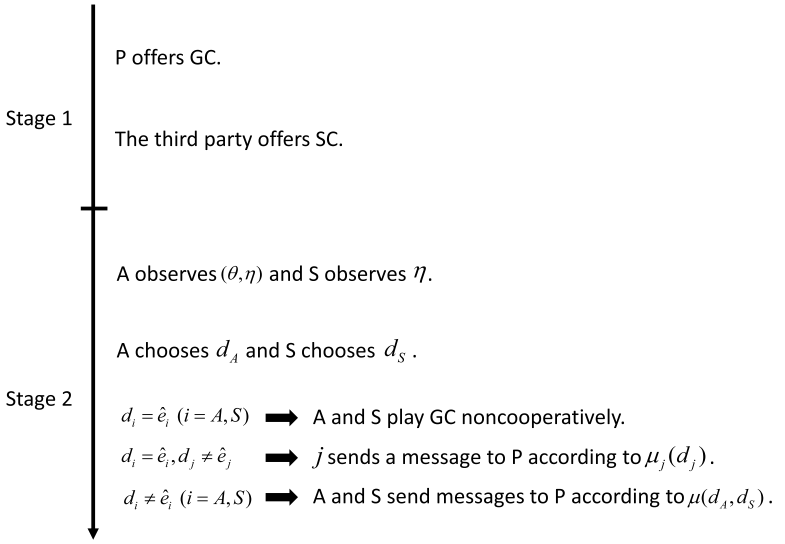

Fig. 1. The Timeline of the Game.

exchanged at the *ex ante* stage; they can only be exchanged at the *ex post* stage after payments from P have been received. The side contract cannot be renegotiated at the interim or *ex post* stage, while either party can decide to withdraw from the agreement at the interim stage. The side contract allows A and S to exchange messages privately among one another, which determine a side payment and joint set of messages they respectively send to P. Since message spaces include exit as well as type reports, collusion takes the *ex ante* form studied in Mookherjee et al. (2020).

The stages of the game are as follows (the timeline is shown in Fig. 1). Following the choice of GC by P, at stage 1 (the ex ante stage) the third party offers a side contract SC to S and A. A null side contract (NSC) could also be offered.

Next at stage 2 (the interim stage) S observes η and A observes (θ, η). If a NSC was offered, they play the GC noncooperatively based on their prior beliefs, just as in a game without any collusion. If a non-null side contract was offered, S and A independently decide whether to accept it. Specifically, the game proceeds as follows. $i = A, S$ selects a message $d_i \in D_i$ $(i = A, S)$ where D_i is i's message set specified in the side-contract; these messages are mutually observed. D_i includes i's exit option $\hat{e}_i$ from the side-contract. If $d_A \neq \hat{e}_A$ and $d_S \neq \hat{e}_S$, their reports to P are selected according to $\mu(d_A, d_S) \in \Delta(M)$, and side payments to A and S are determined according to functions $t_A(d_A, d_S)$ and $t_S(d_A, d_S)$ respectively.

If $d_A = \hat{e}_A$ and $d_S = \hat{e}_S$, A and S play GC non-cooperatively. What happens when one accepts and the other does not? Then SC specifies a reporting strategy of the party that accepted it, which can be interpreted as a threat that party commits to. The party that rejected it then plays a best response to this threat. Hence if $d_i \neq \hat{e}_i$ and $d_j = \hat{e}_j$ $(i, j = A, S)$, i's message to P is selected according to $\mu_i(d_i) \in \Delta(M_i)$, and the side payment to i is $t_i(d_i)$.[15] On the other hand, j plays GC without any constraint imposed by the side contract, and without any side transfer.[16]

15 Owing to the condition that the coalition cannot make a deficit, $t_i(d_i) \leq 0$.

16 When collusion is weak instead, the side contract ceases to apply for the subsequent messages for either player when one of them exits — S and A play GC noncooperatively.

We focus on Perfect Bayesian Equilibrium (PBE) of this extortionary collusion game induced by the grand contract GC and bargaining weight parameter α.[17] However, there may be multiple PBE in a given game. We assume collusion permits parties to coordinate the choice of a PBE: in the event of multiple PBE, the third party can specify a selected PBE to maximize the welfare-weighted sum of ex ante payoffs of S and A. The resulting equilibrium concept is denoted by PBE(c). In case there are multiple PBE(c) where the third party receives the same payoff, we assume that P can select the most desirable one.

In the economic environment, an allocation evaluated in the expected value conditional on (θ, η) (simply called an allocation in later part) is denoted by

$$(X, u_A, u_S, q) = \{(X(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), q(\theta, \eta)) \mid (\theta, \eta) \in K\}$$

where $X \equiv X_A + X_S$ denotes the total payment from P to the coalition. The budget balance condition implies $X(\theta, \eta) \geq u_A(\theta, \eta) + u_S(\theta, \eta) + \theta q(\theta, \eta)$ for any $(\theta, \eta) \in K$.

Achievable allocations in extortionary collusion can now be defined:

Definition 1. An allocation (X, u_A, u_S, q) is achievable in extortionary collusion with α if it is realized in PBE(c) under α for some $GC \in \mathcal{GC}$.

4. Collusion proof principle

In this section we establish a version of the Collusion Proof Principle (CPP, hereafter). It turns out that there are a number of additional complications involved in establishing CPP in the presence of extortion. Below we start with an informal description of these complications. Then we outline the argument we employed to overcome these problems, ending with a formal statement of CPP (whose proof is provided in the Appendix).

To simplify the exposition we abstract temporarily from the issue of agents' participation incentives, and focus on their incentives to report information to P (conditional on agreeing to participate in the mechanism). In the following discussion, it will be helpful to recall the distinction between collusion and extortion. 'Collusion' per se refers to the potential for joint deviation of the two agents (A and S respectively) in messages sent to P combined with a private side-transfer, which makes both better off relative to the status quo (i.e., what they would attain in the mechanism in the absence of the deviation). In contrast, 'extortion' refers to a potential for a joint deviation which results in redistribution of payoffs between the two agents. One agent then succeeds in driving down the payoff of the other below the latter's status quo payoff, with a suitable threat to send an 'off-equilibrium-path' report to P which would punish the latter if she were to veto the side-contract. Even if there is no scope for collusion (i.e., the mechanism selects a payoff vector on the utility possibility frontier of the coalition), it can still be vulnerable to extortion which results in a move *along* the frontier.

Extending the CPP to such a context would require showing that the outcome of any equilibrium of an arbitrary mechanism can be mimicked by a 'truthful' equilibrium of a some revelation mechanism which is immune to individual and joint deviations (which now includes both collusion and extortion possibilities). In a revelation mechanism each agent is provided a message space consisting of various possible reports of their private information — a report of (θ, η) for A, and η for S. When these respective reports are consistent, i.e., the same η is reported by both,

[17] For definition of PBE, see Fudenberg and Tirole (1991).

the mechanism selects the outcome resulting in the allocation in the equilibrium of the original mechanism GC when the true state is (θ, η). When the η reports are inconsistent, the mechanism specifies some 'off-equilibrium-path' allocation which deters agents from lying about their information regarding η. The standard argument for the CPP in the absence of extortion imposes 'large enough' penalties on either or both agents when their η reports are inconsistent. By reducing both individual and joint payoffs, such penalties discourage both individual and collective incentives for deviating when there is no scope for extortion. However, in the presence of extortion, a larger penalty imposed on one agent (S, say) for inconsistent reports can enlarge the scope for S to be extorted by A with a suitable 'off-equilibrium-path' threat which would drive down S's payoff sufficiently if S were to veto the side-contract. Hence the presence of extortion complicates the construction of 'cross-checking' mechanisms that ensure incentives for truthful reporting of η.

A somewhat different problem in establishing the CPP can arise when the original mechanism contains a 'larger' set of message options than are present in a revelation mechanism. In the latter, there is a message option corresponding to each possible report of the concerned agent's private information. These mimic the outcome of sending various messages that are sent on the equilibrium path of the original mechanism. But there may be other message options in the original mechanism which were never chosen in the concerned equilibrium. In particular, the maxmin reporting strategy of an agent may put weight on some off-equilibrium path messages. Since such message options are effectively deleted from message spaces in the corresponding revelation mechanism, the resulting maxmin payoffs could end up lower than in the original mechanism, thereby expanding the scope for extortion.

Despite these complications we nevertheless show that the CPP holds, provided the revelation mechanism is augmented to allow for some 'off-equilibrium' or 'non-type' message reports. The augmentation is minimal, consisting of one additional 'non-type' report provided to S. This is combined with a 'cross-checking' mechanism in which penalties for inconsistent reports are imposed only on S. These penalties are large enough to deter both individual deviations and joint deviations involving 'collusion'. Such penalties could conceivably create additional scope for A to extort from S relative to the original mechanism. The purpose of providing S an additional 'non-type' message option is to provide her with 'protection' against this possibility. Specifically, S is guaranteed the same minmax payoff as in the original mechanism whenever she selects this non-type message option, irrespective of whatever message is sent by A.

The preceding remarks provide an informal account of the main complications resulting from the presence of extortion in establishing the CPP, and the main idea underlying the way we address these problems. Many additional details need to be provided to make these more concrete and precise. We turn to these details next.

An important first step is to characterize the scope for extortion in any given mechanism, by defining the outside-option payoffs of S and A in coalitional bargaining in any given mechanism. For S, this is given by the minmax payoff $\underline{w}_S(GC)$. The minimax theorem cannot, however, be applied to A's payoffs owing to the presence of private information of A regarding the realization of θ. Given grand contract GC, for a report of S: $\mu_S \in \Delta(M_S)$ and a type of A: $\theta \in \Theta$, define the maximum payoff that A can attain in response:

$$\hat{u}_A(\theta, \mu_S, GC) \equiv \max_{\mu_A \in \Delta(M_A)} \bar{X}_A(\mu_A, \mu_S) - \theta \bar{q}(\mu_A, \mu_S)$$

and $\mu_A(\theta, \mu_S, GC)$ as a maximizer of the above problem. These payoffs obviously depend on the realization of θ, and on the specific 'threat' μ_S that may be chosen by S (which is independent of θ).

Take a given mechanism GC, and consider the side contract design problem corresponding to a specific allocation (X, u_A, u_S, q) which is achieved in extortionary collusion in GC. Let the associated equilibrium reporting strategy be $\mu(\theta, \eta) \in \Delta(M)$, so that $q(\theta, \eta) = \hat{q}(\mu(\theta, \eta))$ and $X(\theta, \eta) = \hat{X}_A(\mu(\theta, \eta)) + \hat{X}_S(\mu(\theta, \eta))$. It can be shown (see Lemma A1 in the Appendix) that the third party designing the side contract can devise a cross-checking mechanism to elicit the realization of η. Hence the side-contract design problem is 'as if' the third party designer observes the realization of η, and can be simplified as follows. For each state η, solve the following problem (hereafter denoted $P^E(\alpha : \eta, GC)$): choose

$$(\tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta), \tilde{\mu}(\theta, \eta), \tilde{\mu}_S(\eta))$$

to maximize

$$E[\alpha \tilde{u}_A(\theta, \eta) + (1 - \alpha)\tilde{u}_S(\theta, \eta) \mid \eta]$$

subject to the constraint that for all $\theta \in \Theta$:

(i) $\tilde{u}_A(\theta, \eta) \in \Re, \tilde{u}_S(\theta, \eta) \in \Re, \tilde{\mu}(\theta, \eta) \in \Delta(M_A \times M_S), \tilde{\mu}_S(\eta) \in \Delta(M_S)$
(ii) $\tilde{u}_A(\theta, \eta) \geq \tilde{u}_A(\theta', \eta) + (\theta' - \theta)\hat{q}(\tilde{\mu}(\theta', \eta))$ for any $\theta' \in \Theta$
(iii) $\tilde{u}_A(\theta, \eta) + \tilde{u}_S(\theta, \eta) \leq \hat{X}_A(\tilde{\mu}(\theta, \eta)) + \hat{X}_S(\tilde{\mu}(\theta, \eta)) - \theta\hat{q}(\tilde{\mu}(\theta, \eta))$
(iv) $\tilde{u}_A(\theta, \eta) \geq \hat{u}_A(\theta, \tilde{\mu}_S(\eta), GC)$ and
(v) $E[\tilde{u}_S(\theta, \eta) \mid \eta] \geq \underline{w}_S(GC)$.

The side contract in state (θ, η) generates payoffs $\tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta)$ for A and S respectively. It stipulates a joint deviation to a report $\tilde{\mu}(\theta, \eta)$, and is supported by threats $\tilde{\mu}_S(\eta)$ and $\underline{\mu}_A$ by S and A respectively in the event that the other agent refuses to participate in it. These threats determine outside payoffs of A and S reflected in the right-hand sides of participation constraints (iv) and (v) respectively. Owing to the exclusive private information of A over the realization of θ, the side contract must satisfy the constraint (ii) that A reports truthfully. Moreover, the side contract is sustained by a suitable side payment which is implicitly reflected in (iii), a budget balance condition within the coalition.

Lemma A1 in the Appendix verifies that if (X, u_A, u_S, q) is achieved in extortionary collusion in GC, then $(\tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta), \tilde{\mu}(\theta, \eta)) = (u_A(\theta, \eta), u_S(\theta, \eta), \mu(\theta, \eta))$ solves the side contract problem $P^E(\alpha : \eta, GC)$ for each η. In other words there does not exist some side contract satisfying (i)-(v) that deviates from the allocation and achieves a higher α-welfare-weighted sum of expected payoffs for A and S. The Lemma also establishes that the coalition's budget is balanced, i.e., (iii) holds as an equality: $X(\theta, \eta) - \theta q(\theta, \eta) = u_A(\theta, \eta) + u_S(\theta, \eta)$. As outside options in (iv) and (v) are non-negative, an achievable allocation satisfies standard individual incentive and participation constraints:

(a) (IC_A): truthful telling condition for A

$$u_A(\theta, \eta) \geq u_A(\theta', \eta) + (\theta' - \theta)q(\theta', \eta)$$

for any $\theta, \theta' \in \Theta$ and any $\eta \in \Pi$;

(b) (PC_A): interim participation constraint for A

$$u_A(\theta, \eta) \geq 0$$

for any $(\theta, \eta) \in K$; and

(c) (PC_S): interim participation constraint for S

$$E[u_S(\theta,\eta) \mid \eta] \geq 0$$

for any $\eta \in \Pi$.

We say that (u_A, u_S, q) satisfies *individual incentive compatibility* (IIC) if and only if it satisfies IC_A, PC_A and PC_S.

Now we describe the problems that arise in establishing CPP in the presence of extortion. CPP requires that every achievable allocation in an arbitrary mechanism GC is also achievable in a suitable revelation mechanism where A and S agree to participate, report truthfully and not deviate to any non-null side contract. We represent a revelation mechanism where participation decisions are included as part of the message space of each agent, with the constraint that both agents are offered a null contract if either of them decides not to participate. Hence the message spaces are as follows: $M_A^R = \{e_A\} \cup \Pi \times \Theta$ and $M_S^R = \{e_S\} \cup \Pi$. Let the outcome function be denoted by $(X_A^R(m_A, m_S), X_S^R(m_A, m_S), q^R(m_A, m_S)) \in \mathcal{GC}$. Then $(X_A^R, X_S^R, q^R) = (0, 0, 0)$ if either $m_A = e_A$ or $m_S = e_S$. Moreover, in any state (θ, η), the requirement that truthful reporting results in the allocation (u_A, u_S, q) requires $(X_A^R, X_S^R, q^R) = (u_A(\theta,\eta) + \theta_A q(\theta,\eta), u_S(\theta,\eta), q(\theta,\eta))$ when $m_A = (\theta, \eta), m_S = \eta$. It remains to specify the 'cross-checking mechanism', i.e., outcomes when the reports submitted by the agents are inconsistent, i.e., $\eta_A \neq \eta_S$. In the absence of extortion, inconsistent reports result in the imposition of penalties on either or both agents, that are large enough to deter individual and joint deviations. For instance, suppose the penalties are imposed on S alone, as in the following revelation mechanism (denoted by GC^R):

$$(X_A^R, X_S^R, q^R) = (u_A(\theta_A, \eta_S) + \theta_A q(\theta_A, \eta_S), u_S(\theta_A, \eta_A) - L(\eta_A, \eta_S), q(\theta_A, \eta_S))$$

for messages $((\theta_A, \eta_A), \eta_S) \in M_A^R \times M_S^R$, while $(X_A^R, X_S^R, q^R) = (0, 0, 0)$ if either $m_A = e_A$ or $m_S = e_S$, where the penalty imposed on S for inconsistent reports satisfies $L(\eta_A, \eta_S) = 0$ when $\eta_A = \eta_S$ and a large positive number for $\eta_A \neq \eta_S$. Panel (a) in Fig. 2 outlines this revelation mechanism GC^R where elements in the matrix denote (X_A^R, X_S^R, q^R) for different combinations of S and A's messages.

Suppose that GC^R is played by A and S non-cooperatively. Truthful reporting of η is guaranteed by the property of GC^R that cross-checks reports of η between A and S, imposing a penalty L on S when these do not match.[18] IIC ensures that A does not benefit from misreporting θ, and both parties have an incentive to participate in the mechanism. Thus, allocation (u_A, u_S, q) is achieved on the continuation game following a null side-contract between S and A.

Next, CPP requires the third party side-contract designer to not want to deviate to a non-null side contract. This requires the scope for collusion and extortion in GC^R should not be larger when compared to the original GC.

Note first that the scope for collusion does not expand if the reports are inconsistent $(\eta_A \neq \eta_S)$, since this causes them to collectively lose $L(\eta_A, \eta_S)$ in transfers from P. Restricting attention to consistent reports that satisfy $\eta_A = \eta_S$, for any message choice in GC^R there exists a message in GC which induces the same allocation. Thus the change from GC to GC^R does not expand the scope for collusion.

However, large penalties imposed on S could expand the scope for A to extort from S. For instance, suppose that S's minmax payoff in GC is positive, and a constant penalty $L(\eta_A, \eta_S) =$

[18] Since A's payoff does not depend on η_A, A does not have an incentive to misreport η. And given truthful reporting of η by A, S does not have an incentive to misreport owing to the penalty imposed on S for inconsistent reports.

(a)

$S \setminus A$	e_A	$(\theta_A,\eta_A)\in K$
e_S	$(0,0,0)$	$(0,0,0)$
$\eta_S \in \Pi$	$(0,0,0)$	$(u_A(\theta_A,\eta_S)+\theta_A\, q(\theta_A,\eta_S), u_S(\theta_A,\eta_A)-L(\eta_A,\eta_S), q(\theta_A,\eta_S))$

(b)

$S \setminus A$	e_A	$(\theta_A,\eta_A)\in K$
e_S	$(0,0,0)$	$(0,0,0)$
η_0	$(0,\omega,0)$	$(0,\omega,0)$
$\eta_S \in \Pi$	$(0,0,0)$	$(u_A(\theta_A,\eta_S)+\theta_A\, q(\theta_A,\eta_S), u_S(\theta_A,\eta_A)-L(\eta_A,\eta_S), q(\theta_A,\eta_S))$

(c)

$S \setminus A$	e_A	$(\theta_A,\eta_A)\in K$
e_S	$(0,0,0)$	$(\bar{X}_A(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S),0,\bar{q}(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S))$
η_0	$(0,\omega,0)$	$(\bar{X}_A(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S),\omega,\bar{q}(\mu_A(\theta_A,\bar{\mu}_S,GC),\bar{\mu}_S))$
$\eta_S \in \Pi$	$(0,0,0)$	$(u_A(\theta_A,\eta_S)+\theta_A\, q(\theta_A,\eta_S), u_S(\theta_A,\eta_A)-L(\eta_A,\eta_S), q(\theta_A,\eta_S))$

Fig. 2. Revelation Mechanism GC^R and Its Modification.

$\bar{L} > 0$ is imposed on S whenever $\eta_A \neq \eta_S$. If $\bar{L}$ is sufficiently large, the minmax payoff of S in GC^R would fall below the minmax payoff of S in GC. If A has all the bargaining power in the coalition ($\alpha = 1$), the side-contract designer would want to deviate to a non-null side contract which pushes S to a lower payoff than in the allocation achievable in GC. This would be sustained by a threat of A reverting to the reporting strategy that minmaxes S's payoff if S should refuse the side contract.

This illustrates the problem that extortion creates in designing a suitable cross-checking mechanism with large penalties for inconsistent reports. Could this problem be avoided by imposing small or no penalties at all for inconsistent reports? The answer is no, for the following reason. Even if $L(\eta_A, \eta_S) = 0$ for all (η_A, η_S), the maxmin payoff of S in GC^R could be smaller than that in the original mechanism GC because S may have some message option in GC that she actually does not use on the equilibrium path, but it receives a positive weight in her maxmin strategy.[19] Since (conditional on agreeing to participate), S's message space in GC^R consists only of messages that mimic 'equilibrium path' messages in the original mechanism, S's maxmin payoff in GC^R could fall below that in the original mechanism GC, rendering her more vulnerable to extortion.

Nevertheless, this problem can be avoided if we augment GC^R by providing S with an auxiliary 'non-type' report option which guarantees her the same minmax payoff as in the original mechanism. Using such an augmentation of GC^R, we can show that CPP holds. We outline the argument below.

The following two conditions (a) and (b) suffice to ensure that the scope for extortion does not expand:

[19] Of course other problems may also arise if inconsistent reports result in no penalties: e.g., the side contract designer may be tempted to select a reporting strategy with inconsistent reports.

(a) $\underline{w}_S(GC^R) \geq \underline{w}_S(GC)$
(b) For any $\mu_S \in \Delta(M_S^R)$, there exists $\mu'_S \in \Delta(M_S)$ such that $\hat{u}_A(\theta, \mu_S, GC^R) \geq \hat{u}_A(\theta, \mu'_S, GC)$ for all $\theta \in \Theta$.

Conditions (a) and (b) imply that outside option payoffs are not any lower in the augmented mechanism compared to GC, for S and A respectively. We now explain how the augmented revelation mechanism can be designed to ensure these two conditions.

Hereafter, let $\underline{w}_S(GC)$ be denoted by ω to simplify the exposition. To ensure (a), we provide S with the auxiliary message option η_0. Defining $\bar{\Pi} \equiv \Pi \cup \{\eta_0\}$, S's message space is modified to $M_S^R = \{e_S\} \cup \bar{\Pi}$. For the moment, let us select (X_A^R, X_S^R, q^R) equal to $(0, \omega, 0)$ for any (m_A, m_S) such that $m_S = \eta_0$. With this modified GC^R, the minmax value of S's payoff is equal to ω, ensuring condition (a). This modified GC^R is illustrated in Fig. 2(b).

Observe that

$$(X_A^R(m_A, e_S), q^R(m_A, e_S)) = (X_A^R(m_A, \eta_0), q^R(m_A, \eta_0)) = (0, 0)$$

in this modified GC^R, i.e., choice of either e_S or η_0 by S pushes A's outside payoff down to zero. This may cause condition (b) to be violated, unless $\hat{u}_A(\theta, \mu'_S, GC) = 0$ for any $\theta \in \Theta$ and for some $\mu'_S \in \Delta(M_S)$ in the original GC. It therefore necessitates an additional modification of GC^R in order to protect A from the use of such a threat by S. Hence we modify $(X_A^R((\theta_A, \eta_A), e_S), q^R((\theta_A, \eta_A), e_S))$ and $(X_A^R((\theta_A, \eta_A), \eta_0), q^R((\theta_A, \eta_A), \eta_0))$ from $(0, 0)$ to

$$(\bar{X}_A(\mu_A(\theta_A, \bar{\mu}_S, GC), \bar{\mu}_S), \bar{q}(\mu_A(\theta_A, \bar{\mu}_S, GC), \bar{\mu}_S))$$

where it may be recalled $\mu_A(\theta_A, \bar{\mu}_S, GC)$ denotes the best response of type θ of A to S's maxmin report $\bar{\mu}_S$ in GC. The resulting mechanism is illustrated in Fig. 2(c).

With these two modifications, GC^R satisfies all the required conditions for CPP to hold. We sketch the argument for the benefit of interested readers; others can skip directly to the definition of ECP allocations below. It is evident that these modifications do not affect the non-cooperative equilibrium of the mechanism, since Lemma A1 implies $E[u_S(\theta, \eta) \mid \eta] \geq \omega$ and S does not benefit from selecting η_0 when A reports truthfully. Moreover, since the minimax theorem implies

$$\bar{X}_S(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) \geq \bar{X}_S(\underline{\mu}_A, \bar{\mu}_S) = \omega \geq 0$$

for any θ, a report of $(m_A, m_S) = ((\theta, \eta), e_S)$ or $((\theta, \eta), \eta_0)$ in this modified GC^R does not result in a higher total payment $\hat{X}_A^R + \hat{X}_S^R$, as compared to the report of $(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S)$ in the original GC. Hence this modification of the mechanism does not expand the scope for collusion, relative to GC.

Next, we claim that condition (b) also holds. Since the use of e_S and η_0 as a threat in GC^R has the same impact on A's payoff, it suffices to consider what can be achieved in terms of lowering A's participation constraint in side contracting by supplementing S's message options by η_0 alone. Let any such threat be represented by a probability function $P(\cdot)$ defined over $\bar{\Pi}$, i.e., where $P(\eta)$ is the probability that S reports $\eta \in \bar{\Pi}$.

Now observe that conditional on S reporting $\eta_S = \eta \in \Pi$, type θ of A ends up with the same payoff that she would have attained in the original GC upon playing a best response to $\mu_S(\eta)$. And conditional on S reporting $\eta_S = \eta_0$, she would end up with the same payoff as by playing a best response to $\bar{\mu}_S$ in GC. Since A's payoff does not depend on her own report of η_A, and reporting θ truthfully is a best response to any report $\eta_S \in \Pi$, it follows that reporting truthfully is a dominant strategy for A in GC^R given any threat $P(.)$. It follows that the expected payoff that A attains in GC^R when S executes his threat represented by $P(.)$, is at least as large as what she would have attained in the original GC if S reported according to a mixed strategy which corresponds to the following two stage lottery: (1) at the first stage η_S is chosen according to $P(\cdot)$ over $\bar{\Pi}$, and then (2) at the second stage S reports according to $\mu_S(\eta)$ if at the first stage $\eta_S = \eta \in \Pi$, and according to $\bar{\mu}_S$ if instead $\eta_S = \eta_0$. Hence (b) holds, establishing the validity of CPP.

In the preceding argument, the mechanism augmented S's message space to include η_0 in order to protect S from additional scope of extortion relative to the original GC. No such augmentation was needed for A's message space, owing to the absence of any penalty on A for inconsistencies between A and S's reports in the underlying revelation mechanism. As explained above, A can protect herself from the possibility of any additional extortion, by reporting truthfully in response to any threat $P(.)$.

We conclude this section with a formal statement of the CPP. For this we use the following definition of an extortionary collusion-proof (ECP) allocation. In words, this refers to an allocation that is achievable in a revelation mechanism which is augmented in the manner described above, utilizing strategies in which A and S always agree to participate and report truthfully, and in response to which it is optimal for the coalition to select a null side-contract.

Definition 2. An allocation (u_A, u_S, q) is extortionary collusion-proof (or ECP) for $\alpha \in [0, 1]$, if (u_A, u_S, q) is IIC, there exists $\omega \geq 0$, and an augmentation of (u_A, u_S, q) on $\bar{K} \equiv \Theta \times \bar{\Pi}$ with $u_S(\theta, \eta_0) = \omega$ for any $\theta \in \Theta$ and $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ satisfying (IC_A) and (PC_A), such that for any $\eta \in \Pi$,

$$(\mu(\theta, \eta), \tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta), P(\cdot \mid \eta)) = (I_1(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), I_2(\eta))$$

solves problem $P^E(\alpha : \eta)$:

$$\max E[\alpha \tilde{u}_A(\theta, \eta) + (1 - \alpha)\tilde{u}_S(\theta, \eta) \mid \eta]$$

subject to $(\mu(\theta, \eta), \tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta), P(\cdot \mid \eta))$ satisfies for all $\theta \in \Theta$:

(i) $\mu(\theta, \eta) \in \Delta(\bar{K} \cup \{e\}), \tilde{u}_A(\theta, \eta) \in \Re, \tilde{u}_S(\theta, \eta) \in \Re, P(.|\eta) \in \Delta(\bar{\Pi})$
(ii) $\tilde{u}_A(\theta, \eta) \geq \tilde{u}_A(\theta', \eta) + (\theta' - \theta)q(\mu(\theta', \eta))$ for any $\theta' \in \Theta$
(iii) $\tilde{u}_A(\theta, \eta) + \tilde{u}_S(\theta, \eta) \leq X(\mu(\theta, \eta)) - \theta q(\mu(\theta, \eta))$
(iv) $\tilde{u}_A(\theta, \eta) \geq \Sigma_{\eta' \in \bar{\Pi}} P(\eta' \mid \eta) u_A(\theta, \eta')$
(v) $E[\tilde{u}_S(., \eta) \mid \eta] \geq \omega$,

where $X(\theta, \eta) \equiv u_A(\theta, \eta) + u_S(\theta, \eta) + \theta q(\theta, \eta)$ for $(\theta, \eta) \in \bar{K}$ and $(X(e), q(e)) \equiv (0, 0)$, and $I_1(\theta, \eta)$ denotes a probability measure on $\bar{K} \cup \{e\}$ where (θ, η) is selected with probability one, while $I_2(\eta)$ denotes a probability function on $\bar{\Pi}$ where η is selected with probability one (i.e., $P(\eta \mid \eta) = 1$ and $P(\eta' \mid \eta) = 0$ for any $\eta \neq \eta'$).

$P^E(\alpha : \eta)$ refers to the side-contracting problem faced by the coalition in response to the augmented revelation mechanism. A feasible side-contract corresponding to η generates payoffs $\tilde{u}_A(\theta, \eta), \tilde{u}_S(\theta, \eta)$, and is supported by reporting strategies $\mu(\theta, \eta)$ and S's threat to report according to the probability distribution $P(\cdot \mid \eta)$ if A should refuse the side-contract. S's outside option payoff is set equal to ω, the payoff that S is guaranteed by the auxiliary message option η_0. This obviates the need to specify the reporting threat assigned to A in the event that S refuses the side-contract.

Note that an ECP allocation is described by properties that do not depend on the original GC. Moreover, for any ECP allocation satisfying Definition 2, we can easily construct a (modified) revelation mechanism GC^R as described above by adding the cross-checking scheme about reports of η by A and S, and then augmenting S's message space. This leads us to the formal statement of CPP; the proof is provided in the Appendix.

Proposition 1. ***(Collusion Proof Principle (CPP))***

An allocation (u_A, u_S, q) is achievable in extortionary collusion with α if and only if it is extortionary collusion-proof (ECP) for α.

It is useful to contrast ECP allocations with allocations achievable with weak collusion (see e.g., Mookherjee et al. (2020)) where extortion is absent. In the weak collusion setting, failure of A and S to agree on a non-null side contract always triggers noncooperative play of GC. Let A and S's noncooperative payoffs be denoted respectively by $u_A^N(\cdot, GC)$ and $u_S^N(\cdot, GC)$. These serve as outside options in the bargaining over the side-contract. Hence the characterization of allocations achievable in weak collusion differs with respect to the right-hand sides of constraints (iv) and (v) in problem $P^E(\alpha : \eta, GC)$: these are replaced respectively by $u_A^N(\theta, \eta, GC)$ and $E[u_S^N(\theta, \eta, GC) \mid \eta]$. Moreover, there is no need to augment message spaces in weak collusion, owing to the absence of extortion.

5. Effects of varying bargaining power

As explained in the Introduction, earlier studies have established that a set of achievable allocation under weak collusion does not depend on α. The main result of this paper is that the allocation of bargaining power does matter in extortionary collusion. In particular, there is a sharp dichotomy between contexts where S has greater ($\alpha < 1/2$) and lower ($\alpha \geq 1/2$) bargaining power than A.

5.1. When S has less bargaining power than A

We begin with the case where S has lower bargaining power, and the side contract is designed to maximize the scope for A to extort from S. In the following, we use NS to denote a setting where S is not hired, and P contracts with A alone.

Proposition 2. *If $\alpha \geq 1/2$, P does not benefit from hiring S, i.e., cannot attain a higher welfare than in NS where S is not hired.*

To explain the intuition underlying this result, it is helpful to first consider the case where $\alpha = 1$, i.e., the optimal side contract is chosen to maximize the payoff of A, subject to S's participation constraint — it is as if A is the 'sub-principal' within the (A-S) coalition. This side contract problem is not subject to any information friction, since A is better informed than S. The only constraint that effectively matters is S's participation constraint, i.e., that S must be assured a payoff of at least ω. The optimal side contract from A's perspective is then one in which S is provided a lumpsum payment of ω, all the residual rents accrue to A, and a joint reporting strategy is selected which maximizes A's rents. The same logic applies when $\alpha \in (\frac{1}{2}, 1)$, since a transfer of surplus from S to A still raises the objective of the side-contract designer. Hence it is optimal for the side contract to be designed so that the reporting strategy maximizes the sum of A and S's payoffs, i.e., the coalition will achieve efficient collusion:

$$\mu^*(\theta) \in \arg\max_{\mu \in \Delta(\bar{K} \cup \{e\})} [X(\mu) - \theta q(\mu)] \tag{1}$$

This reporting strategy is evidently independent of the actual realization of η, implying that S merely acts as a 'rubberstamp' and submits the same audit report irrespective of his true signal which best serves A's interest. Hence the presence of S adds no value to the principal P.

The detailed argument (presented in the Appendix) also needs to check that efficient collusion is feasible in the side contracting problem. Here is an outline of the argument. Feasibility requires the existence of threats which ensure that S and A's participation constraints are satisfied in the side contracting problem when efficient collusion is chosen. Given any grand contract GC, let $\mu^{GC}(\theta)$ denote the reporting strategy μ that maximizes $[\hat{X}_A(\mu) + \hat{X}_S(\mu) - \theta\hat{q}(\mu)]$ subject to $\mu \in \Delta(M_A \times M_S)$. Then observe that using threats $\underline{\mu}_A$ by S and $\bar{\mu}_S$ by A respectively (if the other rejects the side-contract), the sum of their respective payoffs is at least as large as the sum of their corresponding outside options:

$$\begin{aligned}
&\hat{X}_A(\mu^{GC}(\theta)) + \hat{X}_S(\mu^{GC}(\theta)) - \theta\hat{q}(\mu^{GC}(\theta)) \\
&\geq \hat{X}_A(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) + \hat{X}_S(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) - \theta\hat{q}(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) \\
&\geq \hat{X}_A(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) + \hat{X}_S(\underline{\mu}_A, \bar{\mu}_S) - \theta\hat{q}(\mu_A(\theta, \bar{\mu}_S, GC), \bar{\mu}_S) \\
&= \hat{u}_A(\theta, \bar{\mu}_S, GC) + \underline{w}_S(GC)
\end{aligned}$$

The first inequality follows from the definition of $\mu^{GC}(\theta)$, while the second inequality is the result of the minimax theorem. Hence with appropriate side-transfers, efficient collusion can be sustained by the specified threats.

5.2. *When S has greater bargaining power*

Now consider the opposite extreme, where S has all the bargaining power ($\alpha = 0$), i.e., S becomes the 'sub-principal' within the coalition. As S is less informed than A, the side contracting problem is subject to asymmetric information frictions. As a consequence, the coalition does not select an allocation with efficient collusion, giving rise to the conjecture that the resulting weakening of collusion would benefit P. However, it is not easy to verify this at the current level of generality. The next subsection examines this question in an example with a specific technology and information structure.

One result which nevertheless can be established quite generally provides an indication of how optimal side contracts can vary as the bargaining power allocation shifts in favor of S. Recall that when A has more bargaining power, side contracts achieve efficient collusion and S's participation constraint is binding, with all residual rents accruing to A. We now show that when S has greater bargaining power, S's participation constraint will not bind and can therefore be dropped in the formulation of the side contracting problem.

Lemma 1. *If $\alpha \in [0, 1/2)$, ECP allocation (u_A, u_S, q) solves the relaxed version of $P^E(\alpha : \eta)$ where S's participation constraint $E[\tilde{u}_S(\theta, \eta) \mid \eta] \geq \omega$ is dropped.*

The reasoning is as follows. If this lemma were false, the solution to the relaxed version of problem $P^E(\alpha : \eta)$ must violate S's participation constraint, implying that S ends up with an expected payoff below his minmax payoff ω. The coalition then has the option of switching to the side-contract (used in the proof of Proposition 2) in which S receives a constant payoff of ω and A receives the rest of the aggregate coalitional rent. This side-contract induces ex post efficient reporting strategies, thereby (weakly) expanding the aggregate rent in every state. Given $\alpha < \frac{1}{2}$, A must also benefit from deviating to this side-contract. But this leads to a contradiction, as S and A both benefit from this deviation.

5.3. Example: procuring an indivisible good

In this subsection we consider a specific example involving procurement of an indivisible good, and show that hiring S can be valuable when she has greater bargaining power.

The quantity to be procured is $q \in Q \equiv \{0, 1\}$. P obtains a zero gross benefit if $q = 0$, and a benefit of $V > 0$ otherwise. There are two possible signal realizations $\eta_i, i = L, H$; $a_i(\theta) \equiv a(\eta_i \mid \theta) > 0$ on Θ for $i \in \{L, H\}$, and $a_H(\theta)$ is increasing in θ, i.e., H is more likely to occur for higher values of θ. $F_i(\theta) \equiv F(\theta \mid \eta_i)$ denotes the distribution of θ conditional on signal realization i, which is assumed to have a positive density $f_i(\theta) \equiv f(\theta \mid \eta_i)$ for all $\theta \in \Theta$. Let $\kappa_i \equiv p(\eta_i) = \int_{\underline{\theta}}^{\bar{\theta}} f(\theta)a_i(\theta)d\theta \in (0, 1)$ denote the ex ante probability of signal i. These assumptions imply $F_L(\theta) > F(\theta) > F_H(\theta)$ and $F_L(\theta)/f_L(\theta) > F(\theta)/f(\theta) > F_H(\theta)/f_H(\theta)$ on $(\underline{\theta}, \bar{\theta})$. In addition, assume that (i) $h_i(\theta) = \theta + \frac{F_i(\theta)}{f_i(\theta)}$ and $l_i(\theta) = \theta + \frac{F_i(\theta)-1}{f_i(\theta)}$ are increasing in θ for each $i = L, H$, (ii) $\theta + \frac{F(\theta)}{f(\theta)}$ is increasing in θ, and (iii) $V \in (\underline{\theta}, \bar{\theta} + \frac{1}{f(\bar{\theta})})$.

In the *No Supervision* (NS) context, P offers a non-contingent price p^{NS} to maximize $F(p)[V - p]$. Let $W^{NS} \equiv F(p^{NS})[V - p^{NS}]$ denote the resulting expected payoff of P. Assumption (iii) above guarantees an interior solution $p^{NS} \in (\underline{\theta}, \bar{\theta})$ and $W^{NS} > 0$ in the optimal NS contract. In the *second-best* context, P can costlessly access S's signal, eliminating any scope for collusion between A and S. Here for each $i \in \{L, H\}$, P offers A a price p_i^{SB} which maximizes $(V - p_i)F_i(p_i)$ subject to $p_i \in \Theta$. Then (iii) also implies $p_H^{SB} > p^{NS} > p_L^{SB}$.

With binary output $Q = \{0, 1\}$, it may be beneficial to randomize output assignments. However, we restrict attention to deterministic mechanisms, i.e., where the output assignment in the direct revelation mechanism corresponding to the GC is deterministic following any message combination. Our main purpose is to show that hiring S can be valuable when $\alpha < 1/2$, and it suffices to show this using deterministic mechanisms. Our earlier working paper (Mookherjee and Tsumagari (2022)) studies the corresponding problem of finding the optimal deterministic mechanism.

5.3.1. ECP allocations

We now use the CPP to characterize deterministic ECP allocations in this setting with $0 \leq \alpha < 1/2$. Consider an ECP allocation (u_A, u_S, q) with deterministic output $q(\theta, \eta) \in \{0, 1\}$ for any $(\theta, \eta) \in K$. Since this allocation satisfies IC_A and PC_A, there exists $p_i \in \Theta$ such that $q(\theta, \eta_i) = 1$ on $[\underline{\theta}, p_i)$ and 0 on $(p_i, \bar{\theta}]$ for any $i \in \{L, H\}$, resulting in $u_A(\theta, \eta_i) = \max\{p_i - \theta, 0\} + u_i$ with $u_i \geq 0$. Applying Proposition 1, we augment the mechanism by providing S an additional message option η_0, where $(u_A(\theta, \eta_0), q(\theta, \eta_0))$ satisfies (IC_A) and (PC_A), $q(\theta, \eta_0) \in \{0, 1\}$, and $u_S(\theta, \eta_0) = \omega \geq 0$. This implies the existence of $p_0 \in \Theta$ and $u_0 \geq 0$ such that $u_A(\theta, \eta_0) = \max\{p_0 - \theta, 0\} + u_0$.

Given an augmented allocation (u_A, u_S, q) defined on the domain $\bar{K}$, consider problem $P^E(\alpha : \eta)$ for $\eta \in \{\eta_L, \eta_H\}$, with $\alpha < 1/2$. Lemma 1 implies S's participation constraint (v) can be dropped: there is no need to protect S from A's use of minimax strategies as a threat. It is easy to check that ω can be set to be equal to zero without loss of generality. However, the auxiliary message η_0, can be used as part of S's strategy to threaten A.

Next, observe that collusion-proofness implies that the aggregate payment to the coalition cannot vary across message combinations that result in the same output assignment. Let X_0 and $X_1 \equiv X_0 + b$ denote the aggregate payment corresponding to output $q = 0$ and 1 respectively, so X_0 is interpreted as a lumpsum payment and b a bonus for delivering $q = 1$. X_0 must be non-negative, otherwise A and S would both benefit from a joint decision to exit in states with $q = 0$. Let $\tilde{q}(\theta, \eta) \in [0, 1]$ denote $q(\mu(\theta, \eta))$: this is the probability of being assigned to deliver the output in state (θ, η) upon using the reporting strategy $\mu(\theta, \eta)$. The coalition's choice of a (mixed) reporting strategy $\mu(\theta, \eta)$ reduces to choosing a lottery between the two outcomes $(X, q) = (X_0 + b, 1)$ and $(X, q) = (X_0, 0)$. A's incentive constraint (ii) within the coalition then implies the following two constraints must hold:

$$\tilde{u}_A(\theta, \eta) = \int_{\theta}^{\bar{\theta}} \tilde{q}(y, \eta) dy + \tilde{u}_A(\bar{\theta}, \eta) \tag{a}$$

$$\tilde{q}(\theta, \eta) \text{ is non-increasing in } \theta. \tag{b}$$

Moreover, when S selects η_0, S is paid nothing, implying that the aggregate payment to the coalition equals the payment made to A: i.e., $X(\theta, \eta_0) = p_0 + u_0$ if $q(\theta, \eta_0) = 1$ and u_0 if $q(\theta, \eta_0) = 0$. Collusion-proofness requires the coalition to never benefit from S reporting η_0, which implies the following two constraints must hold:

$$X_0 + b \geq p_0 + u_0 \tag{$CP1$}$$

and

$$X_0 \geq u_0. \tag{$CP2$}$$

Observe that as $u_0 \geq 0$, (CP2) implies the requirement $X_0 \geq 0$ is automatically satisfied and can therefore be dropped.

Define $\beta \equiv (1 - 2\alpha)/(1 - \alpha) \in (0, 1]$. Since constraint (iii) is satisfied with equality in the solution, there is no loss of generality in rewriting it as $\tilde{u}_S(\theta, \eta) + \tilde{u}_A(\theta, \eta) = \tilde{q}(\theta, \eta)(b - \theta) + X_0$. Therefore, the side-contract problem $P^E(\alpha : \eta_i), i \in \{L, H\}$ reduces to the following problem (hereafter denoted $\bar{P}^E(\alpha : \eta_i)$):[20]

$$\max E[\tilde{q}(\theta, \eta_i)(b - \theta) + X_0 - \beta \tilde{u}_A(\theta, \eta_i) \mid \eta_i]$$

subject to (a), (b), and

$$\tilde{u}_A(\theta, \eta_i) \geq \Sigma_{j \in \{L,H,0\}} P(\eta_j \mid \eta_i)[\max\{p_j - \theta, 0\} + u_j]. \tag{2}$$

For the allocation to be ECP, we need $\tilde{q}(\theta, \eta_i) = q(\theta, \eta_i)$, $\tilde{u}_A(\theta, \eta_i) = u_A(\theta, \eta_i)$ and $P(\cdot \mid \eta_i) = I_2(\eta_i)$ to solve $\bar{P}^E(\alpha : \eta_i)$. Let this property be called (CP3).

---

[20] Note that the third party's payoff in η_i is

$$E[(1 - \alpha)\tilde{u}_S(\theta, \eta_i) + \alpha \tilde{u}_A(\theta, \eta_i) \mid \eta_i] = (1 - \alpha)E[\tilde{u}_S(\theta, \eta_i) + \tilde{u}_A(\theta, \eta_i) - \beta \tilde{u}_A(\theta, \eta_i) \mid \eta_i]$$
$$= (1 - \alpha)E[\tilde{q}(\theta, \eta)(b - \theta) + X_0 - \beta \tilde{u}_A(\theta, \eta_i) \mid \eta_i].$$

To summarize, an allocation is described by a tuple $(b, X_0, p_H, p_L, p_0, u_H, u_L, u_0)$. This allocation is ECP if and only if it satisfies (CP1), (CP2), (CP3) and $u_j \geq 0$ for $j \in \{L, H, 0\}$. The associated output assignment is $q = 1$ if $\theta \leq p_i$ and 0 otherwise. A receives a bonus of p_i for delivering the output, while S receives $b - p_i$. It results in payoffs for A and S: $u_A(\theta, \eta_i) = \max\{p_i - \theta, 0\} + u_i$, while $u_S(\theta, \eta_i) = b - p_i + X_0 - u_i$ for $\theta \leq p_i$ and $X_0 - u_i$ for $\theta > p_i$, for $i \in \{L, H\}$. P's resulting ex-ante payoff equals

$$(\kappa_L F_L(p_L) + \kappa_H F_H(p_H))(V - b) - X_0. \tag{3}$$

5.3.2. Value of supervisor

P's problem of finding an optimal deterministic mechanism reduces to maximizing (3) over the set of ECP allocations. Here we focus only on the question regarding the existence of ECP allocations that help P achieve a higher payoff than W^{NS}.

We start with the optimal mechanism in NS, which corresponds to $b = p_L = p_H = p_0 = p^{NS}$ and $X_0 = u_H = u_L = u_0 = 0$. Since S is not involved in this mechanism there is obviously no scope for collusion, so this allocation is ECP. Consider a small variation on this mechanism which involves raising p_H (the price offered to A for delivering the good, following a report of $\eta_S = \eta_H$) above p^{NS}, while all other prices are left unchanged: $b = p_L = p_0 = p^{NS}$. The following lemma shows that a small rise in p_H ensures the allocation remains ECP.

Lemma 2. *For p_H larger than but sufficiently close to p^{NS}, an allocation characterized by $b = p_L = p_0 = p^{NS}$, $u_H = 0$, and*

$$X_0 = u_L = u_0 = \frac{F_H(p_H)(p_H - p^{NS}) - (1-\beta)\int_{p^{NS}}^{p_H} F_H(\theta)d\theta}{\beta} > 0 \tag{4}$$

is ECP for $\alpha \in [0, 1/2)$. It results in an ex ante payoff for P of

$$(\kappa_L F_L(p^{NS}) + \kappa_H F_H(p_H))(V - p^{NS}) - \frac{F_H(p_H)(p_H - p^{NS}) - (1-\beta)\int_{p^{NS}}^{p_H} F_H(\theta)d\theta}{\beta} \tag{5}$$

As p_H is raised slightly above p^{NS}, it necessitates a corresponding (equal) rise in u_L, u_0 and X_0 above 0 to preserve the ECP property. We outline the argument for this; for details see the proof of Lemma 2 in the Appendix. In the constructed allocation, a coalitional report of η_L results in the same outcome as a report of η_0 by S, so it suffices to compare the outcomes of sending a joint report of η_L relative to η_H. Raising p_H above $p_L = p^{NS}$ implies a higher payoff for A for delivering the output following a joint report of η_H rather than η_L. This creates additional scope for extortion of A, supported by S's threat to misrepresent the signal η_H when it arises as η_L instead (if A refuses the side-contract), which would lower A's outside option payoff in state (θ, η_H) from $\max\{p_H - \theta, 0\}$ to $\max\{p^{NS} - \theta, 0\}$. Conditional on signal realization η_H, the third-party earns a payoff of

$$(1-\alpha)F_H(p_H)(b - p_H) + \alpha \int_{\underline{\theta}}^{p_H} F_H(\theta)d\theta$$

if this signal is reported truthfully, compared with

$$(1-\alpha)F_H(p^{NS})(b-p^{NS})+\alpha\int_{\underline{\theta}}^{p^{NS}} F_H(\theta)d\theta$$

if S reports η_L instead. These two payoffs are the same when $p_H = p^{NS}$, but the former payoff from truthful reporting is locally decreasing in p_H at $p_H = p^{NS}$ if $\alpha < \frac{1}{2}$ (given that $b = p^{NS}$). Since the payoff from misrepresentation is independent of p_H, a small rise in p_H above p^{NS} would motivate the side-contract designer to deviate to a non-null contract with this misrepresentation.

In order to deter this deviation, A needs to be 'protected' by raising the lumpsum payment to A, u_j, in $j = L, 0$ to accompany the rise in p_H so that

$$(1-\alpha)(-u_j)+\alpha[\int_{\underline{\theta}}^{p^{NS}} F_H(\theta)d\theta+u_j]=(1-\alpha)F_H(p_H)(b-p_H)+\alpha\int_{\underline{\theta}}^{p_H} F_H(\theta)d\theta$$

which reduces to the expression (4) for u_j. Condition (CP2) requires X_0 to be raised in step with u_0. This explains the construction of the allocation in Lemma 2.

The Lemma also requires p_H to not be raised too much above p^{NS} as that would generate an opposite incentive for S to threaten A by misrepresenting the signal η_L as η_H instead. It also creates an incentive to lower the price offered to A below p_H following signal η_H and compensating A with an accompanying side-payment. These details are spelt out in the proof, which also checks that the constructed allocation in Lemma 2 is ECP.

Finally, we explain the effect of this variation on P's payoff (5), which equals W^{NS} at $p_H = p^{NS}$. A small increase in p_H from p^{NS} has two opposite effects on this payoff. It increases the payoff by alleviating the underproduction problem in NS:

$$\frac{d[\kappa_H F_H(p_H)(V-b)]}{dp_H}\Big|_{p_H=b=p^{NS}}=\kappa_H f_H(p^{NS})(V-p^{NS})=a_H(p^{NS})F(p^{NS}),$$

where we use the first-order condition satisfied by p^{NS}: $V = p^{NS} + F(p^{NS})/f(p^{NS})$. On the other hand, the variation lowers P's payoff by increasing X_0, since

$$\begin{aligned} dX_0/dp_H|_{p_H=p^{NS}} &= \frac{F_H(p_H)+f_H(p_H)(p_H-p^{NS})-(1-\beta)F_H(p_H)}{\beta}\Big|_{p_H=p^{NS}} \\ &= F_H(p^{NS}). \end{aligned}$$

We therefore obtain a sufficient condition for the variation to raise P's payoff:

Proposition 3. *Suppose $F(p^{NS})a_H(p^{NS}) > F_H(p^{NS})$. Then hiring S is valuable for any $\alpha \in [0, 1/2)$.*

Note that this sufficient condition does not depend on α. A specific example where it holds is that of a uniform prior of θ on $[0, 1]$, with $a_H(\theta)$ satisfying $a_H(\theta) = a_L(1-\theta)$, $a_H(\underline{\theta}) = 0$, and $a_H(\theta)$ strictly convex on $[0, 1/2]$.[21] With a uniform prior, the condition stated in Proposition 3 reduces to

[21] $a_H(\theta)$ also needs to satisfy assumption (i) listed at the beginning of this subsection. All these conditions hold when $a_H(\theta) = 2\theta^2$ for $\theta \in [0, 1/2]$ and $1 - 2(1-\theta)^2$ for $\theta \in (1/2, 1]$.

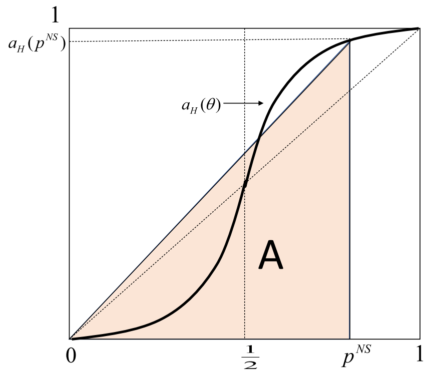

Fig. 3. Example where hiring S is valuable.

$$p^{NS}a_H(p^{NS})/2 > \int_0^{p^{NS}} a_H(\theta)d\theta.$$

Fig. 3 shows an example of $a_H(\theta)$ satisfying these sufficient conditions. For some $p^{NS} \in (0, 1)$, $p^{NS}a_H(p^{NS})/2$ is equal to the area of the triangle A in Fig. 3, and $\int_0^{p^{NS}} a_H(\theta)d\theta$ is the area under $a_H(\theta)$ between 0 and p^{NS}. We can confirm that the latter is less than the area A for any choice of $p^{NS} = V/2 \in (0, 1)$. Thus, in this example, S is valuable for any $V \in (0, 2)$.

It is interesting to note the contrast between settings of weak collusion and extortionary collusion. The online Appendix in Mookherjee et al. (2020) shows that, in the context of weak collusion, hiring S is unconditionally valuable under the same model assumptions. This is based on a small variation in b as well as p_H on the optimal allocation in NS, which satisfies $b < p_L = p^{NS} < p_H$. The bonus b is lowered, which is used by P to extract the collusion rent generated by raising p_H. This variation is not feasible in the presence of extortion. Condition (CP1) for collusion-proofness requires a lower b to be accompanied with a corresponding reduction in p_0. In turn this generates a coalitional incentive where S threatens to report η_0 in both states of L and H. A further increase in u_0 and X_0 is required to deter such a deviation. This reduces the profits generated for P from the variation in the allocation, reflected in the condition imposed in Proposition 3 for hiring S to be valuable. In the working paper version (Mookherjee and Tsumagari (2022)), it is shown that extortionary collusion always results in a lower payoff for P, as compared to weak collusion.

6. Concluding remarks

We have not discussed the question of optimality of delegation, which has been studied in a number of preceding papers on mechanism design with weak collusion that ignored the possibility of extortion (e.g., Faure-Grimaud et al. (2003), Celik (2009), Mookherjee et al. (2020)). Delegation refers to a setting where P contracts only with S, and delegates the authority to (sub-)contract with A to S. This corresponds to a special case of our analysis where P offers a null side-contract to A and α is equal to zero. It is easy to show in the setting of our current model that there is no value of hiring S in a delegation setting. The reasoning is as follows. In the setting of *ex ante* weak collusion, Mookherjee et al. (2020) show there is no value of hiring S in the organization with delegation to S. Since A's outside payoff while contracting with S is identically equal to zero under delegation, any allocation that is achievable with weak collusion does not leave any scope for extortion of A by S. Hence, with delegation to S, any allocation that is achievable in weak collusion is also achievable in extortionary collusion and hiring S cannot be valuable.

Next we explain consequences of relaxing some of the assumptions of our model. We assumed S's signal has full support, i.e., every $\eta \in \Pi$ occurs with positive probability for any $\theta \in \Theta$. If we relax this, CPP can be shown to hold. Let $\Theta(\eta)$ denote the support of θ given η. An achievable allocation (u_A, u_S, q) is defined only on (θ, η) such that $\theta \in \Theta(\eta)$. This can aggravate the problem of extortion, necessitating a variation in how revelation mechanisms need to be augmented. Specifically, the revelation mechanism needs to specify an allocation assigned for messages $(m_A, m_S) = ((\theta_A, \eta_A), \eta_S)$ such that $\theta_A \notin \Theta(\eta_S)$. If such inconsistent messages are penalized, it provides an additional instrument of extortion. However, it is still possible to augment the mechanism to deal with this problem. Moreover, Proposition 2 continues to hold, and Proposition 3 can also be shown to hold for specific information structures without full support. The details are available upon request.

What happens when A cannot observe S's signal? In this case, it is difficult to establish CPP for the following reason. Our proof relied on the property that $u_A(\theta, \eta) \geq \hat{u}_A(\theta, \mu_S(\eta), GC)$ for each η, which was used to ensure that the scope of extortion is not enlarged when the mechanism is altered from GC to GC^R. When the signal is privately observed by S, this property becomes difficult to ensure with the kind of augmentation employed in this paper. It is possible that some more complicated augmentation may suffice to generate the same result, but we leave this as an open question.

Our results extend to the context of interim collusion, in which S and A cannot collude on participation decisions, but extortion is possible. Extending Propositions 1 and 2 is relatively straightforward. With $\alpha < 1/2$, in the context of the example of an indivisible good studied in Section 5.3, the characterization of collusion-proof allocation in interim collusion differs from that in ex-ante collusion in two respects. First, X_0 could be negative as S and A can no longer coordinate their participation decisions. Second, since collusion occurs after S and A accept GC, S's exit option from GC cannot be used to threaten A. This implies that we do not need to add an auxiliary message η_0 in constructing a collusion-proof revelation mechanism. If $\alpha = 0$, we can establish almost the same results as in the weak collusion setting of Mookherjee et al. (2020): the optimal allocation differs between interim collusion and ex-ante collusion if and only if delegation to S is optimal under interim collusion. Since A's outside payoff equals zero with delegation to S, the optimal allocation achieved under weak collusion is also ECP. On the other hand, when the optimal allocation does not differ between interim collusion and ex ante collusion, we can show that extortionary collusion lowers P's payoff as compared to weak collusion.[22] Thus extortionary collusion expands the range of environments where delegation to S becomes optimal.

Finally, our model assumed that the information structure is exogenous. In some settings, P may have the capacity to control information available to S and A respectively, which can be an important instrument for controlling corruption, as studied by Ortner and Chassang (2018), Asseyer (2020) and von Negenborn and Pollrich (2020). Extending their analyses to contexts with extortion is an interesting question for future research.

[22] See Proposition 3(3) in the working paper version (Mookherjee and Tsumagari (2022)).

Data availability

No data was used for the research described in the article.

Appendix A

Proof of Proposition 1

We begin with the proof of the following lemma.

Lemma A1. *If (X, u_A, u_S, q) is achieved as a PBE(c) outcome in GC, then for any $\eta \in \Pi$:*

$$(\tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), \tilde{\mu}(\theta,\eta)) = (u_A(\theta,\eta), u_S(\theta,\eta), \mu(\theta,\eta))$$

solves $P^E(\alpha : \eta, GC)$ for some $\tilde{\mu}_S(\eta) = \mu_S(\eta) \in \Delta(M_S)$, and

$$X(\theta,\eta) - \theta q(\theta,\eta) = u_A(\theta,\eta) + u_S(\theta,\eta).$$

Proof of Lemma A1

Since (X, u_A, u_S, q) is an achievable allocation, it is straightforward to check that it is feasible in $P^E(\alpha : \eta, GC)$. Here for a reporting strategy $\mu_S(\eta)$ for S in this GC, $\hat{u}_A(\theta, \mu_S(\eta), GC)$ is interpreted as the A's maximum payoff in the event that A is of type θ and exits from the side-contract, whence S chooses $\mu_S(\eta)$.

If $(u_A(\theta,\eta), u_S(\theta,\eta), \mu(\theta,\eta), \mu_S(\eta))$ does not solve problem $P^E(\alpha : \eta, GC)$ for some η, we shall now show that there exist another side-contract and a continuation equilibrium in which the third party can achieve a higher payoff, which will contradict the hypothesis that the allocation resulted from a PBE(c) of GC. Suppose that for some η, the solution of $P^E(\alpha : \eta, GC)$ is instead some $(\tilde{u}^*_A(\theta,\eta), \tilde{u}^*_S(\theta,\eta), \tilde{\mu}^*(\theta,\eta), \tilde{\mu}^*_S(\eta)) \neq (u_A(\theta,\eta), u_S(\theta,\eta), \mu(\theta,\eta), \mu_S(\eta))$.

Construct a side-contract SC' as follows. Conditioned on the acceptance of SC' by both A and S, the third party requests a report from A of $(\theta_A, \eta_A) \in K$, and a report from S of $\eta_S \in \Pi$. The report to P is subsequently selected according to $\tilde{\mu}^*(\theta_A, \eta_S)$, while side-transfers are selected as follows.

$$t_A(\theta_A,\eta_A,\eta_S) = \tilde{u}^*_A(\theta_A,\eta_S) - [\hat{X}_A(\tilde{\mu}^*(\theta_A,\eta_S)) - \theta_A \hat{q}(\tilde{\mu}^*(\theta_A,\eta_S))] - l(\eta_A,\eta_S)$$

and

$$t_S(\theta_A,\eta_A,\eta_S) = \tilde{u}^*_S(\theta_A,\eta_A) - \hat{X}_S(\tilde{\mu}^*(\theta_A,\eta_S))$$

where $l(\eta_A,\eta_S)$ is zero for $\eta_A = \eta_S$ and a large positive number for $\eta_A \neq \eta_S$. These transfers satisfy the budget balance conditions: $t_A(\theta_A,\eta_A,\eta_S) + t_S(\theta_A,\eta_A,\eta_S) \leq 0$ for any (θ_A,η_A,η_S), because of constraint (iii) for $\eta_A = \eta_S$ and sufficiently large $l(\eta_A,\eta_S)$ for $\eta_A \neq \eta_S$.

If A were to accept and S were to reject SC', A would threaten to play $\underline{\mu}_A$. Conversely, if S accepts and reports η_S while A rejects SC', S threatens to play $\tilde{\mu}^*_S(\eta_S)$. It is easy to check that there exists a continuation equilibrium where nobody rejects SC' on the equilibrium path, and both A and S report truthfully to the third party, resulting in the allocation $(\tilde{u}^*_A(\theta,\eta), \tilde{u}^*_S(\theta,\eta), \hat{q}(\tilde{\mu}^*(\theta,\eta)))$. The third party attains a higher payoff, contradicting the hypothesis that we started with a PBE(c), completing the proof of the first part in the lemma.

Finally, if $X(\theta,\eta) - \theta q(\theta,\eta) > u_A(\theta,\eta) + u_S(\theta,\eta)$, we can find the choice of feasible control variables with $u_S^{**}(\theta,\eta) = X(\theta,\eta) - \theta q(\theta,\eta) - u_A(\theta,\eta)$ instead of $u_S(\theta,\eta)$, taking the other parts of solution as given, which improves the third party's payoff. It implies $X(\theta,\eta) - \theta q(\theta,\eta) = u_A(\theta,\eta) + u_S(\theta,\eta)$. ■

Now we return the proof of Proposition 1.

Proof of Necessity

Suppose that (u_A, u_S, q) is an achievable allocation in GC in extortionary collusion with α. Evidently (u_A, u_S, q) is IIC. Consider an augmentation of (u_A, u_S, q) to the domain $\bar{K}$ with the selection of

$$(u_A(\theta,\eta_0), u_S(\theta,\eta_0), q(\theta,\eta_0)) \equiv (\hat{u}_A(\theta,\bar{\mu}_S, GC), \omega, \bar{q}(\mu_A(\theta,\bar{\mu}_S, GC), \bar{\mu}_S))$$

where $\mu_A(\theta,\bar{\mu}_S, GC)$ maximizes $\bar{X}_A(\mu_A,\bar{\mu}_S) - \theta\bar{q}(\mu_A,\bar{\mu}_S)$ subject to $\mu_A \in \Delta(M_A)$, and $\omega \equiv \underline{w}_S(GC)$. By the definition, $\omega \geq 0$ and $(u_A(\theta,\eta_0), q(\theta,\eta_0))$ satisfies (IC_A) and (PC_A).

Now consider the problem $P^E(\alpha:\eta)$ defined by the augmented allocation (u_A, u_S, q) on $\bar{K}$. Note that this problem differs from the one considered in Lemma A1 ($P^E(\alpha:\eta, GC)$), as it no longer refers to the original GC.

We show that

$$(\mu(\theta,\eta), \tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), P(\cdot \mid \eta)) = (I_1(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), I_2(\eta))$$

solves problem $P^E(\alpha:\eta)$.

It is straightforward to check that $(I_1(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), I_2(\eta))$ satisfies all constraints of $P^E(\alpha:\eta)$, and generates a payoff for the third party of

$$E[\alpha u_A(\theta,\eta) + (1-\alpha)u_S(\theta,\eta) \mid \eta].$$

Suppose otherwise: that there exists some alternative choice of controls

$$(\mu^*(\theta,\eta), u_A^*(\theta,\eta), u_S^*(\theta,\eta), P^*(\cdot \mid \eta))$$

which is feasible in $P^E(\alpha:\eta)$, such that

$$E[\alpha u_A^*(\theta,\eta) + (1-\alpha)u_S^*(\theta,\eta) \mid \eta] > E[\alpha u_A(\theta,\eta) + (1-\alpha)u_S(\theta,\eta) \mid \eta].$$

We show that in such a case there would exist

$$\tilde{\mu}(\theta,\eta) \in \Delta(M_A \times M_S), \tilde{u}_A(\theta,\eta), \tilde{u}_S(\theta,\eta), \tilde{\mu}_S(\eta) \in \Delta(M_S)$$

which would be feasible in $P^E(\alpha:\eta, GC)$ and generate a higher value in that problem compared to $(\mu(\theta,\eta), u_A(\theta,\eta), u_S(\theta,\eta), \mu_S(\eta))$, thereby contradicting the result established at Lemma A1.

$\mu^*(\theta,\eta)$, which is a probability measure on $\bar{K} \cup \{e\}$, divides its weight between reports either in $K \cup \{e\}$ or satisfying $\eta = \eta_0$. The former event corresponds to an outcome of GC that results when S and A's reports are chosen from M_S and M_A respectively. And the latter event corresponds (by specification of $(u_A(\theta,\eta_0), u_S(\theta,\eta_0), q(\theta,\eta_0)))$ to an outcome of GC resulting when S reports $\bar{\mu}_S \in \Delta(M_S)$ and A reports according to $\mu_A(\theta,\bar{\mu}_S) \in \Delta(M_A)$. In this case,

$$\bar{q}(\mu_A(\theta,\bar{\mu}_S), \bar{\mu}_S) = q(\theta,\eta_0)$$

while

$$X(\theta,\eta_0)=\omega+\bar{X}_A(\mu_A(\theta,\bar{\mu}_S),\bar{\mu}_S)\le \bar{X}_S(\mu_A(\theta,\bar{\mu}_S),\bar{\mu}_S)+\bar{X}_A(\mu_A(\theta,\bar{\mu}_S),\bar{\mu}_S)$$

since ω is S's minmax payoff in GC. Hence the outcome of $\mu^*(\theta,\eta)$ in $P^E(\alpha:\eta)$ can be attained by the coalition as an outcome of GC resulting from some reporting strategy $\tilde{\mu}(\theta,\eta)\in\Delta(M_A\times M_S)$ that satisfies

$$\hat{X}_A(\tilde{\mu}(\theta,\eta))+\hat{X}_S(\tilde{\mu}(\theta,\eta))\ge X(\mu^*(\theta,\eta))$$

and

$$\hat{q}(\tilde{\mu}(\theta,\eta))=q(\mu^*(\theta,\eta)).$$

Let $\mu_S(\eta)$ denote the optimal threat chosen by S in the event that A does not participate in the side-contract, in the solution to problem $P^E(\alpha:\eta,GC)$. Let us select $\mu_S(\eta_0)\equiv\bar{\mu}_S$. Then $u_A(\theta,\eta)\ge\hat{u}_A(\theta,\mu_S(\eta),GC)$ for any $\eta\in\bar{\Pi}$. Define $\tilde{\mu}_S(\eta)\in\Delta(M_S)$ as the composite of the measures $\mu_S(\eta')$ and $P^*(\eta'\mid\eta)$. Then by the definition of $\hat{u}_A(\theta,\mu_S,GC)$,

$$\Sigma_{\eta'\in\bar{\Pi}}P^*(\eta'\mid\eta)u_A(\theta,\eta')\ge\Sigma_{\eta'\in\bar{\Pi}}P^*(\eta'\mid\eta)\hat{u}_A(\theta,\mu_S(\eta'),GC)\ge\hat{u}_A(\theta,\tilde{\mu}_S(\eta),GC).$$

Since $u_A^*(\theta,\eta)\ge\Sigma_{\eta'\in\bar{\Pi}}P^*(\eta'\mid\eta)u_A(\theta,\eta')$, it follows that $u_A^*(\theta,\eta)\ge\hat{u}_A(\theta,\tilde{\mu}_S(\eta),GC)$. Defining $\tilde{u}_A(\theta,\eta)\equiv u_A^*(\theta,\eta)$ and

$$\tilde{u}_S(\theta,\eta)\equiv\hat{X}_A(\tilde{\mu}(\theta,\eta))+\hat{X}_S(\tilde{\mu}(\theta,\eta))-\theta\hat{q}(\tilde{\mu}(\theta,\eta))-u_A^*(\theta,\eta),$$

we infer that $(\tilde{u}_A(\theta,\eta),\tilde{u}_S(\theta,\eta),\tilde{\mu}(\theta,\eta),\tilde{\mu}_S(\eta))$ is feasible in the problem $P^E(\alpha:\eta,GC)$, and $\tilde{u}_S(\theta,\eta)\ge u_S^*(\theta,\eta)$. Hence it generates a higher payoff for the third party than $E[\alpha u_A(\theta,\eta)+(1-\alpha)u_S(\theta,\eta)\mid\eta]$, and we obtain a contradiction to the result of Lemma A1. So $(I_1(\theta,\eta), u_A(\theta,\eta),u_S(\theta,\eta),I_2(\eta))$ must be a solution of $P^E(\alpha:\eta)$, establishing the necessity of the statement.

Proof of Sufficiency

Let (u_A,u_S,q) be an augmentation of allocation for which the latter satisfies the ECP property. By Definition 2, $u_S(\theta,\eta_0)=\omega$ for any $\theta\in\Theta$ and $(u_A(\theta,\eta_0),q(\theta,\eta_0))$ satisfies (IC_A) and (PC_A). P can construct a revelation grand contract GC^R as follows:

$$(X_A(m_A,m_S),X_S(m_A,m_S),q(m_A,m_S):M_A,M_S)$$

where $M_A=K\cup\{e_A\}$, $M_S=\bar{\Pi}\cup\{e_S\}$ and

- for any $(\theta,\eta)\in K$ and $\eta'\in\Pi$, choose $(X_A((\theta,\eta),\eta'),X_S((\theta,\eta),\eta'),q((\theta,\eta),\eta'))=(u_A(\theta,\eta')+\theta q(\theta,\eta'),u_S(\theta,\eta)-L(\eta,\eta'),q(\theta,\eta'))$ where $L(\eta,\eta')=0$ for $\eta=\eta'$ and $L>0$ (and sufficiently large) for $\eta\neq\eta'$
- $(X_A((\theta,\eta),e_S),X_S((\theta,\eta),e_S),q((\theta,\eta),e_S))=(u_A(\theta,\eta_0)+\theta q(\theta,\eta_0),0,q(\theta,\eta_0))$.
- $(X_A((\theta,\eta),\eta_0),X_S((\theta,\eta),\eta_0),q((\theta,\eta),\eta_0))=(u_A(\theta,\eta_0)+\theta q(\theta,\eta_0),\omega,q(\theta,\eta_0))$.
- $(X_A(e_A,m_S),X_S(e_A,m_S),q(e_A,m_S))=(0,0,0)$ for any $m_S\neq\eta_0$
- $(X_A(e_A,\eta_0),X_S(e_A,\eta_0),q(e_A,\eta_0))=(0,\omega,0)$.

It is easy to check that $(m_A,m_S)=((\theta,\eta),\eta)$ is a non-cooperative equilibrium of GC^R, and S's minmax payoff in GC^R is ω. The ECP property of (u_A,u_S,q) implies there is no room for the third party to improve its payoff by offering a deviating side-contract, so (u_A, u_S, q) is realized as the outcome of a PBE(c) under GC^R. ■

Proof of Proposition 2

Consider an ECP allocation (u_A, u_S, q) for $\alpha \geq \frac{1}{2}$. By Proposition 1, there exists $\omega \geq 0$ and an incentive compatible augmentation of this allocation satisfying $u_S(\theta, \eta_0) = \omega$, such that for any η, $(I_1(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), I_2(\eta))$ solves $P^E(\alpha : \eta)$.

Here, for (X, q) in Definition 2, we show that

$$X(\theta, \eta) - \theta q(\theta, \eta) \geq X(\mu) - \theta q(\mu) \tag{6}$$

for any $(\theta, \eta) \in K$ and for any $\mu \in \Delta(\bar{K} \cup \{e\})$. Suppose otherwise that, for some subset of K with positive measure, there exists $\mu(\theta, \eta) \in \Delta(\bar{K} \cup \{e\})$ such that $X(\mu(\theta, \eta)) - \theta q(\mu(\theta, \eta)) > X(\theta, \eta) - \theta q(\theta, \eta)$. Then, if we choose

$$\mu^*(\theta) \in \arg \max_{\mu \in \Delta(\bar{K} \cup \{e\})} [X(\mu) - \theta q(\mu)]$$

for each θ,

$$E[X(\mu^*(\theta)) - \theta q(\mu^*(\theta)) \mid \eta] > E[X(\theta, \eta) - \theta q(\theta, \eta) \mid \eta].$$

We claim that

$$(\mu(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta)) = (\mu^*(\theta), X(\mu^*(\theta)) - \theta q(\mu^*(\theta)) - \omega, \omega)$$

is a solution of $P^E(\alpha : \eta)$ for any η.

To establish the claim, we first derive an upper bound for the objective function in the problem $P^E(\alpha : \eta)$. From the constraint $E[\tilde{u}_S(\theta, \eta) \mid \eta] \geq \omega$ and the assumption that $\alpha \geq 1/2$, for any reporting strategy $\mu(\theta, \eta)$, the following is true:

$$\begin{aligned} & E[\alpha \tilde{u}_A(\theta, \eta) + (1 - \alpha)\tilde{u}_S(\theta, \eta) \mid \eta] \\ \leq\ & E[\alpha\{X(\mu(\theta, \eta)) - \theta q(\mu(\theta, \eta))\} + (1 - 2\alpha)\tilde{u}_S(\theta, \eta) \mid \eta] \\ \leq\ & \alpha E[X(\mu^*(\theta)) - \theta q(\mu^*(\theta)) \mid \eta] + (1 - 2\alpha)\omega. \end{aligned}$$

This upper bound can be attained in $P^E(\alpha : \eta)$ by choosing $\mu(\theta, \eta) = \mu^*(\theta)$,

$$\tilde{u}_A(\theta, \eta) = X(\mu^*(\theta)) - \theta q(\mu^*(\theta)) - \omega$$

and $\tilde{u}_S(\theta, \eta) = \omega$ for any $\theta \in \Theta$, and $P(\eta_0 \mid \eta) = 1$ and $P(\eta' \mid \eta) = 0$ for any $\eta' \neq \eta_0$. This allocation satisfies A's participation constraint (v), since

$$\begin{aligned} & \tilde{u}_A(\theta, \eta) = X(\mu^*(\theta)) - \theta q(\mu^*(\theta)) - \omega \\ \geq\ & X(\theta, \eta_0) - \theta q(\theta, \eta_0) - u_S(\theta, \eta_0) = u_A(\theta, \eta_0). \end{aligned}$$

As the other constraints are obviously satisfied, the claim is established. It follows that $(I_1(\theta, \eta), u_A(\theta, \eta), u_S(\theta, \eta), I_2(\eta))$ does not solve $P^E(\alpha : \eta)$, generating a contradiction. Hence (6) holds.

We now show that the same payoff as in the given ECP allocation can be attained in NS. Observe that (6) implies

$$X(\theta, \eta) - \theta q(\theta, \eta) \geq X(\theta', \eta') - \theta q(\theta', \eta')$$

and

$$X(\theta,\eta)-\theta q(\theta,\eta)\geq 0,$$

for any $(\theta,\eta),(\theta',\eta')\in K$. For each θ, let $\eta(\theta)$ denote a choice of η which maximizes P's payoff over the set of allocations $\{(X(\theta,\eta),q(\theta,\eta))\mid \eta\in\Pi\}$, and define $(\tilde{X}(\theta),\tilde{q}(\theta))\equiv(X(\theta,\eta(\theta)),q(\theta,\eta(\theta)))$. Then by construction:

$$\tilde{X}(\theta)-\theta\tilde{q}(\theta)\geq\tilde{X}(\theta')-\theta\tilde{q}(\theta')$$

and

$$\tilde{X}(\theta)-\theta\tilde{q}(\theta)\geq 0.$$

The allocation which assigns $\tilde{X}(\theta)$ paid to A and expected output $\tilde{q}(\theta)$ for θ is achievable in NS and generates an expected payoff for P which is at least as large as the payoff resulting from the given ECP allocation. ■

Proof of Lemma 1

Suppose that for some η, $(I_1(\theta,\eta),u_A(\theta,\eta),u_S(\theta,\eta),I_2(\eta))$ does not solve the relaxed version of $P^E(\alpha:\eta)$ where the constraint $E[\tilde{u}_S(\theta,\eta)\mid\eta]\geq\omega$ is dropped. It implies $E[\tilde{u}^r_S(\theta,\eta)\mid\eta]<\omega$ in the optimal solution of the relaxed problem represented by

$$(\mu^r(\theta,\eta),\tilde{u}^r_A(\theta,\eta),\tilde{u}^r_S(\theta,\eta),P^r(\cdot\mid\eta)).$$

As shown in the proof of Proposition 2, side contract $\tilde{SC}$ defined as follows is feasible in $P^E(\alpha:\eta)$, hence also in the relaxed problem:

- $\tilde{\mu}(\theta,\eta)=\mu^*(\theta)$ which maximizes $X(\mu)-\theta q(\mu)$ subject to $\mu\in\Delta(\bar{K}\cup\{e\})$
- $P(\eta_0\mid\eta)=1$ and $P(\eta'\mid\eta)=0$ for any $\eta'\neq\eta_0$
- $\tilde{u}_A(\theta,\eta)=X(\mu^*(\theta))-\theta q(\mu^*(\theta))-\omega$ (denoted by $u^+_A(\theta,\eta)$ in later part)
- $\tilde{u}_S(\theta,\eta)=\omega$

Hence

$$\begin{aligned}&E[\alpha\tilde{u}^r_A(\theta,\eta)+(1-\alpha)\tilde{u}^r_S(\theta,\eta)\mid\eta]\\&=E[(1-\alpha)\{X(\mu^r(\theta,\eta))-\theta q(\mu^r(\theta,\eta))\}-(1-2\alpha)\tilde{u}^r_A(\theta,\eta)\mid\eta]\\&\geq E[(1-\alpha)\{X(\mu^*(\theta))-\theta q(\mu^*(\theta))\}-(1-2\alpha)u^+_A(\theta,\eta)\mid\eta].\end{aligned}$$

But since $E[X(\tilde{\mu}^r(\theta,\eta))-\theta q(\tilde{\mu}^r(\theta,\eta))\mid\eta]\leq E[X(\mu^*(\theta))-\theta q(\mu^*(\theta))\mid\eta]$ by the definition of $\mu^*(\theta)$, $\alpha<\frac{1}{2}$ implies that $E[u^+_A(\theta,\eta)\mid\eta]\geq E[\tilde{u}^r_A(\theta,\eta)\mid\eta]$. This implies that the side contract $\tilde{SC}$ creates a Pareto improvement over the solution to the relaxed problem, yielding a strictly higher value of the third party's expected payoff, a contradiction. ■

Proof of Lemma 2

For $\beta\in(0,1]$, define

$$L_i(p,\tilde{p}:\beta)\equiv-\frac{\int_{\tilde{p}}^{p}[\tilde{p}-(\theta+\beta\frac{F_i(\theta)}{f_i(\theta)})]dF_i(\theta)}{\beta}=\frac{F_i(p)(p-\tilde{p})-(1-\beta)\int_{\tilde{p}}^{p}F_i(\theta)d\theta}{\beta}.$$

This is positive for $p > \tilde{p}$, and equals zero for $p = \tilde{p}$.

We begin with the following lemma.

Lemma A2. *For each $\alpha \in [0, 1/2)$, there exists $\bar{p}_H > p^{NS}$ such that*

$$p^{NS} > p_H + \beta \frac{F_H(p_H) - 1}{f_H(p_H)},$$

and

$$L_H(p_H, p^{NS} : \beta) < (p_H - p^{NS}) F_L(p^{NS})$$

for any $p_H \in (p^{NS}, \bar{p}_H)$.

Proof of Lemma A2

Since $p^{NS} < \bar{\theta}$ and $\beta > 0$, $p^{NS} > p^{NS} + \beta \frac{F_H(p^{NS})-1}{f_H(p^{NS})}$. By the continuity of $[p + \beta \frac{F_H(p)-1}{f_H(p)}]$, there exists $p'_H > p^{NS}$ such that $p^{NS} > p + \beta \frac{F_H(p_H)-1}{f_H(p_H)}$ for any $p_H \in (p^{NS}, p'_H)$. On the other hand, since

$$\frac{\partial L_H(p_H, p^{NS} : \beta)}{\partial p_H}|_{p_H = p^{NS}} = F_H(p^{NS}) < F_L(p^{NS}) = \frac{\partial (p_H - p^{NS}) F_L(p^{NS})}{\partial p_H}|_{p_H = p^{NS}}$$

and

$$L_H(p_H, p^{NS} : \beta)|_{p_H = p^{NS}} = (p_H - p^{NS}) F_L(p^{NS})|_{p_H = p^{NS}} = 0,$$

there exists $p''_H > p^{NS}$ such that $L_H(p_H, p^{NS} : \beta) < (p_H - p^{NS}) F_L(p^{NS})$ for any $p_H \in (p^{NS}, p''_H)$. Then the statement holds with $\bar{p}_H \equiv \min\{p'_H, p''_H\}$. ∎

For $p_H \in (p^{NS}, \bar{p}_H)$, consider allocation $(b, X_0, p_H, p_L, p_0, u_H, u_L, u_0)$ specified in Lemma 2. Evidently (CP1), (CP2) and $u_j \geq 0$ for $j \in \{L, H, 0\}$ are satisfied in this allocation. We therefore focus hereafter on the condition (CP3). The following lemma is useful in confirming this condition.

Lemma A3. *For $i \in \{L, H\}$, $\tilde{q}(\theta, \eta_i) = q(\theta, \eta_i)$, $\tilde{u}_A(\theta, \eta_i) = u_A(\theta, \eta_i)$ and $P(\cdot \mid \eta_i) = I_2(\eta_i)$ solve $\bar{P}^E(\alpha : \eta_i)$, if there exists non-decreasing $\Lambda_i : \Theta \to [0, 1]$ with $\Lambda_i(\underline{\theta}) = 0$ and $\Lambda_i(\bar{\theta}) = 1$ such that*

$$p_i = \arg\max_{\tilde{p}_i \in \Theta} \int_{\underline{\theta}}^{\tilde{p}_i} [b - W_i(\theta : \beta)] dF_i(\theta) \tag{7}$$

and

$$\int_{\underline{\theta}}^{p_i} \Lambda_i(\theta) d\theta + u_i \leq \int_{\underline{\theta}}^{p_j} \Lambda_i(\theta) d\theta + u_j \tag{8}$$

for $j \in \{L, H, 0\}$, where

$$W_i(\theta : \beta) \equiv \theta + \beta \frac{F_i(\theta) - \Lambda_i(\theta)}{f_i(\theta)}.$$

Proof of Lemma A3

First, (7) implies that $\tilde{q}(\cdot,\eta_i)=q(\cdot,\eta_i)$ maximizes

$$\int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta,\eta_i)[b-W_i(\theta:\beta)]dF_i(\theta)$$

subject to $\tilde{q}(\theta,\eta_i)\in[0,1]$ which is non-increasing in θ, since

$$\int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta,\eta_i)[b-W_i(\theta:\beta)]dF_i(\theta)=\int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta,\eta_i)\frac{d\int_{\underline{\theta}}^{\theta}[b-W_i(\theta':\beta)]dF_i(\theta')}{d\theta}d\theta$$

$$=\tilde{q}(\bar{\theta},\eta_i)\int_{\underline{\theta}}^{\bar{\theta}}[b-W_i(\theta:\beta)]dF_i(\theta)-\int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{\theta}[b-W_i(\theta':\beta)]dF_i(\theta')d\tilde{q}(\theta,\eta_i)$$

$$\leq\tilde{q}(\bar{\theta},\eta_i)\int_{\underline{\theta}}^{p_i}[b-W_i(\theta:\beta)]dF_i(\theta)-\int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{p_i}[b-W_i(\theta':\beta)]dF_i(\theta')d\tilde{q}(\theta,\eta_i)$$

$$=\tilde{q}(\underline{\theta},\eta_i)\int_{\underline{\theta}}^{p_i}[b-W_i(\theta:\beta)]dF_i(\theta)\leq\int_{\underline{\theta}}^{p_i}[b-W_i(\theta:\beta)]dF_i(\theta)$$

$$=\int_{\underline{\theta}}^{\bar{\theta}} q(\theta,\eta_i)[b-W_i(\theta:\beta)]dF_i(\theta).$$

The first inequality uses (7) and the non-increasing property of $\tilde{q}(\theta,\eta_i)$, while the second inequality comes from $\tilde{q}(\underline{\theta},\eta_i)\leq 1$ and $\int_{\underline{\theta}}^{p_i}[b-W_i(\theta:\beta)]dF_i(\theta)\geq 0$ implied by (7).

Next, for Λ_i which satisfies conditions in the lemma, we obtain the following inequalities for the choice of control variables satisfying the constraints of $\bar{P}^E(\alpha:\eta_i)$:

$$E[\tilde{q}(\theta,\eta_i)(b-\theta)+X_0-\beta\tilde{u}_A(\theta,\eta_i)\mid\eta_i]$$

$$\leq E[\tilde{q}(\theta,\eta_i)(b-\theta)+X_0-\beta\tilde{u}_A(\theta,\eta_i)\mid\eta_i]$$

$$+\beta\int_{\underline{\theta}}^{\bar{\theta}}[\tilde{u}_A(\theta,\eta_i)-\Sigma_{j\in\{L,H,0\}}P(\eta_j\mid\eta_i)\{\max\{p_j-\theta,0\}+u_j\}]d\Lambda_i(\theta)$$

$$=\int_{\underline{\theta}}^{\bar{\theta}}\tilde{q}(\theta,\eta_i)[b-W_i(\theta:\beta)]dF_i(\theta)+X_0-\beta\Sigma_{j\in\{L,H,0\}}P(\eta_j\mid\eta_i)[\int_{\underline{\theta}}^{p_j}\Lambda_i(\theta)d\theta+u_j]$$

$$\leq\int_{\underline{\theta}}^{\bar{\theta}} q(\theta,\eta_i)[b-W_i(\theta:\beta)]dF_i(\theta)+X_0-\beta[\int_{\underline{\theta}}^{p_i}\Lambda_i(\theta)d\theta+u_j].$$

The first inequality comes from constraint (2) in $\bar{P}^E(\alpha : \eta_i)$ and non-decreasing $\Lambda_i(\theta)$. The next equality uses constraint (a) of $\bar{P}^E(\alpha : \eta_i)$, and integration by parts:

$$\int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} \tilde{q}(\tilde{\theta}, \eta_i) d\tilde{\theta} dF_i(\theta) = \int_{\underline{\theta}}^{\bar{\theta}} \tilde{q}(\theta, \eta_i) F_i(\theta) d\theta$$

and

$$\int_{\underline{\theta}}^{p_j} (p_j - \theta) d\Lambda_i(\theta) = \int_{\underline{\theta}}^{p_j} \Lambda_i(\theta) d\theta.$$

(7) and (8) are used to obtain the last inequality. Since the last equation is equal to

$$E[q(\theta, \eta_i)(b - \theta) + X_0 - \beta u_A(\theta, \eta_i) \mid \eta_i],$$

we can conclude that the objective function of the problem is maximized at the choice of $\tilde{q}(\theta, \eta_i) = q(\theta, \eta_i)$, $\tilde{u}_A(\theta, \eta_i) = u_A(\theta, \eta_i)$ and $P(\cdot \mid \eta_i) = I_2(\eta_i)$. ■

Now we construct $\Lambda_i(\theta)$ such that (7) and (8) are satisfied for $i \in \{L, H\}$.

Case of $i = L$

Let us select

$$\Lambda_L(\theta) = \begin{cases} 0 & \text{for } \theta = \underline{\theta} \\ F_L(p^{NS}) & \text{for } \theta \in (\underline{\theta}, \bar{\theta}) \\ 1 & \text{for } \theta = \bar{\theta}. \end{cases}$$

It is evident that $\Lambda_L(\theta)$ is non-decreasing in θ with $\Lambda_L(\underline{\theta}) = 0$ and $\Lambda_L(\bar{\theta}) = 1$. Evidently (7) is satisfied, since $b - W_L(\theta : \beta) = p^{NS} - W_L(\theta : \beta)$ is positive on $[\underline{\theta}, p^{NS})$ and negative on $(p^{NS}, \bar{\theta}]$. With choice of $p_L = p_0 = p^{NS}$, $u_L = u_0 = L_H(p_H, p^{NS} : \beta)$ and $u_H = 0$,

$$\int_{\underline{\theta}}^{p_L} \Lambda_L(\theta) d\theta + u_L = \int_{\underline{\theta}}^{p_0} \Lambda_L(\theta) d\theta + u_0 = (p^{NS} - \underline{\theta}) F_L(p^{NS}) + L_H(p_H, p^{NS} : \beta)$$

and

$$\int_{\underline{\theta}}^{p_H} \Lambda_L(\theta) d\theta + u_H = (p_H - \underline{\theta}) F_L(p^{NS}).$$

By Lemma A2, for $p_H \in (p^{NS}, \bar{p}_H)$, the latter is greater than the former, implying that (8) is satisfied.

Case of $i = H$

Let us select $\Lambda_H(\theta)$ as follows:

$$\Lambda_H(\theta) = \begin{cases} 0 & \text{for } \theta = \underline{\theta} \\ F_H(p^{NS}) & \text{for } \theta \in (\underline{\theta}, p^*) \\ F_H(p_H) + (p_H - p^{NS}) \frac{f_H(p_H)}{\beta} & \text{for } \theta \in [p^*, \bar{\theta}) \\ 1 & \text{for } \theta = \bar{\theta} \end{cases}$$

for some $p^* \in (p^{NS}, p_H)$. p^* is specified explicitly below. Since $p^{NS} > p_H + \beta\frac{F_H(p_H)-1}{f_H(p_H)}$ by Lemma A2,

$$0 < F_H(p^{NS}) < F_H(p_H) + (p_H - p^{NS})\frac{f_H(p_H)}{\beta} < 1.$$

Thus, $\Lambda_H(\theta)$ is non-decreasing in θ with $\Lambda_H(\underline{\theta}) = 0$ and $\Lambda_H(\bar{\theta}) = 1$.

By the assumption that $h_H(\theta)$ and $l_H(\theta)$ are increasing in θ, $W_H(\theta : \beta)$ is also increasing in θ on each of two intervals $(\underline{\theta}, p^*)$ and $(p^*, \bar{\theta})$, even though it has a discontinuous decline at p^*. We can also check that $W_H(p^{NS} : \beta) = W_H(p_H : \beta) = p^{NS} = b$. Therefore, $b - W_H(\theta : \beta)$ takes positive value on each of $(\underline{\theta}, p^{NS})$ and (p^*, p_H), while negative value on each of (p^{NS}, p^*) and $(p_H, \bar{\theta})$. Thus

$$\int_{\underline{\theta}}^{\tilde{p}_H} [b - W_H(\theta : \beta)]dF_H(\theta)$$

takes a local maximum value at $\tilde{p}_H = p^{NS}$ and p_H. On the other hand, by this property of $b - W_H(\theta : \beta)$ on $[p^{NS}, p_H]$, we can find $p^* \in (p^{NS}, p_H)$ such that

$$\int_{p^{NS}}^{p_H} [b - W_H(\theta : \beta)]dF_H(\theta) = 0. \tag{9}$$

With this selection of p^*, the maximum value is achieved at both p^{NS} and p_H, implying (7) is satisfied. By the definition of $L_i(p, \tilde{p} : \beta)$,

$$\int_{p^{NS}}^{p_H} [b - W_H(\theta : \beta)]dF_H(\theta) = \beta[L_H(p_H, b : \beta) - \int_{p^{NS}}^{p_H} \Lambda_H(\theta)d\theta].$$

Thus, (9) can be also rewritten as

$$\int_{p^{NS}}^{p_H} \Lambda_H(\theta)d\theta = L_H(p_H, p^{NS} : \beta).$$

With $u_L = u_0 = L_H(p_H, p^{NS} : \beta)$, $u_H = 0$, and $p_L = p_0 = p^{NS}$, we obtain

$$\int_{\underline{\theta}}^{p_H} \Lambda_H(\theta)d\theta + u_H = \int_{\underline{\theta}}^{p_j} \Lambda_H(\theta)d\theta + u_j$$

for $j = 0, L$. This implies that (8) is satisfied. ■

Proof of Proposition 3

Provided in the text.

References

Andrianova, S., Melissas, N., 2008. Corruption, extortion and boundaries of the law. J. Law Econ. Organ. 25 (2), 442–471.

Asseyer, A., 2020. Collusion and delegation under information control. Theor. Econ. 15 (4), 1547–1586.

Ayres, I., 1997. The twin faces of judicial corruption: extortion and bribery. Denver Univ. Law Rev. 74, 1231–1254.

Cam, H., 1930. The Hundred and the Hundred Rolls. London Methuen and Co.

Celik, G., 2009. Mechanism design with collusive supervision. J. Econ. Theory 144 (1), 69–95.

Che, Y.K., Kim, J., 2009. Optimal collusion-proof auctions. J. Econ. Theory 144 (2), 565–603.

Dequiedt, V., 2007. Efficient collusion in optimal auctions. J. Econ. Theory 136 (1), 302–323.

Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2013. Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. Q. J. Econ. 128 (4), 1499–1545.

Faure-Grimaud, A., Laffont, J.J., Martimort, D., 2003. Collusion, delegation and supervision with soft information. Rev. Econ. Stud. 70 (2), 253–279.

Furnivall, J.S., 1956. Colonial Policy and Practice: A Comparative Study of Burma and Netherlands India. New York University Press.

Fudenberg, D., Tirole, J., 1991. Perfect Bayesian equilibrium and sequential equilibrium. J. Econ. Theory 53 (2), 236–260.

Hindricks, J., Keen, M., Muthoo, A., 1999. Corruption, extortion and evasion. J. Public Econ. 74 (3), 395–430.

Khalil, F., Lawarree, J., Yun, S., 2010. Bribery versus extortion: allowing the lesser of the two evils. Rand J. Econ. 41 (1), 179–198.

Klitgaard, R., 1988. Controlling Corruption. University of California Press.

Laffont, J.J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press, Cambridge, MA.

Mookherjee, D., 1997. Incentive reforms in developing country bureaucracies: lessons from tax administration. In: Pleskovic, B., Stiglitz, J. (Eds.), Proceedings of the Annual World Bank Conference on Development Economics. World Bank Press, Washington DC.

Mookherjee, D., Motta, A., Tsumagari, M., 2020. Consulting collusive experts. Games Econ. Behav. 122, 290–317.

Mookherjee, D., Tsumagari, M., 2022. Regulatory mechanism design with extortionary collusion. Working paper. https://people.bu.edu/dilipm/publications/HierarchyExtortCollWorkingPaperDec2022v1.pdf.

Nikaido, H., 1954. On von Neumann's minimax theorem. Pac. J. Math. 4 (1), 65–72.

Ortner, J., Chassang, S., 2018. Making corruption harder: asymmetric information, collusion, and crime. J. Polit. Econ. 126 (5), 2108–2133.

Quesada, L., 2003. Modeling collusion as an informed principal problem. In: Game Theory and Information. University Library of Munich, 0304002.

Tirole, J., 1986. Hierarchies and bureaucracies: on the role of collusion in organizations. J. Law Econ. Organ. 2 (2), 181–214.

Tirole, J., 1992. Collusion and the theory of organizations. In: Laffont, J.J. (Ed.), Advances in Economic Theory, Sixth World Congress, vol. 2. Cambridge University Press, Cambridge, pp. 151–206.

Vanutelli, S., 2020. From Lapdogs to Watchdogs: Random Auditor Assignment and Municipal Fiscal Performance in Italy. Working paper. Boston University.

von Negenborn, C., Pollrich, M., 2020. Sweet lemons: mitigating collusion in organizations. J. Econ. Theory 189, 105074.