Bribing and Signalling in Second Price Auctions*

Péter Eső† James Schummer‡

October 10, 2002

Abstract

We examine whether a two-bidder, second-price auction for a single good (with private, independent values) is immune to a simple form of collusion, where one bidder may bribe the other to commit to stay away from the auction (i.e. submit a bid of zero). First, we consider a situation in which only a bribe of a fixed size may be offered. There are precisely two equilibria in this extended game: a "bribing" and a "no-bribing" equilibrium. While the bribing equilibrium is naturally stable, the no-bribing equilibrium is shown to fail several standard refinements on out-of-equilibrium beliefs. Second, we consider the case in which bribes of any size may be offered. Robust equilibria in this situation involve low briber-types revealing themselves through the amount they offer, while high types "pool" by offering the same bribe. Only one such equilibrium involves a continuous offer strategy. Bribing equilibria in all cases lead to inefficiency.

Keywords Second-Price Auction, Collusion, Bribing, Signalling.

JEL classification numbers: C72 (noncooperative games), D44 (auctions), D82 (asymmetric information)

---

*We are grateful for conversations with David Austen-Smith, Sandeep Baliga, Eddie Dekel, Balázs Szentes, and seminar participants at the University of Wisconsin-Madison and Northwestern University. We thank Joel Sobel for detailed comments.

†Kellogg School of Management, MEDS Department, Northwestern University. Email: eso@kellogg.northwestern.edu.

‡Kellogg School of Management, MEDS Department, Northwestern University. Email: schummer@kellogg.northwestern.edu.

1 Introduction

Our goal is to examine the extent to which an auction mechanism is immune to a simple form of collusion in which one bidder may bribe another to leave the auction. Specifically, we consider a second-price or English auction where two buyers have private information about their valuations for a good. Before the auction begins, one of the buyers has the opportunity to offer the other a bribe in exchange for the other's commitment to remove himself from the auction (or bid zero). We analyze two versions of this model: one in which the amount of the bribe is exogenously fixed (e.g. representing a fixed "favor" whose transfer is undetectable), and one in which the bribe can be chosen to be any amount (e.g. offering money). With respect to a given equilibrium concept for this extended game, we examine whether the second-price auction is "bribe-proof" in the following sense.

We say that the auction is *strongly* bribe-proof if bribing does not occur in *any* equilibrium of the extended game. We show that the second-price auction fails this requirement under any reasonable equilibrium concept: under both fixed and variable bribes, there exists a robust equilibrium in which bribing occurs.

We say that the auction is *weakly* bribe-proof if *there exists* an equilibrium in which bribing does not occur. While there is a sequential equilibrium in which no bribe is offered, this equilibrium turns out not to be robust. We provide necessary and sufficient conditions such that the no-bribing equilibrium does not survive iterated deletion of dominated strategies (or the Intuitive Criterion) in the fixed bribes model. Furthermore, the no-bribing equilibrium always fails other common refinements.

The concept of bribe-proofness is a practical and reasonable requirement, as bribing agreements represent the simplest and crudest form of collusion. A bribing contract like ours is relatively easy to enforce; participation in the auction is often verifiable, and the contract does not rely on post-auction payments.

The bribing contracts we study certainly do not represent all possible collusive arrangements. However, the availability of even these can induce collusion and inefficiencies in the second-price auction. One interpretation of our results is that, in the private-values auction environment, no efficient and strategy-proof mechanism is resistant to very simple forms of bidder collusion, even if the buyers have incomplete information regarding each others' valuations.

1.1 Related literature

Bribing contracts have been analyzed by Schummer (2000) in the context of dominant strategy implementation. In a general collective decision problem, he calls a mechanism bribe-proof if, given player i's type, player j has no incentive to pay i to commit to misreport his type, even when j reports truthfully. Schummer (2000) shows that only constant mechanisms are bribe-proof. In this paper, we extend this type of analysis to a Bayesian setting, where players do not know each others' types, and where the decision problem of allocating an object is being solved with a second-price auction.

Our paper also contributes to a growing literature on collusion in auctions, including Graham and Marshall (1987), Mailath and Zemsky (1991), McAfee and McMillan (1992), and Marshall and Marx (2002).[1] These authors model collusion by assuming that a subset of buyers congregates before the auction, and play some kind of "collusive mechanism" or "knock-off auction." Graham and Marshall (1987) show that a group of bidders can collude in an incentive compatible and ex-ante budget balanced way by simply asking low-valuation bidders in the group to drop out of the auction. Payments are made to all group members before determining who should drop out, while after the auction, the group's high-valuation member makes a payment back to the group only if the manipulation produced ex-post gains for him. Mailath and Zemsky (1991) provide a more sophisticated mechanism that also achieves ex-post budget balance, and identify the optimal collusive contract subject

[1] Laffont and Martimort (2000) have a two-agent public goods setup where the modelling of collusion is similar to that of this literature.

to this constraint.

The main difference between our extended game and the way collusion is modelled in this literature is that we consider a different (and particular) bribing stage. Instead of the agents jointly designing a collusive side-contract, one of our agents is fixed as having the opportunity to offer a contract to the other agent.[2] This is important because in our model the "designer" of the mechanism, bidder j, has private information, and his goal is not the maximization of the joint surplus, but rather his own. The result of this difference is that in our game, signalling is an issue, and the bribing equilibrium is not efficient.

In previous work on bidding rings, the ring serves as a device to siphon profits from the seller to the ring members, and overall efficiency is not lost (under ex-ante symmetry) as a consequence. In our model, though, bribing leads to a loss in social surplus. We do not assert that our way of modelling collusion is better, but we think that it is an interesting alternative, especially, that inefficiencies arising from bribing have not been considered before.

1.2 Outline of Results

In Section 3 we start with a model in which the briber may only offer an exogenously fixed bribe amount b. We show that in this model, there are precisely two equilibria in pure strategies: (i) a bribing equilibrium in which high briber types offer the bribe, and low acceptor types accept it, and (ii) a no-bribing equilibrium in which the bribe is never offered.

Since the bribing equilibrium has full support on the action space, it is robust to the usual equilibrium refinements of signalling games. We argue that the no-bribing equilibrium, however, is not robust. First, we show that it fails the iterated deletion of dominated strategies if and only if the amount of the bribe is sufficiently large compared to a certain function of the distribution

[2] Furthermore, we restrict the set of available contracts to "bribing contracts," that is, a transfer from j to i conditional on i bidding zero.

of types.[3]

Second, regardless of the distribution of types, the no-bribing equilibrium does not survive standard equilibrium refinements (such as D1 or Perfect Sequential Equilibrium). It also fails an intuitive restriction on out-of-equilibrium beliefs that we introduce in Section 3.3. To briefly describe this refinement, we require that the support of i's out-of-equilibrium beliefs regarding the briber's type is restricted to a uniquely defined "briber's club"—a set of briber-types that is equivalent to the set of types for whom it is profitable to be associated with that set, as long as i's beliefs are restricted to that set. We show that beliefs on that support cannot be part of a no-bribing equilibrium.

In Section 4 we turn to the case of "endogenous bribe amounts," i.e., when the briber may choose to offer any amount b. The equilibrium behavior found in the case of a fixed bribe can be supported in this model as well, though sometimes only with unintuitive out-of-equilibrium beliefs. In particular, the no-bribing equilibrium is incompatible with standard refinements for signalling games.[4] We believe that the most plausible, interesting, and robust equilibrium is the unique bribing equilibrium in continuous and weakly monotonic strategies. It is a "mostly-separating" equilibrium, in the sense that any briber type below a certain threshold offers a unique amount b as a function of his type, while all types above the threshold offer the same amount. All bribes are accepted with positive probability, and the highest bribe is always accepted. The allocation of the good in this equilibrium is inefficient with positive probability.

Section 5 numerically illustrates our results for the uniform distribution of types. Proofs are collected in Appendix I, while a discussion of standard refinements is relegated to Appendix II.

[3] This necessary and sufficient condition holds if the distribution function is convex.

[4] In Appendix II, we show that in the variable bribes model, any no-bribing equilibrium fails Cho and Sobel's (1990) D1 criterion.

2 The Bribing Contract

Consider a second-price (Vickrey) auction for a single indivisible good, with two risk-neutral bidders i and j. The buyers have private valuations, $\theta_i, \theta_j \in [0, 1]$, drawn independently according to the same differentiable c.d.f. F. We assume that $0 < F'(x) < \infty$ for all $x \in [0, 1]$. Everything is commonly known except the valuations, which are privately known by the buyers who hold them.

We modify the second-price auction to model *bribing* in the following way. After the buyers learn their valuations, but before the auction starts, bidder j has an opportunity to offer a bribe b to bidder i in exchange for i's commitment not to bid. If i accepts the bribe, he is committed to making a bid of 0 in the auction; we are assuming that the bribing contract is enforceable. If i rejects the bribe or if j doesn't offer a bribe in the first place, then the game proceeds as a second-price auction.

We provide results for two cases: (Section 3) when b is given exogenously, so j decides whether to bribe but not how much to offer; and (Section 4) when j may also choose the amount of the bribe b. One interesting aspect of this game is that buyer j's decision whether or not to offer a bribe (and the amount offered) reveals information regarding his type. This signalling effect adds much complication to a Bayesian model, and makes out-of-equilibrium beliefs an important issue.[5] In a second-price auction, however, if a bribe is offered but it is declined, then the players' beliefs about each other's type becomes irrelevant, since bidders have an incentive to bid truthfully regardless of their information.

Formally, the game we describe above involves three stages: a stage where bidder j decides whether to offer a bribe, a stage where i decides whether to accept an offer (if made), and the second-price auction stage. In order to simplify the presentation, however, we do not explicitly model the bidders' behavior in the auction stage. We assume that bidders bid truthfully in

[5] In our model, out-of-equilibrium beliefs affect what would happen if j offered a bribe that is not expected.

the second-price auction, except of course when bidder i accepts a bribe, in which case he is forced to bid zero.[6] If a bribe b is offered and accepted, then the payoffs to i and j are b and $\theta_j - b$, respectively. Otherwise, the payoff to the bidder with the highest type is $\max(\theta_i, \theta_j) - \min(\theta_i, \theta_j)$, while the other bidder receives zero. We formalize the definitions of strategies and equilibrium concepts in each of the following two sections.

3 Fixed Bribe Amount

In this section we assume that the amount that j can offer, $b \leq \mathbf{E}(\theta_i)$, is exogenously fixed, and that the briber chooses only whether to offer it. One interpretation of this model is that j has a car, and on the day of the auction he tells i "Take the keys to my car and leave town."

In this model, a (pure) *strategy* for bidder j prescribes for each type θ_j a decision of whether to offer the bribe b. Hence it can be represented by the set $\mathbb{B} \subseteq [0, 1]$ of types that offer the bribe. A strategy for bidder i prescribes for each type θ_i a decision of whether to accept b if offered; it can be represented by the set $\mathbb{A} \subseteq [0, 1]$ of types that would accept the bribe if it were offered. We make the innocuous (see Proposition 1) assumption that these strategy-representing sets are measurable.

A *sequential equilibrium* is a pair of strategies $(\mathbb{A}, \mathbb{B})$ and a posterior belief distribution, μ, which satisfy the usual consistency and rationality conditions for each type.[7]

Some of our results involve equilibria whose description includes a partition of the set of types. As in many such games with a continuum of types, a pair of equilibria may exist which differ only in the behavior of a single

[6] This assumption is innocuous since our most interesting results concern equilibria in undominated strategies. It does, however, rule out equilibria (in weakly dominated strategies) in which bidder j threatens to bid the maximum amount in the second-price auction, forcing i (if he believes this) to accept the bribe regardless of his type. Since our emphasis is on when bribing equilibria are the only "reasonable" ones, our results are not weakened by this assumption.

[7] See Sect. 8.3, Fudenberg and Tirole (2000).

(borderline) type. In order to describe such equilibria more concisely, we introduce the following notation. For any $0 \leq a \leq 1$, we write $[0, a\rangle$ to mean "$[0, a]$ or $[0, a)$." Similarly, $\langle a, 1]$ means "$[a, 1]$ or $(a, 1]$." This notation facilitates the description of "essentially unique" equilibria, where certain types on interval boundaries may behave in indeterminate (and irrelevant) ways.

Our first result describes the structure of all sequential equilibria in the model with a fixed bribe b. Strategies are described by sets which are 2-partitions of $[0, 1]$.

Proposition 1 *In any sequential equilibrium, the set of types that offer a bribe is of the form $\langle B, 1]$ and the set of types that accept the bribe is of the form $[0, A\rangle$, where $B < 1$ implies $b < B < A \leq 1$.*

Proof: For a given equilibrium, denote the set of types that offer the bribe as $\mathbb{B}$, and the set of types that accept the bribe as $\mathbb{A}$. When $\mathbb{B}$ is non-empty, if player i accepts the bribe then it must exceed the profit he would get in the auction, given that $\theta_j \in \mathbb{B}$. In other words, if $\theta_i \in \mathbb{A}$ then

$$b \geq \mathbf{E}_{\theta_j}[(\theta_i - \theta_j)\mathbf{1}_{\{\theta_j \leq \theta_i\}} \mid \theta_j \in \mathbb{B}] \tag{1}$$

where $\mathbf{1}_X$ is the indicator function for event X. If this inequality holds for some θ_i then it holds for any $\theta_i' < \theta_i$. Therefore $\mathbb{A} = [0, A\rangle$. If $\mathbb{B}$ is empty then a similar argument (in which the posterior based on F is replaced by the out-of-equilibrium beliefs) shows that for any beliefs supporting the sequential equilibrium, $\mathbb{A}$ must be an interval.

To show that $\mathbb{B}$ is also an interval, define $B = \inf \mathbb{B}$. If $B = 1$ then we are done. Otherwise, since i can infer $\theta_j \geq B$ from the fact that the bribe was offered, he has an incentive to accept the bribe if his type is less than $B + b$. This follows because i's profit in the second-price auction is at most $\theta_i - B \leq b$. Therefore $A \geq \min\{1, B + b\} > B$.

For any $\theta_j \in \mathbb{B}$, the payoff from offering the bribe must be at least as great as his unconditional payoff in the second-price auction, that is,

$$F(A)(\theta_j - b) + \mathbf{E}_{\theta_i}[(\theta_j - \theta_i)\mathbf{1}_{(A<\theta_i\leq\theta_j)}] \geq \mathbf{E}_{\theta_i}[(\theta_j - \theta_i)\mathbf{1}_{(\theta_i\leq\theta_j)}]. \qquad (2)$$

Differentiating both the left and right hand sides,

$$\frac{\partial LHS(\theta_j)}{\partial\theta_j} = \max\{F(A), F(\theta_j)\} \geq F(\theta_j) = \frac{\partial RHS(\theta_j)}{\partial\theta_j}.$$

When $\theta_j < A$, the left hand side increases in θ_j strictly faster than the right hand side does. Therefore, for any $\theta_j \in \mathbb{B}$ for which $B \leq \theta_j < A$, and any $\theta'_j > \theta_j$, eqn. (2) holds strictly with respect to θ'_j. This implies $\theta'_j \in \mathbb{B}$, and therefore $\mathbb{B}$ is of the form $\langle B, 1]$. Furthermore, eqn. (2) cannot hold at $\theta_j = b$, hence $B > b$. □

3.1 The Bribing Equilibrium

Our next result states that regardless of the distribution and the amount of the bribe, an essentially unique bribing equilibrium exists. In it, high types offer the bribe while low types accept it. Since there is an overlap between these sets, inefficiency occurs with positive probability.

Proposition 2 *For any $b \in (0, \mathbf{E}(\theta_i)]$, there exists a sequential equilibrium in which bribing occurs. Moreover, all equilibria in which a bribe is offered with positive probability are essentially equivalent: there exist A^b, B^b such that in any equilibrium where bribing occurs, the sets of bribers and acceptors are $\langle B^b, 1]$ and $[0, A^b\rangle$, respectively.*

This equilibrium can be shown to be the unique one to satisfy Grossman and Perry's (1986) Perfect Sequential Equilibrium. Since both of j's actions are used in equilibrium, it clearly satisfies any reasonable refinement. We discuss refinements in more detail in Section 3.3 and Appendix II.

3.2 The No-Bribing Equilibrium

An equilibrium in which bribing does not occur (a "no-bribing equilibrium") can be supported when bidder i believes that only type $\theta_j = 0$ would offer a bribe. In this case, his optimal strategy is to accept the bribe when his type is such that $\theta_i \in [0, b\rangle$. Then, bidder j never could benefit from offering the bribe, hence bribing would not occur. These out-of-equilibrium beliefs are unreasonable, though, since for types $\theta_j < b$, offering the bribe is a strictly dominated strategy (as long as it is accepted with positive probability).

We examine two refinements that rule out such unreasonable beliefs. First, we consider iteratively deleting weakly dominated strategies. Proposition 3 provides a necessary and sufficient condition under which this refinement rules out no-bribing equilibria. In Section 3.3, we introduce an intuitive refinement that is in the spirit of Cho and Kreps' (1987) Intuitive Criterion, but slightly stronger. Proposition 4 shows that a no-bribing equilibrium must fail this refinement.[8]

To discuss these refinements, it is helpful to define the briber-type who would be indifferent between offering the bribe and not offering it, given that every acceptor type $\theta_i \in [0, 1]$ would accept the bribe.

Definition 1 For any $b \in [0, \mathbf{E}(\theta_i)]$ define θ^b to satisfy

$$\theta^b - b = \int_0^{\theta^b} (\theta^b - \theta_i)\, dF(\theta_i). \tag{3}$$

One can check that θ^b is unique and well-defined by this equation.

The following result considers the consequences of iteratively eliminating weakly dominated strategies. Since "order matters" when eliminating weakly dominated strategies, for simplicity we restrict attention to the case of eliminating *every* weakly dominated strategy in each round of deletion. We call this *maximal elimination of weakly dominated strategies*.

[8] In Appendix II, we show that no-bribing equilibria fail Cho and Sobel's (1990) D1 criterion and Grossman and Perry's (1986) concept of Perfect Sequential Equilibrium.

Proposition 3 *For all $b \in (0, \mathbf{E}(\theta_i))$, there exists a no-bribing sequential equilibrium that survives the iterated maximal elimination of weakly dominated strategies if and only if*

$$b > \mathbf{E}[\theta_i \mid \theta_i \leq \theta^b + b] \tag{4}$$

Condition (4) fails to hold, for example, if the distribution function is convex, in which case iterated dominance rules out the no-bribing equilibrium. On the other hand, for example, if $F(x) = x^\alpha$ with $0 < \alpha < 1$, then eqn. (4) holds for small b.

Remark 1 Reasoning similar to that used in the proof of Proposition 3 can be used to show that there is a no-bribing equilibrium satisfying Cho and Kreps' (1987) Intuitive Criterion if and only if eqn. (4) is satisfied. Roughly speaking, the Intuitive Criterion requires the acceptor to form out-of-equilibrium beliefs that place no probability on any briber type who could not hope to gain a payoff higher than his equilibrium payoff, as long as the acceptor plays some best response strategy. Since a best response for the acceptor must involve an interval $[0, A]$ of accepting types, no briber with type $\theta_j \leq \theta^b$ could hope to do better offering the bribe than he does when not offering it (as in equilibrium). Hence, (out-of-equilibrium) beliefs for the acceptor must have support only on $[\theta^b, 1]$, and a conclusion similar to that of Proposition 3 is reached. It may also be noted that in our model, the Intuitive Criterion is equivalent to the (stronger) iterated version of that condition, defined by Fudenberg and Tirole (1991, p. 449).

3.3 A Refinement on Out-of-Equilibrium Beliefs

When eqn. (4) holds, we construct a no-bribing equilibrium (in the proof of Proposition 3) by using out-of-equilibrium beliefs for bidder i that are "skewed downward" in the following sense. When j unexpectedly offers the bribe, i believes that j is very likely to have a type close to the *lowest* one that could conceivably offer the bribe after iterated deletion of weakly dominated strategies. Such beliefs run counter to the intuition (established by differentiating eqn. (2)) that if type θ_j has an incentive to offer a bribe, then so does any type $\theta'_j > \theta_j$. Although these skewed beliefs are permissible, one may find a sequential equilibrium whose existence *depends* on them to be less appealing than other, more robust equilibria.

Cho and Sobel's (1990) D1 criterion eliminates such beliefs by requiring that only types that are "most likely to gain" from deviating be given weight in i's posterior beliefs. It turns out (see Appendix II) that in a no-bribing equilibrium, the types of j that are most likely gain from offering the bribe are the ones whose expected payment in a second-price auction exceeds the bribe.[9] By inducing such beliefs, it is not hard to see that the no-bribing equilibrium is ruled out by this refinement.

These considerations motivate us to introduce what we consider to be a more intuitive refinement on out-of-equilibrium beliefs. It is somewhat in the spirit of Cho and Kreps' (1987) Intuitive Criterion. However, it turns out to have stronger consequence in our model, since a no-bribing equilibrium cannot satisfy it regardless of the distribution F.

To provide the motivation for our refinement, first consider an equilibrium in which a bribe is never offered by j—we generalize and formalize the definition below. Fix a set of "credible deviating types" $\mathcal{C} \subset [0,1]$, and consider the following speech by bidder j.

> "I hereby offer you the bribe, and inform you that my type is a member of $\mathcal{C}$. You should believe this because my type is in $\mathcal{C}$ if and only if I am better off making this speech (than in equilibrium) *for any* best response you play, consistent with believing my type is in $\mathcal{C}$. Furthermore, no other set of types $\mathcal{C}'$ can make a similar speech."

This speech, if true, creates a sort of self-fulfilling prophecy. If bidder i believes $\theta_j \in \mathcal{C}$, then he should form posterior beliefs with support restricted

[9] Such types exist because, for example, the expected payment of type $\theta_j = 1$ is $\mathbf{E}(\theta_i)$, which is larger than b.

to $\mathcal{C}$, and play a corresponding best response. In that case, the only briber types who would profit from making this speech *for any such best response* are those in $\mathcal{C}$. Therefore, bidder i would be justified in believing $\theta_j \in \mathcal{C}$.

The uniqueness of $\mathcal{C}$ merely weakens our requirement, and ensures that bidder i need not worry about why one speech was made instead of another.

If such a speech can be made truthfully, we say that this set of types $\mathcal{C}$ breaks the no-bribing equilibrium. Our result in the fixed-bribe model is that a no-bribing equilibrium is always broken by a unique such set of types. For more general situations and games (e.g. with more than two actions available to the sender in a signalling game), such a speech is defined with respect to a given equilibrium, a set of types, and an out-of-equilibrium action.

Remark 2 Our speech states that $\theta_j \in \mathcal{C}$ whenever θ_j gains from making the speech *for any* corresponding best response by bidder i. The Intuitive Criterion can be described by a similar speech in which $\theta_j \in \mathcal{C}$ whenever θ_j gains *for some* corresponding best response by bidder i. Perfect Sequential Equilibrium also can be described similarly, where $\theta_j \in \mathcal{C}$ whenever θ_j gains when i's best response is formed only with respect to Bayesian updating on $\mathcal{C}$, i.e. i's posteriors are not allowed to vary. Uniqueness of the set $\mathcal{C}$ is not required, however, in the definition of Perfect Sequential Equilibrium (see van Damme's (1991) remarks, p. 291).[10]

To formalize our concept, we first give a loose description in terms of a general signalling game. Then, we more formally interpret the definition to our specific model. A similar formal interpretation is given for the variable-bribe model in Section 4.2.

Consider a 2-agent game in which a Sender has an unknown type and sends a (possibly costly) message to a Receiver. Suppose an equilibrium exists in which some message m is never sent. We say that a closed, measurable set of (Sender's) types, $\mathcal{C}$, is a *set of credible deviating types that breaks the equilibrium with message* m when the following hold.

[10] It is easy to show, as we do in Appendix II, that the no-bribing equilibrium is not a Perfect Sequential Equilibrium.

1. Let $D(\mathcal{C}, m)$ be the set of Sender types who are strictly better off than in the equilibrium whenever the Receiver plays any best response to m consistent with posterior beliefs restricted to $\mathcal{C}$.

2. We require $\mathcal{C} = \text{closure}(D(\mathcal{C}, m))$.

3. We require that there exist no other (measurable) set of types $\mathcal{C}' \neq \mathcal{C}$ for which $\mathcal{C}' = \text{closure}(D(\mathcal{C}', m))$.

In the fixed-bribe model of this section, there are only two types of equilibria: the essentially unique bribing equilibrium discussed in Section 3.1 and equilibria in which bribing does not occur. Since the former has no out-of-equilibrium actions on the part of bidder j, for the sake of brevity we formalize our definition only with respect to the no-bribing equilibrium. Furthermore, briber j has only two "messages" in this game, and "offer b" is the one that is not played in this equilibrium.

For any closed set of types $\mathcal{C}$, in order to calculate $D(\mathcal{C}) \equiv D(\mathcal{C}, \text{offer})$, we need to determine bidder i's set of best responses when he believes $\theta_j \in \mathcal{C}$. Analogous to the result in Proposition 1, a best response for i is to accept the bribe whenever $\theta_i \in [0, A\rangle$, where A is determined by some beliefs of i over $\mathcal{C}$. In fact, for such beliefs, G, A is such that

$$b = G(A)[A - E(\theta_j \mid \theta_j \leq A)]$$

(see eqn. (1)) where the expectation is with respect to G.

Let the set of such best responses be parameterized by

$$\mathbb{A}(\mathcal{C}) = \{A : b = G(A)[A - E(\theta_j \mid \theta_j \leq A)] \text{ for some beliefs } G \text{ over } \mathcal{C}\}$$

i.e. the set of such best responses is $\{[0, A\rangle\colon\ A \in \mathbb{A}(\mathcal{C})\}$. The set of types who would want to make the speech described earlier can now be defined as follows.[11]

$$D(\mathcal{C}) = \{\theta_j \in [0,1] : \text{inequality (2) holds strictly } \forall A \in \mathbb{A}(\mathcal{C})\}$$

Finally, in the fixed-bribe model of this section, a set of types $\mathcal{C}$ is a *credible deviating set* if it is the unique set for which $\mathcal{C} = \text{closure}(D(\mathcal{C}))$.

Lemma 1 *With respect to a no-bribing equilibrium, a set of credible deviating types must be of the form $\mathcal{C} = [c, 1]$, where type $\theta_j = c$ is indifferent between offering the bribe and not when i accepts with types $\theta_i \in [0, c+b\rangle$. In other words,*

$$F(c+b)(c-b) = \int_0^c (c-x)\, dF(x). \tag{5}$$

Our result is that a no-bribing equilibrium cannot be robust to the kind of speech described above, since a (unique) set of credible deviating types always exists.

Proposition 4 *For any $b < \mathbf{E}(\theta_i)$ and any distribution F, any no-bribing sequential equilibrium is broken by a unique set of credible deviating types who offer the bribe.*

Proof: Using the same techniques as in the proof of Proposition 1, eqn. (5) can be uniquely solved for c. □

Hence, only the bribing equilibrium of Section 3.1 survives.

4 Variable Bribes

In this section, we examine the model in which j may offer any amount b to bidder i. As a simplification, we equate the act of offering a bribe of $b = 0$ with the act of offering no bribe.[12] Therefore, a strategy for j is simply a

[11] We omit from the notation the label for the "message" of offering the bribe.

[12] Under any reasonable equilibrium concept, this assumption changes nothing in the analysis.

function mapping types into offers, $b\colon [0,1] \to \mathbb{R}_+$. A strategy for i specifies a measurable set of accepting types for each offer $b \in \mathbb{R}_+$, $\mathbb{A}(b) \subseteq [0,1]$. A sequential equilibrium is defined analogously to the previous section (with i's beliefs over types θ_j conditional on receiving any offer $b \in \mathbb{R}_+$).

Certain results from the previous section carry over to this one. In particular, bidder i's equilibrium strategies must be such that any offer $b \in \mathbb{R}_+$ is accepted by sets of the form $\mathbb{A}(b) = [0, A(b)\rangle$. For bidder j, Proposition 1 generalizes in the following way.

Lemma 2 *In any sequential equilibrium, j's strategy $b(\theta_j)$, is weakly monotonic in θ_j.*

From this, it follows that in any equilibrium, if two amounts $b, b' \in \mathbb{R}_+$ are offered in equilibrium, then $b > b'$ implies $A(b) > A(b')$, where $A()$ defines i's strategy as above.

The type of equilibrium behavior described in the fixed-bribe model can be supported in this model with an appropriate specification of (out-of-equilibrium) beliefs for i. For example, for any given $b < \mathbf{E}(\theta_i)$, the bribing equilibrium described in Proposition 2 can be extended to this model by specifying that whenever a different bribe $b' \neq b$ is offered, i believes that $\theta_j = 0$ with probability 1. The no-bribing equilibrium described in Section 3.2 applies with similar beliefs. Such beliefs are, of course, unappealing, and do not survive typical refinements used in signalling games with continuous type spaces.

On the other hand, there may exist an equilibrium in which j's strategy $b()$ is continuous. Under the assumption that F is log concave, we will prove that there is such equilibrium, and that it is unique (up to the specification of i's out-of-equilibrium behavior). Furthermore, under our refinement, any equilibrium bribing function must at least partially agree with this continuous function.

For the remainder of this section, we make the (widely used) assumption that F is log concave: $d[F(\theta_j)/F'(\theta_j)]/d\theta_j \geq 0$.[13] Bagnoli and Bergstrom (1989) provide an extensive list of distributions that are log concave.

4.1 Continuous Equilibrium Bribing Function

In the continuous bribing equilibrium, $b()$ is strictly increasing on some interval $[0, \bar{\theta})$, and is constant on $[\bar{\theta}, 1]$. Therefore, if i receives some offer $b(\theta_j) < b(\bar{\theta})$, then j's type θ_j is perfectly revealed. In this case, i accepts the offer only if the bribe $b(\theta_j)$ exceeds his (perfectly anticipated) payoff in the auction, $\theta_i - \theta_j$, i.e. when $\theta_i < \theta_j + b(\theta_j)$.

In order for j to have the incentive to reveal his type (e.g. not to pretend to be a slightly higher type), a local incentive compatibility condition must be satisfied. An increase in the amount of bribe offered must be exactly offset by the increase in the set of types θ_i who would accept it. This leads to a differential equation (6) characterizing the bribing function.

Proposition 5 *Suppose F is log concave. In any sequential equilibrium in which bribing occurs, if j's bribing strategy function $b()$ is continuous, then it is the unique equation to satisfy $b(0) = 0$ and*

$$b'(\theta_j) = \begin{cases} \dfrac{F'(\theta_j + b(\theta_j))(\theta_j - b(\theta_j))}{F(\theta_j + b(\theta_j)) - F'(\theta_j + b(\theta_j))(\theta_j - b(\theta_j))} & \text{if } \theta_j + b(\theta_j) < 1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Conversely, there exists a sequential equilibrium in which j's (continuous) strategy $b()$ is described by eqn. (6), with $b(0) = 0$.

From eqn. (6), it follows that $b()$ is strictly increasing up to some $\bar{\theta}$, after which it is constant, where $\bar{\theta} + b(\bar{\theta}) = 1$.

The equilibrium is robust to any reasonable refinement of out-of-equilibrium beliefs: The only out-of-equilibrium bribe that can occur is $b > b(\bar{\theta})$. Even if

[13] Fudenberg and Tirole (2000) relate log concavity to a monotone hazard rate condition.

all types θ_i accept this offer, no θ_j could benefit from offering it because all types θ_i already accept the smaller (equilibrium) bribe $b(\bar{\theta})$.

When F is not log concave, a continuous bribing function b satisfying (6) may or may not exist. Tedious difficulties arise when F is such that $b'(\theta_j) = \infty$ for some θ_j. Without the log concavity assumption, it may be that any $b()$ satisfying eqn. (6) is discontinuous, in which case there is no sequential equilibrium where j has a continuous strategy. Intuitively, some type θ_j may see "increasing returns" from increasing the amount of his bribe, if his increased expenditure is more than offset by the increase in the density of types θ_i that accept the higher bribe.

Even without log concavity, the equilibrium payoffs to the bidders must be continuous. While this is not a surprising result, we state it here formally, as it is used to prove a later result. To do so, first define j's payoff from offering b when his type is θ_j and i's strategy is $A()$ as

$$\pi(b, \theta_j) = F(A(b))(\theta_j - b) + \mathbf{1}_{\{\theta_j > A(b)\}} \int_{A(b)}^{\theta_j} (\theta_j - x) dF(x). \tag{7}$$

With respect to a given pair of equilibrium strategies for i and j, $A()$ and $b()$, denote j's equilibrium payoff by

$$\pi^e(\theta_j) \equiv \pi(b(\theta_j), \theta_j).$$

Lemma 3 *Let $b()$ and $A()$ be defined with respect to given sequential equilibrium strategies. The briber's equilibrium profit function $\pi^e()$ is continuous in θ_j.*

4.2 Refinements under Variable Bribes

Our main result in this section describes the briber's strategy in any equilibrium that satisfies the refinement introduced in Section 3.3. Such bribing functions $b()$ must agree with the function described in Proposition 5 (eqn. (6)) on some interval $[0, \hat{\theta}_j\rangle$, and remain constant afterwards. Furthermore, there is a discontinuity at $\hat{\theta}_j$ unless $b()$ coincides with the continuous strategy described in Proposition 5.

Before presenting that result, we formalize the concept of our refinement for this variable-bribe model. In the fixed-bribe model, there could be at most one out-of-equilibrium action for bidder j. Here, our refinement not only rules out no-bribing equilibria, but also rules out some equilibria in which bribes are offered. In formalizing the requirement, we need to account for the possible multiplicity of out-of-equilibrium actions (offers) that bidder j may use.

Fix equilibrium strategies for i and j, defining functions $A()$ and $b()$ as above. For any measurable (and closed) set of types $\mathcal{C}$, and offer $b \in \mathbb{R}_+$, we wish to determine the set of types θ_j who are better off offering b (than in equilibrium) whenever i believes $\theta_j \in \mathcal{C}$. To determine this set, we need to determine bidder i's set of best responses to this offer when he believes $\theta_j \in \mathcal{C}$. As in Section 3.3, we can parameterize this set of i's best responses to the offer of b as

$$\mathbb{A}(\mathcal{C}, b) = \{A : b = G(A)[A - E(\theta_j \mid \theta_j \leq A)] \text{ for some beliefs } G \text{ over } \mathcal{C}\}$$

i.e. "accept b if and only if $\theta_i \in [0, A\rangle$" is a best response under beliefs restricted to $\mathcal{C}$ if and only if $A \in \mathbb{A}(\mathcal{C}, b)$.

The set of briber types θ_j who would strictly want to offer b (and make the speech described earlier) can now be defined as

$$D(\mathcal{C}, b) = \{\theta_j \in [0,1] : F(A)(\theta_j - b) + \mathbf{E}_{\theta_i}((\theta_j - \theta_i)\mathbf{1}_{(A<\theta_i \leq \theta_j)}) > \pi^e(\theta_j) \\ \forall A \in \mathbb{A}(\mathcal{C}, b)\}$$

A set of types $\mathcal{C}$ is a *credible deviating set for offer* b if it is the unique set (given b) for which $\mathcal{C} = \text{closure}(D(\mathcal{C}, b))$.[14]

Our result is that if an equilibrium is such that no credible deviating set

[14]We require the uniqueness of $\mathcal{C}$ only for *some* offer b. There may exist another credible deviating set for some other offer b'.

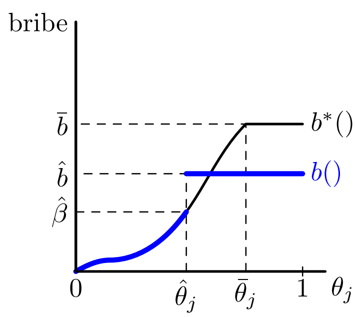

Figure 1: The structure of an equilibrium bribing function $b()$ under our refinement. The function described in Proposition 5 is labelled $b^*()$.

exists for any $b \in \mathbb{R}_+$, then j's strategy must agree with the one described in Proposition 5 for low types, and must involve "pooling" for all other types. See Figure 1.

Proposition 6 *Assume F is log concave. Suppose a sequential equilibrium satisfies our refinement in the variable bribes model: there exists no out-of-equilibrium offer $b \in \mathbb{R}_+$ and set of types $\mathcal{C}$ such that $\mathcal{C}$ is a credible deviating set for offer b. Denote the bribing function described in Proposition 5 (eqn. (6)) by $b^*()$.*

Then bidder j's strategy $b()$ is such that for some $\hat{\theta}_j \leq \bar{\theta}$,

$$b(\theta_j) = \begin{cases} b^*(\theta_j) & \text{if } \theta_j < \hat{\theta}_j \\ \hat{b} \equiv \hat{\theta}_j - F(\hat{\theta}_j + \hat{\beta})(\hat{\theta}_j - \hat{\beta}) & \text{if } \hat{\theta}_j < \theta_j \end{cases}$$

where $\hat{\beta} = \lim_{\theta_j \nearrow \hat{\theta}_j} b(\theta_j)$. Furthermore, $\mathbf{E}[\theta_j \mid \theta_j \geq \hat{\theta}_j] + \hat{b} \geq 1$, implying both $\hat{\theta}_j > 0$ and that $\hat{b}$ is accepted by all types θ_i.

Conversely, any function $b()$ satisfying these conditions is part of a sequential equilibrium satisfying our refinement.

5 Numerical Example: Uniform Distribution

In this section, we illustrate some of our results with a numerical example in which the valuations are drawn from a uniform distribution. We consider the case of a fixed bribe amount b.

As a preliminary, we derive θ_i's expected payoff in the SPA conditional on $\theta_j \geq B$.

$$\mathbf{E}[\pi_i(\theta_i \mid \theta_j \geq B)] = \frac{\int_B^{\theta_i}(\theta_i - x)\,dx}{1-B} = \frac{1}{2}\frac{(\theta_i - B)^2}{1-B} \quad \text{if } \theta_i > B, \tag{8}$$

and 0 if $\theta_i \leq B$. Note that the unconditional profit is $\mathbf{E}[\pi_i(\theta_i \mid \theta_j \geq 0)] = \int_0^{\theta_i}(\theta_i - x)\,dx = \frac{1}{2}\theta_i^2$.

First, we verify that in the case of uniform distributions the no-bribing equilibrium is ruled out by iterated dominance. Given $b > 0$, we define θ^b according to eqn. (3) that is, $\theta^b - b = \frac{1}{2}(\theta^b)^2$, so $b = \theta^b - \frac{1}{2}(\theta^b)^2$. By Proposition 2, the no-bribing equilibrium is dominated if and only if eqn. (4) holds, i.e., $b < \frac{1}{2}(\theta^b + b)$ for the uniform distribution. Since

$$b \equiv \theta^b - \frac{1}{2}(\theta^b)^2 < \theta^b - \frac{1}{4}(\theta^b)^2 \equiv \frac{1}{2}(\theta^b + b),$$

eqn. (4) holds and the no-bribing equilibrium is dominated.

Second, we calculate the fixed-bribe equilibrium (b, B, A) where $A = 1$. From eqns. (10) and (8) at $\theta_i = 1$,

$$b = \frac{1}{2}(1 - B).$$

From eqn. (9), using $A = 1$,

$$B - b = \int_0^B (B - x)\,dx = \frac{1}{2}B^2.$$

Adding these two and rearranging, we get $B^2 - 3B + 1 = 0$, which yields $B = \frac{3-\sqrt{5}}{2} \approx 0.382$, and in turn $b \approx 0.309$. In the equilibrium where all θ_i

b	B	A
.01	.01036	.15104
.05	.05403	.36160
.15	.17208	.67046
.20	.23505	.78820
.25	.30068	.89200
.30	.36925	.98443

Table 1: Equilibrium values of B and A for various bribe levels b, under the uniform distribution.

accept, the bribe is $b \approx 0.309$, and $\theta_j \geq B \approx 0.382$ offer it.

There is a continuum of other "semi-pooling" equilibria (b, B, A), which solve the following counterparts of eqns. (10) and (9).

$$(1 - B)b = \frac{1}{2}(A - B)^2$$
$$A(B - b) = \frac{1}{2}B^2.$$

Some solutions are given in Table 1.

Finally, the "mostly-separating" equilibrium of Proposition 5 is even easier to compute. Since the likelihood-ratio function is $L(x) \equiv F'(x)/F(x) = 1/x$, the differential equation for $b(\theta_j)$—eqn. (6)—simplifies to

$$b'(\theta_j) = \frac{[\theta_j - b(\theta_j)]/[\theta_j + b(\theta_j)]}{1 - [\theta_j - b(\theta_j)]/[\theta_j + b(\theta_j)]} = \frac{\theta_j - b(\theta_j)}{2b(\theta_j)},$$

which admits a linear solution, $b(\theta_j) = \frac{1}{2}\theta_j$ (the initial condition is $b(0) = 0$). This bribe function is valid for $\theta_j \in [0, \frac{2}{3}]$; $b(\theta_j) \equiv \frac{1}{3}$ for all $\theta_j \in [\frac{2}{3}, 1]$. In the "separating" equilibrium j uses the bribe function

$$b(\theta_j) = \begin{cases} \frac{1}{2}\theta_j & \text{if } \theta_j \in [0, \frac{2}{3}) \\ \frac{1}{3} & \text{if } \theta_j \in [\frac{2}{3}, 1] \end{cases} .$$

Clearly, the briber is better off in the bribing equilibrium than in the no-bribing equilibrium because his surplus in the no-bribing equilibrium (ordinary SPA) would be equal to that in a first-price auction where he would bid $\theta_j/2$ (instead of bribing with that amount) and win much less often.

6 Conclusion

We have examined a simple, specific form of collusion among two bidders in a second price auction, where one of the bidders is permitted to pay the other to commit to leave the auction (bid zero). Regardless of whether the bribing agent is permitted to offer only an exogenously fixed payment or is permitted to choose any payment, a robust equilibrium exists in which bribing occurs. In an equilibrium for the latter case, a bribe is offered with probability one. In these equilibria, the object is allocated inefficiently with positive probability.

Equilibria in which bribing does not occur are not robust to intuitive refinements. Therefore, depending on the solution concept an auctioneer uses, he may be unable to rationalize the use of this auction even on the basis of the *existence* of a collusion-free equilibrium.

Our approach differs from much of the collusion literature in a few ways. Foremost, we do not model a "collusion design" problem for the agents. In that literature, it is typical to assume that an uninterested third party designs and administers a revelation mechanism, making or receiving payments from bidders based on various information.[15] The third-party design assumption is a way to escape the issue of information transmission in the design stage: If we make the more-realistic assumption that bidders already have some idea about their types at the design stage, then when a bidder proposes the use of a particular collusive mechanism, information about his type could be inferred from that proposal.[16]

[15] See Marshall and Marx (2002) for a case-by-case analysis of the types of post-auction information that could be used.

[16] See Jackson and Wilkie (2001) for one approach to modelling this issue, where agents simultaneously propose and commit to contingent transfer mechanisms, and final payments

This type of inference is a part of what we model. In our model, there is common knowledge about whether agent j desired to seek collusion, while that information exists about i whenever j makes the offer. Analogously to the way information can be lost in moving from an extensive-form game to a normal-form game, a collusive revelation mechanism may or may not "naturally" model the information transmission that occurs when the "uninterested third party" does not actually exist.

We believe that our approach to modelling collusion can provide useful insights in more general settings as well. Our game is decidedly simple in that we consider a dominance-solvable mechanism (a second-price auction), where informational inferences do not directly affect the strategic behavior of any given type. We also restricted attention to two players, which greatly simplifies calculations without losing the essence of the bribing-signalling game.

Appendix I: Proofs

Proof of Proposition 2: For any $b \in (0, \mathbf{E}(\theta_i)]$, Proposition 1 implies that in any bribing equilibrium, briber and acceptor types are of the form $\langle B, 1]$ and $[0, A\rangle$. Since $B \geq b$, standard continuity arguments imply that type $\theta_j = B$ must be indifferent between offering the bribe and not, i.e., by eqn. (2) and $B \leq A$,

$$F(A)(B-b) = \int_0^B (B-x)\, dF(x). \tag{9}$$

Note that this holds even if $B = 1$ since bribing is occurring by assumption.

Also, either $A = 1$, or type $\theta_i = A$ is indifferent between accepting the bribe and not. By eqn. (1),

$$b \geq \frac{\int_B^A (A-x)\, dF(x)}{1-F(B)}, \tag{10}$$

are the sum of those proposals.

and if $A < 1$ then eqn. (10) holds with equality. (From Proposition 1, if $B = 1$ then $A = 1$ and eqn. (10) becomes $b \geq 0$; if $B < 1$ then $A > B$ and the right-hand side is positive.)

It is helpful to define the following functions for $A, B \in [0, 1]$, $A \geq B$.

$$b_1(A, B) = B - \frac{\int_0^B (B - x)\, dF(x)}{F(A)},$$
$$b_2(A, B) = \frac{\int_B^A (A - x)\, dF(x)}{1 - F(B)}.$$

Observe that since $B > 0$, $b_1(A, B) < B$.

To prove the existence of an equilibrium, we need to find A, B such that eqns. (9) and (10) hold, that is, $b = b_1(A, B) \geq b_2(A, B)$, with equality if $A < 1$.

Define

$$H = \{(A, B) : A \geq B,\ b_1(A, B) - b_2(A, B) \geq 0, \text{ with equality if } A < 1\}.$$

We claim that (i) for all B, there exists A such that $(A, B) \in H$, and (ii) for all $A < 1$, there exists a unique B such that $(A, B) \in H$.

To see (i), first note that $(0, 0) \in H$. For $B > 0$, $b_1(B, B) - b_2(B, B) = E[\theta_i \mid \theta_i \leq B] - 0 > 0$ so either $(1, B) \in H$ or by continuity, $(A, B) \in H$ for some $A \in (B, 1)$. To see (ii) for $A \in (0, 1)$, note that $b_1(A, 0) - b_2(A, 0) = 0 - \int_0^A (A - x)\, dF(x) < 0$, while $b_1(A, A) - b_2(A, A) > 0$. Continuity implies that $b_1(A, B) - b_2(A, B) = 0$ for some $B \in (0, A)$, hence $(A, B) \in H$. Furthermore, this B is unique because $b_1(A, B) - b_2(A, B)$ is strictly decreasing in B (see eqns. (12) and (14) below).

Define the correspondence $h\colon [0, 1] \twoheadrightarrow [0, 1]$ such that $h(A) = \{B : (A, B) \in H\}$. By (ii), h is non-empty and if $A < 1$ then h is single-valued. It can be shown (e.g., by an application of the Maximum Theorem) that h is upper hemi-continuous. Therefore, for $A < 1$, $h(A)$ is a continuous function, and its graph, H, is connected. Define $\hat{B} = \lim_{A \uparrow 1} h(A) \in h(1)$. An example

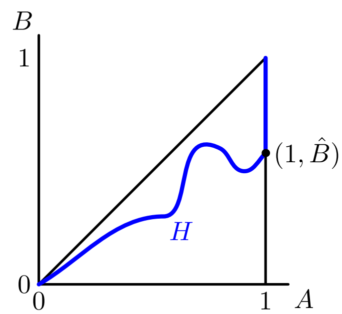

Figure 2: The set H in the proof of Proposition 2.

of h appears in Figure 2.

By differentiating b_1 and b_2, we find that for $0 < B < A \leq 1$,

$$\frac{\partial b_1}{\partial A} = \frac{F'(A)}{F(A)^2} \int_0^B (B - x)\, dF(x) > 0, \tag{11}$$

$$\frac{\partial b_1}{\partial B} = 1 - \frac{F(B)}{F(A)} > 0, \tag{12}$$

$$\frac{\partial b_2}{\partial A} = \frac{F(A) - F(B)}{1 - F(B)} > 0, \tag{13}$$

$$\frac{\partial b_2}{\partial B} = \frac{\left\{ \int_B^A (A - x)\, dF(x) - (A - B)(1 - F(B)) \right\} F'(B)}{(1 - F(B))^2} < 0 \tag{14}$$

where the last inequality follows by $\int_B^A (A - x)\, dF(x) < \int_B^A (A - B)\, dF(x) = (A - B)(F(A) - F(B)) \leq (A - B)(1 - F(B))$.

Consider any $A' > A$. If $h(A') > h(A)$ then by eqns. (11) and (12) we have $b_1(A', h(A')) > b_1(A, h(A))$. If $h(A') < h(A)$ then by eqns. (13) and (14) we have $b_2(A', h(A')) > b_2(A, h(A))$, which implies $b_1(A', h(A')) \geq b_2(A', h(A')) > b_2(A, h(A)) = b_1(A, h(A))$. Therefore, on $A \in [0, 1)$, $b_1(A, h(A))$ is strictly increasing in A.

By continuity, $b_1(1, \hat{B}) = b_2(1, \hat{B})$. Therefore, eqns. (12) and (14) imply that $h(1) = [\hat{B}, 1]$, and $b_1(1, B)$ is strictly increasing in B.

Therefore, $b_1(A, B)$ is strictly increasing on H in the sense that if $A < 1$ then $b_1(A, B)$ is strictly increasing in A, and if $A = 1$ then $b_1(A, B)$ is strictly increasing in B. Since $b_1(0, h(0)) = 0$ (by $h(0) = 0$) and $b_1(1, 1) = \mathbf{E}(\theta_i)$, the strict monotonicity and continuity of $b \equiv b_1(A, B)$ along H implies that there exists a one-to-one mapping between $b \in [0, \mathbf{E}(\theta_i)]$ and $(A, B) \in H$, i.e. for all $b \in [0, \mathbf{E}(\theta_i)]$, there exist unique A, B solving eqns. (9) and (10). $\square$

Proof of Proposition 3: Recall that we assume that players bid their valuations in the second-price auction, if it is ever reached (which is consistent with deleting dominated strategies). We begin the process of iteratively deleting maximal sets of weakly dominated strategies whether or not eqn. (4) holds.

Round 1. In the first round of elimination, we delete any of j's strategies that prescribe type $\theta_j \leq b$ to offer the bribe. This is clear because by offering the bribe, such a type can only obtain a negative payoff if the bribe is accepted. For bidder i, we delete any strategy that prescribes type $\theta_i \leq b$ to reject the bribe because such a type cannot obtain a payoff higher than b in the second-price auction. It is straightforward to check that no other strategies can be eliminated in the first round.

Round 2. Subject to the first round of elimination, we delete any strategy for the briber that prescribes $\theta_j \leq \theta^b$ to offer the bribe. To see this, denote any set of acceptors by $\mathbb{A} \subseteq [0, 1]$. From the first round, we must have $[0, b] \subseteq \mathbb{A}$. If $\theta_j \leq \theta^b$ offers the bribe then his profit would be

$$\mathbf{E}_{\theta_i}[(\theta_j - b)\mathbf{1}_{\{\theta_i \in \mathbb{A}\}} + (\theta_j - \theta_i)\mathbf{1}_{\{\theta_i \in [0,\theta_j] \setminus \mathbb{A}\}}] \leq \theta_j - b$$

where the inequality holds because $\theta_i > b$ for all $\theta_i \notin \mathbb{A}$. Since the LHS of eqn. (3) increases faster in θ^b than the RHS does, we have $\theta_j - b \leq \int_0^{\theta_j} (\theta_j - \theta_i)\, dF(\theta_i)$ for all $\theta_j \leq \theta^b$. Furthermore, for some admissible $\mathbb{A}$, the inequality is strict. Therefore for $\theta_j \leq \theta^b$ not offering the bribe dominates offering the bribe.

Continuing the second round of elimination, we delete any acceptor strategy that prescribes $\theta_i \leq 2b$ to reject the bribe. This follows because for any admissible briber strategy, the bribe is offered only by types $\theta_j > b$, limiting the acceptor's SPA-payoff to no more than $\theta_i - b$. It is again straightforward to check that no other strategies can be eliminated in the second round.

Round 3. Similarly, the acceptor's strategies that we delete in the third round of elimination are precisely those that prescribe $\theta_i \leq \theta^b + b$ to reject the bribe, because the briber's type is greater than θ^b if the bribe is offered.

Non-existence. As the first of two cases, suppose that eqn. (4) does not hold. Let $\theta_j = \min\{1, \theta^b + b\}$. If $\theta_j = 1$ then his payoff is $1 - b$ from offering the bribe and is $1 - \mathbf{E}(\theta_i)$ from not offering the bribe, so θ_j strictly prefers to bribe. If $\theta_j = \theta^b + b < 1$, then his payoff is at least $F(\theta_j)(\theta_j - b)$ from offering the bribe (because each $\theta_i \leq \theta^b + b$ accepts the bribe according to previous rounds of strategy deletion), and his payoff is $F(\theta_j)(\theta_j - \mathbf{E}[\theta_i \mid \theta_i \leq \theta_j])$ from not offering the bribe. Since eqn. (4) does not hold, for any admissible acceptor strategy, type θ_j's payoff from offering the bribe is weakly greater than that from not offering the bribe. Furthermore, since $\theta^b + b < 1$, this inequality is strict when i's strategy is to accept the bribe with any type. We conclude that for $\theta_j = \min\{1, \theta^b + b\}$, offering the bribe weakly dominates not offering the bribe, therefore the no-bribing equilibrium does not survive the iterated maximal elimination of weakly dominated strategies.

Existence. Second, suppose that eqn. (4) holds. We show that no briber strategies can be eliminated in the third round of deletion and that the no-bribing equilibrium can be supported. If the set of acceptors is exactly $[0, 1]$ then for all $\theta_j > \theta^b$, offering the bribe is strictly better than not offering it (because θ^b is indifferent and the LHS of eqn. (3) increases faster in θ^b than the RHS does).

On the other hand, if the set of acceptors is exactly $[0, \theta^b + b]$, which is also admissible even after the third round of deletion, then, for all $\theta_j > \theta^b$, offering the bribe is strictly worse than not offering it. To see this, consider the payoff difference between not offering and offering the bribe,

$$\mathbf{E}[(\theta_j - \theta_i)\mathbf{1}_{\{\theta_i \leq \theta_j\}}] - F(\theta^b + b)(\theta_j - b) - \mathbf{E}[(\theta_j - \theta_i)\mathbf{1}_{\{\theta^b + b \leq \theta_i \leq \theta_j\}}],$$

which has a derivative equal to

$$F(\theta_j) - F(\theta^b + b) - \max\{0, F(\theta_j) - F(\theta^b + b)\}.$$

This derivative is negative for $\theta_j < \theta^b + b$ and zero otherwise, so the payoff difference is minimized at $\theta_j = \theta^b + b$, where it equals

$$F(\theta^b + b)(\mathbf{E}[\theta_i \mid \theta_i \leq \theta^b + b] - b) > 0. \tag{15}$$

Hence for type $\theta_j > \theta^b$ not offering the bribe is strictly better than offering it. Therefore, we can delete no more briber strategies: For any $\theta_j > \theta^b$, either bribing or not bribing may be strictly better, depending on the set of accepting types (and we already established that any $\theta_j \leq \theta^b$ must not offer the bribe).

To support the no-bribing equilibrium in which j never offers the bribe we specify i's out-of-equilibrium beliefs as follows. If the bribe is offered then i believes that $\theta_j = \theta^b + \varepsilon$ with probability 1. His best response is to accept if and only if $\theta_i \leq \theta^b + \varepsilon + b$. For small $\varepsilon > 0$, no θ_j wants to offer the bribe. This can be seen by perturbing eqn. (15) with ε. □

Proof of Lemma 1: From the monotonicity established in the proof of Proposition 1 (eqn. (2)), $\mathcal{C}$ must be of the form $[c, 1]$; if type c gains from offering the bribe so does any type $\theta_j > c$. It should also be clear that regardless of the value of $\theta_j > b$, the "worst" of i's best responses (for j) occurs when i believes $\Pr(\theta_j = c) = 1$, in which case his best response is to accept when $\theta_i \in [0, c + b\rangle$ (in fact, the reader may check that $c + b = \min \mathbb{A}$). The usual continuity arguments imply that type c is indifferent between offering the bribe and not offering it, in this worst case (the reader may also check that $D(\mathcal{C}) = (c, 1])$. □

Proof of Lemma 2: Suppose two bribe amounts b and $b' < b$ are offered in equilibrium. From the arguments of Proposition 1, the set of types θ_i that accept b and b' are $[0, A\rangle$ and $[0, A'\rangle$ respectively. Clearly $A' < A$, otherwise no type would offer b.

Let the infimum type who offers b be denoted $\tilde{\theta}_j = \inf\{\theta_j : b(\theta_j) = b\}$. We show that $\theta_j > \tilde{\theta}_j$ implies that θ_j strictly prefers offering b to offering b', implying monotonicity.

If $\tilde{\theta}_j = 1$, we are done. If $\tilde{\theta}_j < 1$, denote the expected payoff to some type θ_j from offering b as

$$\pi(b, \theta_j) = F(A)(\theta_j - b) + \mathbf{1}_{\{\theta_j > A\}} \int_A^{\theta_j} (\theta_j - x)\, dF(x).$$

As with eqn. (2), we have $\partial \pi(b, \theta_j)/\partial \theta_j = \max\{F(A), F(\theta_j)\}$. Therefore,

$$\partial[\pi(b, \theta_j) - \pi(b', \theta_j)]/\partial \theta_j = \max\{F(A), F(\theta_j)\} - \max\{F(A'), F(\theta_j)\}.$$

Incentive compatibility (and continuity) imply $\pi(b, \tilde{\theta}_j) - \pi(b', \tilde{\theta}_j) \geq 0$. Since $\tilde{\theta}_j < 1$, we have $\tilde{\theta}_j < A$ (as in the second paragraph of the proof of Proposition 1). Therefore, $\partial[\pi(b, \tilde{\theta}_j) - \pi(b', \tilde{\theta}_j)]/\partial \theta_j > 0$, and so for all $\theta_j > \tilde{\theta}_j$, $\pi(b, \theta_j) - \pi(b', \theta_j) > 0$. □

Proof of Lemma 3: Observe that π^e is strictly increasing: For all $\theta'_j < \theta_j$,

$$\pi^e(\theta'_j) < \pi(b(\theta'_j), \theta_j) \leq \pi^e(\theta_j)$$

where the first inequality follows from the definition of π and the fact that $\theta_j > b(\theta_j)$. The second inequality follows from incentive compatibility.

We first show continuity approaching from the right. Suppose towards contradiction that for some $\theta_j \in [0, 1)$, there exists $\delta > 0$ such that for all

$\varepsilon > 0$, $\pi^e(\theta_j) + \delta \leq \pi^e(\theta_j + \varepsilon)$. Observe that

$$\begin{aligned}
& \pi^e(\theta_j + \varepsilon) - \pi(b(\theta_j + \varepsilon), \theta_j \mid A) \\
= \; & F(A(b(\theta_j + \varepsilon)))\,\varepsilon + \mathbf{1}_{\{\theta_j+\varepsilon > A(b(\theta_j+\varepsilon))\}} \int_{A(b(\theta_j+\varepsilon))}^{\theta_j+\varepsilon} (\theta_j + \varepsilon - x)\,dF(x) \\
& - \mathbf{1}_{\{\theta_j > A(b(\theta_j+\varepsilon))\}} \int_{A(b(\theta_j+\varepsilon))}^{\theta_j} (\theta_j - x)\,dF(x) \\
= \; & F(A(b(\theta_j + \varepsilon)))\,\varepsilon + \mathbf{1}_{\{\theta_j > A(b(\theta_j+\varepsilon))\}} \int_{A(b(\theta_j+\varepsilon))}^{\theta_j} \varepsilon\,dF(x) \\
& + \mathbf{1}_{\{\theta_j > A(b(\theta_j+\varepsilon))\}} \int_{\theta_j}^{\theta_j+\varepsilon} (\theta_j + \varepsilon - x)\,dF(x) \\
& + \mathbf{1}_{\{\theta_j+\varepsilon > A(b(\theta_j+\varepsilon)) > \theta_j\}} \int_{A(b(\theta_j+\varepsilon))}^{\theta_j+\varepsilon} (\theta_j + \varepsilon - x)\,dF(x) \\
< \; & 4\varepsilon
\end{aligned}$$

where the inequality follows because each of the four terms is no greater than ε and the last one is strictly less than ε. Therefore,

$$\lim_{\varepsilon \downarrow 0} [\pi^e(\theta_j + \varepsilon) - \pi(b(\theta_j + \varepsilon), \theta_j \mid A)] = 0.$$

But then for $\varepsilon > 0$ sufficiently small, $\pi(b(\theta_j + \varepsilon), \theta_j \mid A) > \pi^e(\theta_j + \varepsilon) - \delta \geq \pi^e(\theta_j)$, contradicting incentive compatibility. Therefore π^e is continuous from the right.

To see continuity from the left, suppose towards contradiction that we have $\lim_{\theta'_j \uparrow \theta_j} \pi^e(\theta'_j) < \pi^e(\theta_j)$. Since $\pi(b(\theta_j), \theta'_j)$ is continuous in θ'_j, we have $\lim_{\theta'_j \uparrow \theta_j} \pi(b(\theta_j), \theta'_j) = \pi(b(\theta_j), \theta_j) = \pi^e(\theta_j)$. But then for θ'_j sufficiently close to θ_j, $\pi(b(\theta'_j), \theta'_j) < \pi(b(\theta_j), \theta'_j)$, contradicting incentive compatibility. □

Proof of Proposition 5: Uniqueness. Consider a sequential equilibrium in which j's offer strategy, $b(\theta_j)$, is continuous and where a positive bribe is offered by some type θ_j.

Since any positive bribe would be accepted with positive probability (à la Proposition 1), it is clear that $b(0) = 0$. Let $\theta'_j = \max\{\theta_j : b(\theta_j) = 0\}$. For any $\delta > 0$ the set of acceptors of a bribe $b_\delta \equiv b(\theta'_j + \delta) > 0$ includes the interval $[0, \theta'_j + b_\delta]$. Hence the payoff for θ'_j from offering b_δ is at least $F(\theta'_j + b_\delta)(\theta'_j - b_\delta)$ while his payoff in equilibrium is $F(\theta'_j)(\theta'_j - \mathbf{E}[x \mid x \leq \theta'_j])$.

For δ sufficiently small, $b_\delta < \mathbf{E}[x \mid x \leq \theta'_j]$, therefore incentive compatibility requires $\theta'_j = 0$. Therefore $b(\theta_j)$ is strictly increasing at $\theta_j = 0$.

We extend this argument to prove that $b()$ can be constant only on some interval whose maximum is 1. For this, suppose that there exists a bribe b such that $\{\theta_j : b(\theta_j) = b\} = [\theta''_j, \theta'_j]$ where $\theta''_j < \theta'_j < 1$. This bribe is accepted by $\theta_i \in [0, A\rangle$ where $A < \theta'_j + b$ because $\theta''_j < \theta'_j$. Define $c = \theta'_j + b - A > 0$.[17] For any $\delta > 0$ (and $\theta'_j + \delta \leq 1$), the set of types that accept a bribe $b_\delta = b(\theta'_j + \delta)$ is an interval $[0, A_\delta\rangle$, where $A_\delta \geq \min\{\theta'_j + b, 1\}$. Hence $A_\delta \geq A + c$. If θ'_j offers b_δ then his payoff is $F(A_\delta)(\theta'_j - b_\delta)$. His equilibrium payoff is $F(A)(\theta'_j - b)$ when $A \geq \theta'_j$, and is at most $F(\theta'_j)(\theta'_j - b)$ when $A < \theta'_j$. As $\delta \to 0$, we have $b_\delta \to b$. However, $A_\delta - A \geq c > 0$ and $A_\delta - \theta'_j \geq \min\{b, 1 - \theta'_j\} > 0$, therefore type θ'_j has a strict incentive to offer b_δ instead of b for δ sufficiently close to 0. We conclude that if $b()$ is constant on a non-degenerate interval $[\theta''_j, \theta'_j]$ then it is constant on $[\theta''_j, 1]$ also.

We have established that the equilibrium bribe-function, $b()$, is strictly increasing on an interval $[0, \hat{\theta}]$ and constant on $[\hat{\theta}, 1]$. For any $\theta_j \in [0, \hat{\theta}]$, if j offers bribe $b(\theta_j)$, then it is accepted by types $\theta_i \in [0, \theta_j + b(\theta_j)]$. His equilibrium payoff is

$$\pi^e(\theta_j) = \pi(b(\theta_j), \theta_j) = F(\theta_j + b(\theta_j))(\theta_j - b(\theta_j))$$

Furthermore, the Envelope Theorem implies

$$\frac{d}{d\theta_j}\pi(b(\theta_j), \theta_j) = F(\theta_j + b(\theta_j))$$

Therefore,

$$F'(\theta_j + b(\theta_j))(1 + b'(\theta_j))(\theta_j - b(\theta_j)) = F(\theta_j + b(\theta_j))b'(\theta_j).$$

Therefore $b'()$ is defined by eqn. 6.

[17] Note that $A = 1$ would require $\theta'_j = 1$, otherwise any $\theta_j > \theta'_j$ should not offer any bribe greater than b. Therefore $A < 1$ also.

Finally, we show that $\hat{\theta} = \bar{\theta}$ where $\bar{\theta}$ is defined to be the lowest type such that $\beta(\theta_j) + \theta_j = 1$. Suppose that $\hat{\theta} < \bar{\theta}$. For $\delta \geq 0$ denote the set of types that accept $b(\hat{\theta} - \delta)$ by $[0, a_\delta]$. Notice that a_δ is discontinuous at $\delta = 0$ because $\hat{\theta} + b(\hat{\theta}) < 1$. Therefore, for sufficiently small $\delta > 0$, type $\hat{\theta} - \delta$ has a strict incentive to offer $b(\hat{\theta})$, which is a contradiction. On the other hand, if $\hat{\theta} > \bar{\theta}$ then type $\bar{\theta} + \varepsilon$ could strictly gain by offering $b(\bar{\theta})$, which is accepted by all types $\theta_i \in [0, 1]$, also a contradiction.

Existence. When F is log concave, eqn. 6 (and the requirement $b(0) = 0$) uniquely defines a continuous function. To see that, note that $b'(0) = 1/2$ by l'Hôpital's rule. Therefore, $0 < b'(\theta_j) < \infty$ on some interval $[0, \varepsilon)$.

Using local arguments, consider (locally) the inverse of $b(\theta_j)$, denoted $\Theta(b)$, defined by $\Theta(0) = 0$ and

$$\Theta'(b) = \frac{F(\Theta(b) + b)}{F'(\Theta(b) + b)(\theta_j(b) - b)} - 1$$

when $\Theta(b) + b \leq 1$.

We claim that $\Theta(b)$ is a well-defined, (weakly) increasing and continuous function, and $\Theta'(b) > 0$ almost everywhere. To see this, note that $\Theta'(0) = 2$. If for some $b > 0$, $\Theta'(b) = 0$, then $\Theta(b) > b$. Furthermore,

$$\Theta''(b) = \left(\frac{F(\Theta(b) + b)}{F'(\Theta(b) + b)}\right)' \frac{1}{\Theta(b) - b} + \frac{F(\Theta(b) + b)}{F'(\Theta(b) + b)} \frac{1}{(\Theta(b) - b)^2},$$

which is strictly positive because $(F/F')' \geq 0$ by log concavity, and $\Theta(b) > b$. Therefore, $\Theta(b)$ is strictly increasing in a right-hand side neighborhood of b. That is, whenever $b'(\theta_j)$ becomes infinite, θ_j is only an inflexion point of $b()$, and $b()$ continues with a positive and finite derivative in the right-hand side neighborhood of θ_j. This demonstrates the existence of a unique continuous $b()$ from eqn. 6.

To finish the proof, we construct a strategy (and beliefs) for i, and show that it and $b()$ form a sequential equilibrium.

We show that $\theta_j > 0$ implies $b(\theta_j) < \theta_j$ to establish that it is rational for θ_j to offer $b(\theta_j)$. Since $b'(0) = 1/2$, $b(\theta_j) < \theta_j$ holds sufficiently close to $\theta_j = 0$. Let $\theta_j' = \min\{\theta_j : b(\theta_j) \geq \theta_j\}$ (assuming by contradiction that the set is nonempty). By continuity, $b(\theta_j') = \theta_j' > 0$. This implies $b'(\theta_j') = 0$. Since b' is also continuous, this implies $b(\theta_j' - \epsilon) \geq \theta_j - \epsilon$, a contradiction.

This implies that $b'(\theta_j)$ is positive whenever $\theta_j + b(\theta_j) < 1$ (i.e. whenever b' is not explicitly defined to be zero). Therefore, $b()$ is invertible on $[0, b(\bar{\theta}_j))$.

To construct the equilibrium, the acceptor i believes that an offer $\hat{b} < b(\bar{\theta})$ comes from type $\theta_j = b^{-1}(\hat{b})$; an offer of $b(\bar{\theta})$ comes from some type in $[\bar{\theta}, 1]$, where i's beliefs are a Bayesian update of F over that interval. Let i's beliefs for any out-of-equilibrium offer $\hat{b} > b(\bar{\theta})$ be the same posterior over $[\bar{\theta}, 1]$. An (obvious) best response for i is to accept an offer $\hat{b}$ if and only if $\theta_i \leq A(\hat{b})$ where

$$A(\hat{b}) = \begin{cases} b^{-1}(\hat{b}) + \hat{b} & \text{if } \hat{b} < b(\bar{\theta}_j) \\ 1 & \text{if } \hat{b} \geq b(\bar{\theta}_j) \end{cases}$$

Note that $A()$ is continuous and differentiable everywhere except $\hat{b} = b(\bar{\theta})$.

For j, offering $\hat{b} > b(\bar{\theta})$ is strictly dominated by offering $b(\bar{\theta}_j)$. Therefore to check incentive compatibility, it suffices to check that no type θ_j prefers to offer $\hat{b} \leq b(\bar{\theta})$, i.e. where $\hat{b} = b(\hat{\theta}_j)$ for some $\hat{\theta}_j$.

To prove this, first consider the quantity

$$\pi(b(\hat{\theta}_j), \theta_j) - \pi(b(\hat{\theta}_j), \hat{\theta}_j) = F(\hat{\theta}_j + \hat{b})(\theta_j - \hat{\theta}_j) + \mathbf{1}_{\{\theta_j > \hat{\theta}_j + \hat{b}\}} \int_{\hat{\theta}_j + \hat{b}}^{\theta_j} (\theta_j - x) dF(x)$$

(see eqn. (7)) which can be written as

$$\begin{cases} \int_{\hat{\theta}_j}^{\hat{\theta}_j + b(\hat{\theta}_j)} F(\hat{\theta}_j + b(\hat{\theta}_j)) dx + \int_{\hat{\theta}_j + b(\hat{\theta}_j)}^{\theta_j} (\theta_j - x) F(x) dx & \text{if } \hat{\theta}_j + b(\hat{\theta}_j) \leq \theta_j \\ \int_{\hat{\theta}_j}^{\theta_j} F(\hat{\theta}_j + b(\hat{\theta}_j)) dx & \text{if } \hat{\theta}_j \leq \theta_j < \hat{\theta}_j + b(\hat{\theta}_j) \\ -\int_{\theta_j}^{\hat{\theta}_j} F(x + b(x)) dx & \text{if } \theta_j < \hat{\theta}_j. \end{cases}$$

Second, consider the quantity $\pi(b(\theta_j), \theta_j) - \pi(b(\hat{\theta}_j), \hat{\theta}_j)$. Since

$$\frac{d}{d\theta_j}\pi(b(\theta_j), \theta_j) = \begin{cases} F(\theta_j + b(\theta_j)) & \text{if } \theta_j < \bar{\theta} \\ 1 & \text{if } \theta_j \geq \bar{\theta} \end{cases}$$

(which can be verified either directly or via the Envelope Theorem), we have

$$\pi(b(\theta_j), \theta_j) - \pi(b(\hat{\theta}_j), \hat{\theta}_j) = \begin{cases} \int_{\hat{\theta}_j}^{\theta_j} F(\min\{x + b(x), 1\})dx & \text{if } \hat{\theta}_j \leq \theta_j \\ -\int_{\theta_j}^{\hat{\theta}_j} F(x + b(x) \wedge 1)dx & \text{if } \theta_j < \hat{\theta}_j. \end{cases}$$

Comparing these two quantities reveals

$$\pi(b(\theta_j), \theta_j) - \pi(b(\hat{\theta}_j), \hat{\theta}_j) \geq \pi(b(\hat{\theta}_j), \theta_j) - \pi(b(\hat{\theta}_j), \hat{\theta}_j)$$

implying incentive compatibility. □

Proof of Proposition 6: The proof involves demonstrating various properties of $b()$ under our refinement. Suppose that $b()$ is discontinuous at some point $\hat{\theta}_j$, so by monotonicity (Lemma 2),

$$b' = \lim_{\theta_j \uparrow \hat{\theta}_j} b(\theta_j) < \lim_{\theta_j \downarrow \hat{\theta}_j} b(\theta_j) = b''.$$

We show (Step 1) that there cannot be pooling to the left of $\hat{\theta}_j$, i.e. $b^{-1}(b')$ is either empty or a singleton. In Step 2 we show that there *has* to be pooling to the right of $\hat{\theta}_j$, i.e. $b^{-1}(b'')$ is a nondegenerate interval. In Step 3 we combine these arguments to demonstrate the first direction of the proof. We prove the converse statement Step 4.

Step 1 (no pooling on the left). With $b()$, θ_j, b, b' defined as above, suppose towards contradiction that $b^{-1}(b')$ is a non-degenerate interval $\langle \theta'_j, \hat{\theta}_j \rangle$. We first prove that there exists a briber type who could improve upon his equilibrium payoff by offering $b' + \varepsilon$ (for $\varepsilon > 0$ sufficiently small) if this offer revealed his type. For an appropriate ε, denote the set of such types by $\mathcal{B}_\varepsilon$.

Then, we prove that briber types near (above) $\inf \mathcal{B}_\varepsilon$ strictly prefer to offer $b' + \varepsilon$ if, as a result of this deviation, the acceptor believes that their type is $\inf \mathcal{B}_\varepsilon$. These types form the credible deviating set whose uniqueness is obvious.

Claim 1: For any $\varepsilon > 0$ sufficiently small, briber type $\hat{\theta}_j$ strictly prefers to offer $b' + \varepsilon$ (versus his equilibrium action) if this act reveals his type, i.e. if $A(b' + \varepsilon) = \hat{\theta}_j + b' + \varepsilon$.

Proof of Claim 1: The equilibrium payoff of a type $\theta_j \in b^{-1}(b')$ who offers b' can be written

$$\pi^e(\theta_j) = F(A(b'))(\theta_j - b') + \int_{A(b')}^{\max\{A(b'),\theta_j\}} (\theta_j - x)\, dF(x).$$

By Lemma 3, the equilibrium payoff of type $\theta_j = \hat{\theta}_j$ can also be expressed with this equation, regardless of whether he actually offers b' in equilibrium. Furthermore, since $b' < A(b')$, $\pi^e(\hat{\theta}_j) \leq \max\{F(A(b')), F(\hat{\theta}_j)\}(\hat{\theta}_j - b')$.

Suppose that type $\hat{\theta}_j$ offered $b' + \varepsilon$ and that this act perfectly revealed his type to bidder i. Then his payoff would be

$$\pi_\varepsilon(\theta_j) \equiv F(\theta_j + b' + \varepsilon)(\theta_j - b' - \varepsilon).$$

Since $b^{-1}(b')$ is a non-degenerate interval with $\hat{\theta}_j = \sup b^{-1}(b')$, we must have $A(b') < \hat{\theta}_j + b'$. Therefore for all sufficiently small $\varepsilon > 0$,

$$\begin{aligned}
\pi^e(\hat{\theta}_j) &\leq \max\{F(A(b')), F(\hat{\theta}_j)\}(\hat{\theta}_j - b') \\
&< F(\hat{\theta}_j + b')(\hat{\theta}_j - b') - \varepsilon \\
&< F(\hat{\theta}_j + b' + \varepsilon)(\hat{\theta}_j - b') - \varepsilon F(\hat{\theta}_j + b' + \varepsilon) \\
&= F(\hat{\theta}_j + b' + \varepsilon)(\hat{\theta}_j - b' - \varepsilon) = \pi_\varepsilon(\hat{\theta}_j)
\end{aligned}$$

proving Claim 1.

Since $b^{-1}(b')$ is a non-degenerate interval with $\theta'_j = \inf b^{-1}(b')$, we must have $\theta'_j < A(b') + b'$. Fix $\varepsilon > 0$ as in Claim 1 so that $\varepsilon < A(b') + b' - \theta'_j$.

Denote as follows the lowest type who would receive a (weakly) higher payoff if, by offering $b' + \varepsilon$, he would reveal his type to bidder i.

$$\theta_j^* = \inf\{\theta_j : \pi_\varepsilon(\theta_j) \geq \pi^e(\theta_j)\}.$$

Following Claim 1, this type is clearly well-defined. Furthermore, continuity implies $\theta_j^* < \hat{\theta}_j$, while the choice of small ε implies $\theta_j' < \theta_j^*$.

We establish that a credible deviating set (offering $b'+\varepsilon$) exists by showing the following claim—the set of types who are strictly better off (than in equilibrium) offering $b' + \varepsilon$ and being (mis-)identified as type θ_j^* has positive measure, and has an infimum of θ_j^*. The credible deviating set is then simply defined as the set of all such types, including θ_j^*.

Denote the deviation payoff to type θ_j from offering $b' + \varepsilon$ and being (mis-)identified as type θ_j^* as

$$\pi^d(\theta_j) = F(\theta_j^* + b' + \varepsilon)(\theta_j - b' - \varepsilon).$$

Claim 2: For sufficiently small $\delta > 0$, we have $\pi^d(\theta_j^* + \delta) > \pi^e(\theta_j^* + \delta)$, while $\theta_j < \theta_j^*$ implies $\pi^d(\theta_j) < \pi^e(\theta_j)$.

Proof of Claim 2: First, on the range $\theta_j \in (\theta_j', \hat{\theta}_j)$, we have

$$\frac{d}{d\theta_j}(\pi^d(\theta_j) - \pi^e(\theta_j)) = F(\theta_j^* + b' + \varepsilon) - \max\{F(A(b')), F(\theta_j)\}.$$

For $\theta_j \in (\theta_j', \theta_j^* + \delta)$ (δ sufficiently small), this difference is positive. That follows from the fact that $A(b') < \theta_j^* + b' + \varepsilon$ (otherwise θ_j^* would be strictly better off in equilibrium than by identifying himself with the more-expensive bribe $b' + \varepsilon$). Therefore, we have proven the first part of the claim, and that $\theta_j \in (\theta_j', \theta_j^*)$ implies $\pi^d(\theta_j) < \pi^e(\theta_j)$.

A similar argument applies to $\theta_j \leq \theta_j'$, because such types receive an equilibrium payoff that is at least as great as the payoff they could receive by behaving as a θ_j'-type, so $\pi^e(\theta_j) > \pi^d(\theta_j)$. Hence θ_j^* is the lowest type who can (weakly) gain by offering $b' + \varepsilon$ and having his type revealed, proving the claim and establishing the existence of a unique credible deviating set.

Step 2 (pooling on the right). Suppose towards contradiction that $b()$ is strictly increasing on a nondegenerate interval $\Theta = (\hat{\theta}_j, \theta'_j)$. Then for all $\theta_j \in \Theta$, $A(b(\theta_j)) = \theta_j + b(\theta_j)$, implying $\pi^e(\theta_j) = F(\theta_j + b(\theta_j))(\theta_j - b(\theta_j))$. Continuity implies $\pi^e(\hat{\theta}_j) = F(\hat{\theta}_j + b'')(\hat{\theta}_j - b'')$, regardless of whether $b(\hat{\theta}_j) = b''$.

We first claim that for $\varepsilon > 0$ sufficiently small, briber type $\hat{\theta}_j$ would strictly prefer to deviate to offering $b'' - \varepsilon$ if that act would reveal his type. As established in eqn. (17), $\pi^e()$ is differentiable from the right at $\hat{\theta}_j$, with $\frac{d\pi^e(\hat{\theta}_j)}{d\theta_j} = F(\hat{\theta}_j + b'')$. Also,

$$\lim_{\delta \downarrow 0} \frac{b(\hat{\theta}_j + \delta) - b''}{\delta} = \frac{F'(\hat{\theta}_j + b'')(\hat{\theta}_j - b'')}{F(\hat{\theta}_j + b'') - F'(\hat{\theta}_j + b'')(\hat{\theta}_j - b'')}.$$

That is, $b()$ must satisfy the usual differential equation from the right due to (local) incentive compatibility. Furthermore,

$$0 < F(\hat{\theta}_j + b'') - F'(\hat{\theta}_j + b'')(\hat{\theta}_j - b'') = -\left. \tfrac{d}{db}[F(\hat{\theta}_j + b)(\hat{\theta}_j - b)] \right|_{b=b''}.$$

where the inequality follows from the strict monotonicity of $b()$. Therefore, for $\varepsilon > 0$ sufficiently small, briber type $\hat{\theta}_j$ strictly prefers to deviate to $b'' - \varepsilon$ if this act reveals his type:

$$\pi^d(\hat{\theta}_j) \equiv F(\hat{\theta}_j + b'' - \varepsilon)(\hat{\theta}_j - b'' + \varepsilon) > F(\hat{\theta}_j + b'')(\hat{\theta}_j - b'') = \pi^e(\hat{\theta}_j).$$

By continuity, the same $\pi^d(\theta_j) > \pi^e(\theta_j)$ holds for some $\theta_j < \hat{\theta}_j$. Let the smallest such type be denoted $\theta_j^* = \inf\{\theta_j : F(\theta_j + b'' - \varepsilon)(\theta_j - b'' + \varepsilon) \geq \pi^e(\theta_j)\}$, and $b^* = \lim_{\theta_j \downarrow \theta_j^*} b(\theta_j)$. Consider the (deviation) payoff of any type θ_j from offering $b'' - \varepsilon$ if he were perceived to be type θ_j^*, denoted

$$\pi_{-\varepsilon}(\theta_j) = F(\theta_j^* + b'' - \varepsilon)(\theta_j - b'' + \varepsilon).$$

Using differentiability arguments similar to previous ones, one can show that (from the right) $\frac{d}{d\theta_j}\pi^e(\theta_j^*) < \frac{d}{d\theta_j}\pi_{-\varepsilon}(\theta_j^*)$. Therefore, there exists a nondegenerate set of types $\mathcal{C}$, with infimum θ_j^*, who strictly gain by offering $b''-\varepsilon$ when the offer is accepted by types $\theta_i \geq \theta_j^* + b''$. This establishes the existence of a credible deviating set, which is a contradiction. We conclude that if $b()$ is discontinuous at some $\hat{\theta}_j$, then it is constant on some nondegenerate interval $(\hat{\theta}_j, \theta_j')$.

Step 3. In the proof of Proposition 5, we established that in any sequential equilibrium, if $b()$ is constant on some nondegenerate interval (X, θ_j), then it is either constant or discontinuous at θ_j; Step 1 rules out a discontinuity in that situation, under our refinement. Therefore, if $b()$ is constant on (X, θ_j), then it is constant on $(X, 1]$. Step 2 implies that if $b()$ is discontinuous at some $\hat{\theta}_j$, then it is constant on some $(\hat{\theta}_j, X)$, hence it is constant on $(\hat{\theta}_j, 1]$.

We conclude that if an equilibrium satisfies our refinement, then there can be at most one discontinuity point in $b()$, hereafter denoted $\hat{\theta}_j$ (if it exists), so that $\theta_j > \hat{\theta}_j$ implies $b(\theta_j) = b''$, hereafter denoted $\hat{b} = b''$. Since $b()$ is continuous on $[0, \hat{\theta}_j)$, it follows from the proof of Proposition 5 (or local incentive compatibility) that $b() = b^*()$ on that range. Furthermore, it is clear that $\hat{\theta}_j \leq \bar{\theta}$.

It remains to be shown that the bribe offered by "high" types, $\hat{b}$, is always accepted: $A(\hat{b}) = 1$. Suppose towards contradiction that $A(\hat{b}) = \mathbf{E}[\theta_j \mid \theta_j \geq \hat{\theta}_j] + \hat{b} < 1$.

Since $\theta_j = 1$ is pooled with lower types when offering $\hat{b}$, he would strictly gain if he could offer $\hat{b} + \varepsilon$ and, in doing so, reveal his type to bidder i. Formally,

$$F(A(\hat{b}))(1 - \hat{b}) + \int_{A(\hat{b})}^{1} (1 - x)\, dF(x) = \pi^e(1) < 1 - \hat{b} - \varepsilon$$

for small ε because $A(\hat{b}) > \hat{b}$.

On the other hand, no type in $(\hat{\theta}_j, A(\hat{b}) - \hat{b} - \varepsilon]$ would want to offer $\hat{b}+\varepsilon$ and reveal his type, since that would involve paying a higher bribe to a smaller set of acceptors. For small ε, this interval is nonempty, so by continuity of π^e, the argument extends to type $\hat{\theta}_j$.

Similarly, letting $0 < \varepsilon < A(\hat{b}) - \hat{\theta}_j - \hat{b}$ (which is possible because $\inf b^{-1}(\hat{b}) = \hat{\theta}_j < \sup b^{-1}(\hat{b})$), type $\theta_j < \hat{\theta}_j$ would not gain by deviating to $\hat{b}+\varepsilon$ and reveal his type. To see this, observe that incentive compatibility and the choice of small ε imply the following.

$$\begin{aligned}\pi^e(\theta_j) &= F(\theta_j + \beta(\theta_j))(\theta_j - \beta(\theta_j)) \geq F(A(\hat{b}))(\theta_j - \hat{b}) \\ &> F(\theta_j + \hat{b} + \varepsilon)(\theta_j - \hat{b} - \varepsilon)\end{aligned}$$

Therefore, the lowest type, θ_j^*, who is willing to offer $\hat{b}+\varepsilon$ and reveal his type, exists, and $\theta_j^* > \hat{\theta}_j$. Using an argument identical to one in the previous step, a unique credible deviating set exists in which θ_j^* is the smallest type, contradicting the supposition that $A(\hat{b}) < 1$ and proving $\mathbf{E}[\theta_j \mid \theta_j \geq \hat{\theta}_j] + \hat{b} \geq 1$.

Finally, the continuity of π^e implies

$$\hat{b} = \hat{\theta}_j - F(\hat{\theta}_j + \beta(\hat{\theta}_j))(\hat{\theta}_j - \beta(\hat{\theta}_j))$$

which proves the first direction of the Proposition.

Step 4 (Converse).

We show that for any $b()$ satisfying the conditions of the Proposition, a sequential equilibrium exists in which j uses that strategy. Bidder i's strategy in response to equilibrium actions is obvious; it can be completely specified as follows. (Without loss of generality, assume $b(\hat{\theta}_j) = \beta$ at the discontinuity.)

$$A(b) = \begin{cases} b^{-1}(b) + b & \text{if } b \leq \beta\hat{b} \\ \hat{\theta}_j + \beta & \text{if } b \in (\beta, \hat{b}) \\ 1 & \text{if } b \geq \hat{b}) \end{cases}$$

Out-of-equilibrium beliefs which justify this strategy can be constructed so that upon receiving an offer $b \in (\beta, \hat{b})$, i believes $\theta_j = \hat{\theta}_j$, and upon receiving an offer $b > \hat{b}$, i believes $\theta_j = 1$.

The non-trivial question remains whether there exists a credible deviating set that breaks our mostly-separating equilibrium. A deviation to $b' > \hat{b}$ is clearly unprofitable for all types, no matter what the receiver's (best) response is, because $\hat{b}$ is already accepted with probability 1, and is cheaper than b'.

No type $\theta_j \leq \hat{\theta}_j$ would want to offer $b' \in (\beta(\hat{\theta}_j), \hat{b})$ and reveal his type. To see this, let θ'_j satisfy $b' = \beta(\theta'_j)$; the deviation payoff is strictly worse than the equilibrium payoff,

$$F(\theta_j + b')(\theta_j - b') < F(\theta'_j + b')(\theta_j - b') \leq F(\theta_j + \beta(\theta_j))(\theta_j - \beta(\theta_j))$$

where the latter follows from incentive compatibility. Therefore, a credible deviating set offering b', denoted $\mathcal{B}$, must satisfy $\theta_j^C \equiv \inf \mathcal{B} > \hat{\theta}_j$ by continuity. Furthermore, continuity implies that θ_j^C is indifferent between deviating and not: $F(\theta_j^C + b')(\theta_j^C - b') = \theta_j^C - \hat{b}$. But then for $\theta_j \in [\hat{\theta}_j, \theta_j^C)$, we have $F(\theta_j^C + b')(\theta_j - b') > \theta_j - \hat{b}$, implying $\theta_j \in \mathcal{B}$, which is a contradiction. Hence the equilibrium satisfies our refinement. $\square$

Appendix II: Supplemental Material on Refinements

Perfect Sequential Equilibrium

At the end of Section 3.1 we claim that the bribing equilibrium is the unique PSE of the extended game under fixed bribes. Here we put forward a formal argument.

In a sender-receiver game, for a given equilibrium and out-of-equilibrium message m, Grossman and Perry (1986) call the beliefs of the receiver (regarding the sender's type, upon seeing m) *consistent* with the equilibrium and the prior distribution of the sender's type, if there exists a mixed strategy α of the receiver that is a best response given these beliefs, and the beliefs are generated from the prior applying Bayes' rule conditional on the sender's type being in the set of types that benefit from sending m when the receiver's response to m is α. A Perfect Sequential Equilibrium is a sequential equilibrium that satisfies this additional consistency requirement.

Clearly, the bribing equilibrium of the extended game (under fixed bribes) satisfies this requirement, as there are no out-of-equilibrium moves for the briber. In light of Propositions 1–3 (in the text), the only other candidate for PSE is the no-bribing equilibrium. We now show that any beliefs supporting the no-bribing equilibrium have to fail the consistency requirement.

In the no-bribing equilibrium of our extended game, Grossman and Perry's consistency implies the following regarding the acceptor's beliefs when he is unexpectedly offered b. His beliefs must come from the prior distribution applying Bayes rule conditional on $\theta_j \in \mathbb{B}$ for some $\mathbb{B} \subseteq [0,1]$. By Proposition 1, his best response is to accept if and only if $\theta_i \in [0, A\rangle$, where $A \leq 1$. Given this response, player j will be better off deviating from the no-bribing equilibrium with type θ_j if and only if

$$F(A)(\theta_j - b) + \mathbf{E}_{\theta_i}[(\theta_j - \theta_i)\mathbf{1}_{\{A<\theta_i\leq\theta_j\}}] \geq \mathbf{E}_{\theta_i}[(\theta_j - \theta_i)\mathbf{1}_{\{\theta_i\leq\theta_j\}}].$$

Again, from Proposition 1, we know that this inequality will be satisfied by types θ_j belonging to an interval $\langle B, 1]$, and therefore by consistency, $\mathbb{B} = \langle B, 1]$. From Proposition 2, we know that for a given $b > 0$ there exist unique A and B satisfying the consistency requirement. In fact, we can conclude that in the no-bribing equilibrium of our game, but off the equilibrium path (i.e., if b is offered), Grossman and Perry's consistency requires the acceptor to behave as in the (unique) bribing equilibrium. Since briber types $\theta_j > B$ are strictly better off in the bribing equilibrium than in the no-bribing equilibrium, they would deviate and bribe if they had consistent beliefs in the no-bribing equilibrium. Therefore the no-bribing equilibrium is not a PSE.

Cho and Sobel's D1

Here we explain in more detail the consequences of D1 on no-bribing equilibria. This refinement restricts the acceptor's beliefs regarding the briber's type when a bribe is offered out of equilibrium. In the discussion, we focus attention on the case of a fixed bribe, since the results are even easier to show in the case of variable bribes. We show that a no-bribing equilibrium cannot satisfy D1 regardless of the value of $b < E(\theta)$.

Recall from eqn. (3) that θ^b is the type of bidder j who would be indifferent between offering the bribe and not offering it, given that it is accepted by every type of bidder i. Whether or not i will accept the bribe depends on his beliefs regarding j's type. For example, if i believes that θ_j is sufficiently high (given the bribe is offered), then i's *best response* is to accept the bribe regardless of his type.

Clearly, if j's type is $\theta_j \geq \theta^b$, then for some best responses of the bribee (i.e., for i's beliefs about the briber's θ_j that are distributed sufficiently close to 1), j would be better off offering the bribe and so deviating from the no-bribing equilibrium. Cho and Sobel's D1 is based on the idea that a sender type that is "more likely" to benefit from a deviation than another type—that is, a type that benefits from a deviation for "more best responses" than another one does—should get "more weight" in the receiver's updating, conditional on observing that deviation.

To make this concept precise in our setting, we will say that θ'_j is *more likely to benefit from bribing* than θ_j is when the following is true:

> For all $A \in [0, 1]$, if θ_j is better off bribing when exactly $\theta_i \in [0, A]$ accept, then θ'_j is also better off bribing when exactly $\theta_i \in [0, A]$ accept.

In other words, the set of i's best responses that would induce θ_j to offer a bribe is a subset of the best responses that would induce θ'_j to do so. We will say that θ'_j is *strictly more likely to benefit from bribing* than θ_j if θ'_j is more likely to benefit from bribing than θ_j but the converse is not true.

Let $D(\theta_j)$ be the acceptor type such that θ_j is indifferent between offering the bribe and not offering it when the bribe is accepted by exactly the types $\theta_i \in [0, D(\theta_j)]$, that is,

$$F(D(\theta_j))(\theta_j - b) + \mathbf{E}_{\theta_i}[(\theta_j - \theta_i)\mathbf{1}_{\{D(\theta_i)<\theta_i\leq\theta_j\}}] = \mathbf{E}_{\theta_i}[(\theta_j - \theta_i)\mathbf{1}_{\{\theta_i\leq\theta_j\}}].$$

Note that $D(\theta_j)$ is undefined on $[0, \theta^b)$, but is unique on $[\theta^b, 1]$, e.g., $D(\theta^b) = 1$.

It is easy to see that θ'_j is strictly more likely to benefit from bribing than θ_j if and only if $D(\theta'_j) < D(\theta_j)$. Furthermore, one may check that the following are true: D has a unique fixed point, which we label θ^{Div}; $\theta^b < \theta^{Div}$; D is strictly decreasing on $[\theta^b, \theta^{Div}]$ and constant on $[\theta^{Div}, 1]$; finally, $b = \mathbf{E}[\theta_i \mid \theta_i \leq \theta^{Div}]$. Note that there is no type strictly more likely to benefit from deviation in the no-bribing equilibrium than any $\theta_j \in [\theta^{Div}, 1]$.

After these preliminaries we can formally describe the concepts D1 in our setting. Here, the acceptor's updating after a deviation must put "infinitely more weight" on sender types that are strictly more likely to benefit from the deviation. Therefore, under D1, i's beliefs regarding θ_j after a deviation must have support on briber types that are the most likely to benefit from deviation, i.e., $\theta_j \in [\theta^{Div}, 1]$. However, any distribution on this support (or a subset of it) is permissible.

Proposition 7 *For any $b < \mathbf{E}(\theta_i)$, there does not exist a no-bribing equilibrium that satisfies D1.*

Proof: Recall that by the definition of $\theta^{Div} = D(\theta^{Div})$, $b = \mathbf{E}_{\theta_i}[\theta_i \mid \theta_i \leq \theta^{Div}]$. Note also that for all $\theta_j > \theta^{Div}$, $b < \mathbf{E}_{\theta_i}[\theta_i \mid \theta_i \leq \theta_j]$.

Suppose that in a no-bribing equilibrium the acceptor's strategy is to accept the bribe if and only if $\theta_i \in [0, a\rangle$. If the equilibrium satisfies Cho and Sobel's D1 criterion, then $a \geq \min\{1, \theta^{Div} + b\}$. This is so because the acceptor believes $\theta_j \geq \theta^{Div}$ conditional on being offered the bribe, hence he must accept when $\theta_i < \theta^{Div} + b$. The payoff to briber type $\theta_j = a$ from offering the bribe is $F(a)(a - b)$, while his payoff from going directly to the second price auction is $F(a)(a - \mathbf{E}_{\theta_i}[\theta_i \mid \theta_i \leq a])$. Since $b < \mathbf{E}(\theta_i)$, we have

$\theta^{Div} < 1$. Therefore $a > \theta^{Div}$, implying $b < \mathbf{E}[\theta_i \mid \theta_i \leq a]$. Hence type $\theta_j = a$ strictly prefers offering the bribe to not offering it. Therefore condition D1 rules out the no-bribing equilibrium. □

References

Bagnoli, M. and T. Bergstrom (1989): "Log-Concave Probability and Its Applications," mimeo, University of Michigan.

Banks, J. and J. Sobel (1987): "Equilibrium Selection in Signalling Games," *Econometrica* **55**, 647–62.

Cho, I. and D. Kreps (1987): "Signalling Games and Stable Equilibria," *Quarterly Journal of Economics* **102**, 179–221.

Cho, I. and J. Sobel (1990): "Strategic Stability and Uniqueness in Signalling Games," *Journal of Economic Theory* **50**, 381–413.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT Press, Cambridge MA.

Graham, D. and R. Marshall (1987): "Collusive Bidder Behavior at Single-Object Second Price and English Auctions," *Journal of Political Economy* **95**, 1217–39.

Grossman, S. J. and M. Perry (1986): "Perfect Sequential Equilibrium," *Journal of Economic Theory* **39**, 97–119.

Jackson, M. and S. Wilkie (2001): "Endogenous Games and Mechanisms: Side Payments among Players," mimeo, Cal Tech.

Laffont, J.-J. and D. Martimort (2000): "Mechanism Design with Collusion and Correlation," *Econometrica* **68**, 309–42.

Mailath, G. and P. Zemsky (1991): "Collusion in Second Price Auctions with Heterogeneous Bidders," *Games and Economic Behaivor* **3**, 467–86.

Marshall, R. and L. Marx (2002): "Bidder Collusion," mimeo, The Pennsylvania State University.

McAfee, P. and McMillan (1992): "Bidding Rings," *American Economic Review* **82**, 579–99.

Schummer, J. (2000): "Manipulation through Bribes," *Journal of Economic Theory* **91**, 180-98.

Van Damme, E. (1991): *Stability and Perfection of Nash Equilibria*, 2nd edition, Springer-Verlag, Berlin—Heidelberg—New York.