Collusion-Robust Auction Design: Posted Price and a Knockout Auction

Karen Wu*

October 17, 2026

Abstract

I study how a revenue-maximizing principal allocating a single good should optimally design her auction in the presence of collusion. The principal evaluates mechanisms according to the worst-case revenue that could arise from collusive or non-collusive play. The principal's optimal mechanism is to post a price and run an efficient knockout auction in-house. Posting a price and running the efficient knockout in-house remains the optimal mechanism when the principal additionally hypothesizes that colluders are maximizing their joint surplus or are constrained by interim truth-telling constraints. With surplus-maximizing colluders, posting a price without running the knockout in-house is also an optimal mechanism.

1 Introduction

Collusion threatens revenue in auctions. While economic theory often imagines that the mechanism designer is able to coordinate bidders to play the non-cooperative equilibrium of her choice, in reality, bidders may collude. In general, auctions raise revenue relative to a posted price by stoking competition between the bidders. By coordinating play between the bidders, collusion can neutralize this competition. There

*University of Chicago. I am very grateful to Ben Brooks, Doron Ravid, and Phil Reny for their guidance and support. I also thank Alex Frankel, Agathe Pernoud, Joe Root, Lars Stole, and the workshop audience at the University of Chicago for useful comments and discussions. This paper was previously circulated under the title "Posted Price and a Knockout Auction: Revenue Maximization Under the Shadow of Collusion."

is clear evidence that collusion is relevant for the mechanism designer in practice. Some large cartels have been revealed and prosecuted, like the Great Electrical Conspiracy, which involved G.E., Westinghouse, and many other firms in the 1950s. Also, a recent empirical literature suggests many more undiscovered cartels exist. Chassang and Ortner (2019), Kawai and Nakabayashi (2022), and Chassang et al. (2022) document the prevalence of patterns in auction bids that are inconsistent with non-cooperative play. They use this to infer the existence of cartels. In particular, Kawai and Nakabayashi (2022) estimate that approximately 37% of auctions in their dataset of Japanese government construction procurement auctions are impacted by collusion.

However, a mechanism designer may find it difficult to anticipate how exactly bidders collude. Many aspects of the real world complicate or restrict the collusive arrangements that bidders can form. For example, in the Great Electrical Conspiracy, bidders deliberately communicated with each other in noisy and constrained ways. Because they feared detection by a regulator, colluders avoided associating with each other in public, referred to the colluding firms only by number codenames on the telephone, and most famously, rotated which firm was given priority in auctions according to a pre-set two-week schedule rather coordinating for each auction Smith (1961a), Smith (1961b). In Smith (1961b), this was referred to as "bidding by the phases of the moon."

In this paper, I study how a revenue-maximizing principal should design an auction given that she does not know exactly how bidders are colluding. Being uncertain about what collusive arrangements might arise, the principal takes the worst case over a large set of possible behaviors. I solve for the optimal mechanism and interpret it as posting the optimal price and then implementing the efficient collusive arrangement on behalf of the bidders. I then conduct a sensitivity analysis of the assumptions in the baseline model. I first add assumptions about the possible collusive arrangements and again solve for the optimal mechanism. The optimal mechanisms identified under the baseline model remain optimal. I then restrict the principal's ability to design the game by requiring that the auction not make payments to the bidders. Under this assumption, I identify the format of the optimal mechanism under additional assumptions. The optimal mechanisms identified under the baseline model are no longer optimal.

The most closely related previous literature models cartels as achieving a Pareto efficient outcome subject to the constraints known to the mechanism designer. One

strand of the literature studies how cartels in auctions would operate given a fixed auction format (Graham and Marshall (1987), Mailath and Zemsky (1991), McAfee and McMillan (1992)). Another strand studies how a revenue-maximizing principal should design in response to collusion (Che and Kim (2006), Che and Kim (2009), Pavlov (2008)). In both of these areas, often, the principal articulates a complete model of cartel formation and functioning. The cartel is then assumed to achieve a Pareto efficient outcome subject to the constraints implied by the environment. In this way, cartels are framed as akin to the theoretical mechanism designer: they are able to implement any institution and equilibrium possible among the ones that satisfy the principal's assumptions. I discuss these papers in greater detail in the related literature section, Section 6.

Unlike the previous literature, I consider a principal from whose perspective bidders implement an inefficient outcome. This principal knows only a subset of the constraints colluding bidders face. Other constraints are unknown because collusion depends on the details of the environment. For example, colluding bidders often cope with a regulator who might detect and punish them. Successful collusion must be profitable enough to outweigh the expected cost of being caught. However, detection, successful conviction, and the extent of fines ultimately depend on the details of how colluders interact and how much evidence they generate of their coordination. As a result, the principal may not know exactly what collusive arrangements are profitable enough to carry out, and the collusive outcomes that arise may appear inefficient to the principal.

Assuming that colluders achieve the Pareto efficient outcome subject to the principal's incomplete list of constraints may lead the principal to incorrectly assess mechanisms as performing well against collusion. Inefficient collusive arrangements may threaten revenue more than efficient collusion does. At a high level, collusion transfers surplus from the principal to the bidders. Efficient collusion ensures that there is no or little loss of total surplus in this transfer. As a result, when bidders collude inefficiently, the principal's revenue can be even lower than when they collude efficiently because total surplus overall may be lower. I explore this possibility in a detailed example in Section 3.

In my baseline model, the principal designs a game that has a non-cooperative equilibrium which corresponds to non-collusive play, but bidders may play collusively instead. The principal is uncertain of the particular arrangement that might arise.

She entertains a large set of possible arrangements. These arrangements satisfy some minimal constraints, which reflect the qualitative knowledge that the principal holds about the process through which collusion forms. In particular, the principal assumes that whatever collusive arrangement forms, it must produce at least as much ex ante surplus for each bidder as playing non-collusively would. She then evaluates her mechanism according to the worst-case revenue that could arise under collusive or non-collusive play.

The principal's optimal mechanism is to post a price and then absorb the efficient collusive arrangement into her mechanism (Theorem 2). The principal's posted price is that which would be optimal if she were facing a single agent whose value is the maximum of the bidders' realized values. From there, the principal can choose from a multiplicity of efficient collusive arrangements to bring in-house. She could absorb the collusive arrangement identified by Mailath and Zemsky (1991) (Corollary 1), or, if bidders are symmetric, she could run the surplus-maximizing pre-auction knockout identified in McAfee and McMillan (1992) (1992) (Corollary 2). While Mailath and Zemsky (1991) and McAfee and McMillan (1992) study efficient collusion against a fixed auction formats—the second-price auction with a reserve price and the first-price auction with reserve, respectively—the compositions of the collusive arrangements they identify and a posted price are optimal mechanisms when the principal can fully design her mechanism in my setting.

The optimal mechanisms have three key qualitative features (Theorem 1). First, the good is allocated to a bidder if and only if the maximum realized value exceeds the posted price. Second, the good is allocated to the high value bidder whenever it is allocated to any bidder. Third, while bidders might make payments to the principal and the principal might make payments to the bidders, the principal collects exactly p^* in total revenue whenever she allocates the good to a bidder. Otherwise, she collects exactly 0 in revenue. The exact details of the transfers in one optimal mechanism are given in Section 4.3.

The proof of the main result reveals the important structure of the problem. The principal's mechanism affects the behavior of colluding bidders through two channels. First, the principal's mechanism determines total bidder surplus from playing the non-cooperative equilibrium. This determines how attractive colluding is to the bidders, since playing non-collusively is the outside option. Second, the principal's mechanism defines the minimum revenue she must collect for selling the good at dif-

ferent probabilities. These minimum revenues are captured by an object I call the *cost function*.

The proof proceeds by doing a nested optimization over the two channels. When two mechanisms induce the same cost function but result in different levels of surplus from non-collusive play, it is possible to show that the principal weakly prefers the mechanism that induces the higher surplus. In the inner problem, the principal holds fixed a cost function and chooses a mechanism that maximizes non-collusive bidder surplus among those within the class defined by the cost function. In the outer problem, the principal maximizes over the cost functions. Given the optimal choice of cost function, the principal's optimal mechanism solves the inner problem induced by that cost function.

The optimal mechanisms under the baseline model remain optimal under additional assumptions on collusive outcomes (Theorems 3, 4). I consider variations of the model where the principal knows that colluding bidders are constrained by interim incentive constraints, maximize their surplus, or cannot dispose of the good. In particular, when bidders are additionally assumed to collude efficiently, i.e., to maximize their surplus subject to the other constraints, simply posting a price is also an optimal mechanism (Corollary 3). Intuitively, when bidders collude efficiently, the principal no longer has to bring the efficient collusive arrangement in-house to ensure that it is done. The value of the principal's problem remains the same as in the baseline model. Corollary 3, in particular, considers the model of collusion in McAfee and McMillan (1992) and demonstrates that posting a price is the optimal mechanism among all possible mechanisms.

When the principal cannot make payments to the bidders, the optimal mechanism identified in the baseline model is no longer optimal. Because the implementations of the baseline model's optimal mechanism identified in Section 4 involve payments to the bidders and this feature is rarely observed in practice, it is important to understand whether it is possible to implement the optimal mechanism without making payments to the bidders. Proposition 1 demonstrates that the qualitative features of the optimal mechanism when the principal cannot pay the bidders differ from the optimal mechanism under the baseline model. While the good is still allocated whenever the maximum realized value exceeds some threshold, the optimal mechanism no longer allocates the good efficiently among the bidders. Instead, the key qualitative feature of the mechanism is that while he does not necessarily get the good, the

highest value bidder pays the principal.

2 Model

In this section, I present the main model and define the principal's problem.

N bidders, indexed $1, ..., n$, participate in a single good auction designed by the principal. Each bidder i has a value for the good v_i, which is drawn from an atomless distribution $F_i[0,1]$ with full support and density f_i. values are drawn independently and are each bidder's private information. I use V to refer to the space of all profiles of values, $[0,1]^N$.

The principal designs a *mechanism*. A mechanism consists of a game form and an associated Bayes-Nash equilibrium. The principal specifies a space of actions for each bidder A_i, a mapping from action profiles to probabilities of allocating the good to each bidder $q : \prod_{i=1}^n A_i \to [0,1]^N$ where $\sum_i q_i(a) \leq 1$ for each a, and mappings from action profiles to transfers paid by each bidder $t_i : \prod_{i=1}^n A_i \to \mathbb{R}$. I use A to refer to the space of profiles of actions, $\prod_{i=1}^n A_i$. The principal also specifies a Bayes-Nash equilibrium of the game (A, q, t), σ, which is how bidders play if they do not collude.

Participation in the principal's auction is voluntary. The principal must set an auction that contains an *opt-out action* for each bidder. When a bidder i plays his opt-out action a_i^0, he guarantees that he gets the good with 0 probability and pays 0 in transfers. Playing action profile a results in payoff $q_i(a)v_i - t_i(a)$ for bidder i with value v_i.

The principal entertains two possible models for how bidders behave. On one hand, bidders may be playing *non-collusively* and follow the strategies given by the BNE σ. On the other hand, bidders may be playing *collusively*, which encompasses a rich set of behaviors.

The key objects in the model that describe collusive behavior capture the outcome of collusion rather than specify a particular timing or structure for the collusion formation game. A *collusive outcome* is given by a tuple (κ, q^c, t^c) where

- $\kappa : V \to \Delta A$ specifies how the collusive arrangement plays the principal's original mechanism given a profile of bidders' values,
- $q^c : V \to [0,1]^N$ with $\sum_i q_i^c(v) \leq 1$ associates each profile of values with a probability bidder i ultimately gets the good,

- $t^c : V \to \mathbb{R}^N$ maps a profile of bidders' values to the transfer a particular bidder i must make to the collusive arrangement in excess of his payments to the principal.

Collusive outcomes are constrained by the games that induce them. For example, if bidders can quit the collusive arrangement after learning their values and default to playing the non-collusive equilibrium, then the interim payoff for each bidder from colluding must exceed the interim payoff for each bidder from playing the Bayes-Nash equilibrium. In contrast, if collusion requires significant planning and the details of the good are unclear ahead of time, then bidders may commit to colluding before they learn their values. In this case, colluding bidders should compare their ex ante payoffs from colluding to their ex ante payoffs from playing non-collusively, rather than comparing their interim payoffs.

The principal considers any collusive outcome that satisfies certain minimal constraints to be possible. Because the principal does not know the exact game through which bidders form their collusive arrangements, the principal does not know all the constraints that outcomes must satisfy. The minimal constraints are necessary conditions on outcomes that are common to all the outcomes of the games the principal considers possible. The constraints are mild enough that it seems reasonable to assume all possible collusive outcomes must satisfy them. Furthermore, while the principal considers any collusive outcome that satisfies the minimal constraints to be possible in the baseline model, the main results are robust to adding more constraints on collusive outcomes, i.e., shrinking the set of possible collusive outcomes. I explore variants of the model with additional assumptions on possible collusive outcomes in the extensions section.

At a minimum, collusive outcomes are constrained by the following frictions:[1]

1. ***Cartel-EABB.*** Collusion is not subsidized by outside sources, so expected

[1]The April 18th, 2025 draft of this paper included an incentive compatibility constraint on collusive outcomes. The May draft of Jagadeesan et al. (2025) did not impose incentive compatibility on collusive arrangements. The April version of the model is now studied in Section 5.1. The relevant constraint is C-IC. This draft relaxes C-IC without changing the main argument from the April draft. Specifically, the previous draft of the paper used it in the proof of Lemma 3, by verifying that when two mechanisms have ranked cost functions (defined in Section 4.2), a collusive outcome against the mechanism with the higher ranked cost function can be replicated in the mechanism with the lower ranked cost function. The author showed that this replication argument can be made while preserving C-IC, but the same argument goes through when the initial collusive outcome is not C-IC. Indeed, proof of Theorem 1 does not explicitly calculate a collusive outcome at any point, so C-IC was easily relaxed.

payments between bidders must be equal to 0.

$$\mathbb{E}_v[\sum_i t_i^c(v)] = 0 \tag{C-EABB}$$

This condition is also necessary if colluding bidders cannot exchange payments or payments between colluders are possible but must be ex post budget balanced.

2. ***Cartel-Feasibility.*** The cartel can only allocate the good to a bidder with at most the probability that bidders acquire the good:

$$\sum_i q_i^c(v) \leq \sum_{a \in \text{supp}(\kappa(\cdot|v))} \sum_i \kappa(a|v) q_i(a) \quad \forall v \tag{C-F}$$

The weak inequality allows the cartel to throw the good away rather than allocate it to a bidder. When the principal is procuring a good from the bidders, throwing away the good could be interpreted as producing the good inefficiently. In this setting, bidders' values are their private costs of producing the good the principal is procuring. If the cartel can mandate that a member produce the good in an inefficient manner, the cartel can essentially make the bidder throw away the profit he would make from producing the good at his true cost.

Additionally, Theorem 1 also holds with the alternative assumption that bidders may be able to reallocate the good but cannot throw it away:

$$\sum_i q_i^c(v) = \sum_{a \in \text{supp}(\kappa(\cdot|v))} \sum_i \kappa(a|v) q_i(a) \quad \forall v \tag{C-F'}$$

See Section 5.1 for more details.

3. ***Cartel-EAIR.*** Bidders only collude if they benefit from colluding over playing the mechanism non-collusively:

$$\mathbb{E}_v[q_i^c(v)v_i - t_i^c(v) - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) t_i(a)] \geq U_i^n \tag{C-EAIR}$$

where U_i^n is the expected payoff of the bidder i under the non-collusive BNE σ chosen by the principal.

This constraint captures that bidders can defect from collusion to return to playing the non-collusive equilibrium. While the constraint could be interpreted literally as capturing a participation decision that happens before bidders' values are realized, the constraint is also a necessary condition if participation decisions are made at the interim stage.

While the model is static, it can also be interpreted as capturing multiple plays or dynamic repeated play of the principal's auction. In this interpretation, the principal fixes an auction format. The cartel plays it over and over, either simultaneously or sequentially. The ex ante participation constraint allows for the possibility that the cartel pools surplus across instances of playing the principal's auction. If the auction is played sequentially, the ex ante constraint can be interpreted as the cartel having a discount factor close to 1.

The principal's model of collusion does not require the cartel to provide its members with incentives to reveal private information or obey the cartel's recommendations. This is consistent with the repeated play or dynamic interpretation. In a dynamic setting where the cartel plays the principal's auction repeatedly, the cartel may be able to set up a complicated mechanism that pools incentives across instances of play. This could take the form of linking truth-telling constraints across periods as in a Jackson and Sonnenschein (2007) (2007) mechanism where participants have quotas on how often they can report particular values over all periods. This could also take the form of grim trigger punishments where the promise of retributive low payoffs in the future incentivizes actions today. The minimal constraints the principal imposes on collusion allow for all of these possible arrangements.

Given a mechanism (A, q, t, σ) and a collusive outcome (κ, q^c, t^c), the following is the revenue that results:

$$Rev(\kappa, q^c, t^c) := \mathbb{E}_v[\sum_{a \in \text{supp}(\kappa(\cdot|v))\kappa(a|v)} \sum_i t_i(a)]$$

The principal evaluates the revenue of a mechanism according to the worst-case revenue that could arise from all possible collusive outcomes and the non-collusive BNE.

Formally, the principal calculates the following for each (A, q, t, σ):

$$\begin{aligned} \text{minRev}(A, q, t, \sigma) := \inf_{\kappa, q^c, t^c} \quad & Rev(\kappa, q^c, t^c) \\ \text{s.t.} \quad & \text{C-EABB}, \text{C-F}, \ \text{C-EAIR} \end{aligned}$$

Note that bidders playing σ is always possible as a "collusive" outcome,[2] so minRev takes the worst case over collusive and non-collusive outcomes.

A mechanism (A, q, t, σ) is *feasible* if it has opt-out actions for each bidder. The principal's problem is

$$\begin{aligned} \max_{(A,q,t,\sigma)} \quad & \text{minRev}(A, q, t, \sigma) \\ \text{s.t.} \quad & (A, q, t, \sigma) \text{ is feasible} \end{aligned} \tag{P}$$

3 Motivating Example

In this section, I present a simple example that previews the main results and demonstrates the key ideas in the proof. I demonstrate how inefficient collusion can undermine the principal's revenue more than efficient collusion.

First, consider a setting with two symmetric bidders 1 and 2 participating in an auction. Each bidder i has a value for the good $v_i \sim U[0, 1]$. Values are independently distributed. Bidders have standard quasilinear utility, so bidder i's payoff is $\mathbb{I}\{\text{get the good}\}v_i - t_i$ when he pays t_i.

Given that bidders play a Bayes Nash equilibrium of the principal's choice, one optimal auction is a first-price auction with a reserve price of $r = \frac{1}{2}$. The bidder with the highest bid above the reserve price wins the good. The winner pays his bid. Ties are broken uniformly at random. This auction induces an equilibrium, which is also unique and symmetric, where bidders bid according to the following strictly

[2] Specifically, the following collusive outcome replicates the BNE:

$$\begin{aligned} q^c(v) &= \mathbb{E}_{a\sim\sigma(v)}[q(a)] && \forall v \\ t^c(v) &= 0 && \forall v \\ \kappa(a|v) &= \prod_i \sigma_i(a_i|v_i) && \forall v \end{aligned}$$

and clearly satisfies C-F, C-EABB, and C-EAIR.

increasing function:

$$\beta_i(v_i) = \frac{v_i}{2} + \frac{1}{v_i} \cdot \frac{1}{8}$$

This equilibrium results in a revenue of $\frac{5}{12}$.

As explored in Graham and Marshall (1987) (1987), McAfee and McMillan (1992) (1991), and Mailath and Zemsky (1991) (1991), bidders could collude by implementing a *pre-auction knockout*. A pre-auction knockout is run before bidders participate in the principal's auction and awards one bidder the right to enter the principal's auction alone. In this sense, it "knocks out" the other bidder from the principal's mechanism. The winner of the knockout can then bid exactly the reserve price to get the good. The loser is compensated for abstaining from the principal's mechanism by a payment from the winner.

Given that they face a first-price auction with reserve price r, a particular form of pre-auction knockout maximizes the joint surplus of the bidders. Bidders could run a sealed-bid first-price auction with a reserve price r among themselves before the principal's auction. If the highest bid in the knockout exceeds r, the highest bidder enters the principal's auction alone and buys the good at the reserve price r. The winner then pays the loser the difference between his bid in the knockout and r. If the highest bid in the knockout is below r, no one enters the principal's auction, and the game ends. McAfee and McMillan (1992) show that there is a Bayes-Nash equilibrium of this game where the bidder with the higher value wins the knockout. In this BNE, the good is allocated efficiently whenever $\max\{v_1, v_2\}$ exceeds r. This collusive arrangement maximizes bidder surplus among all outcomes possible in the model.

If bidders run the McAfee-McMillan knockout, the principal's revenue would be significantly lowered relative to the non-collusive, Myersonian benchmark of $\frac{5}{12}$. The maximum value of the two bidders is above r with probability $1 - r^2$. The principal collects r in revenue whenever this event occurs. As a result, the principal's expected revenue from a FPA with $r = \frac{1}{2}$ if bidders execute the McAfee-McMillan knockout is $r(1 - r^2) = \frac{3}{8}$.

Bidders may be unable to perfectly execute the McAfee-McMillan knockout and instead may have to settle for a *noisy knockout*. The McAfee-McMillan knockout always identifies the bidder with the higher value. Achieving such an outcome may

require a trustworthy center organizing the collusive arrangement or the ability to exchange granular information about bidders' values. A collusive arrangement constrained to a decentralized organization or limited communication due to fear of being detected by a regulator may be unable to exactly carry out a McAfee-McMillan knockout.

In a noisy knockout, bidders still participate in a first-price sealed-bid pre-auction knockout, but with some probability $\alpha \in [0, 0.5)$, the lower bidder i is selected as the winner of the knockout. If b_i is lower than the reserve price r, no one enters the principal's mechanism. If b_i is weakly above r, the lower bidder is the sole participant in the principal's mechanism, buys the good at r, and pays the other bidder $b_i - r$. For all $r \in [0, 1)$ and $\alpha \in [0, 0.5)$, this game has a symmetric Bayes Nash equilibrium where bidders follow a strictly increasing bidding function in the knockout.[3] In this equilibrium, when the cartel operates correctly, the good is allocated to the bidder with the higher value if that value is above r, but when the cartel malfunctions, the good is allocated to the bidder with the lower value if it is above r.

Noisy knockouts with high α are not viable collusive arrangements. At the extreme, $\alpha \approx 0.5$ corresponds to the cartel picking a bidder to enter the principal's auction nearly uniformly at random. In this collusive arrangement, with probability nearly $\frac{1}{2}$, the good either stays with the principal or is allocated to the bidder with the lower value. This misallocation destroys total surplus. Each bidder's payoff is also relatively low at $\frac{1}{16}$.

Rather than form this collusive arrangement, bidders would prefer the outside option of playing the non-collusive BNE, which provides each bidder with a payoff of $\frac{1}{12}$. Figure 1 plots the payoff for a colluder against the probability that the collusive arrangement malfunctions. α_t is the maximum probability of malfunction such that the resulting collusive arrangement produces at least as much surplus as the BNE. Probabilities above α_t correspond to collusive arrangements that cannot form.

In general, for all $r \in [0, 1)$, there is an induced threshold α_t. Cartels with α above the threshold find it more profitable to play the BNE than to carry out their collusive arrangement. Figure 2 shows that α_t is a strictly decreasing function of r.

While the principal's revenue is threatened by a perfectly functioning McAfee-McMillan, it is even more threatened by a moderate probability of noisiness. Figure 3 plots the principal's optimal revenue, given that she is restricted to set a FPA with

[3] The proof of this claim is in the Appendix A.

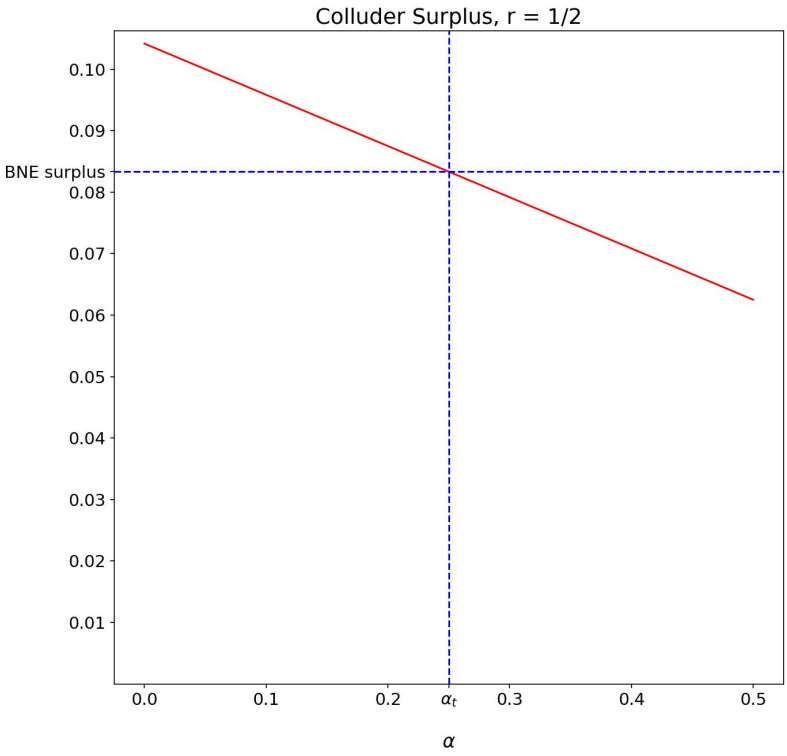


Figure 1: Probability of malfunction α is on the x-axis, and the payoff to an individual bidder participating in the α-noisy knockout is on the y-axis. The blue horizontal dashed line marks the surplus from playing the non-collusive BNE. α_t is the maximum α viable.

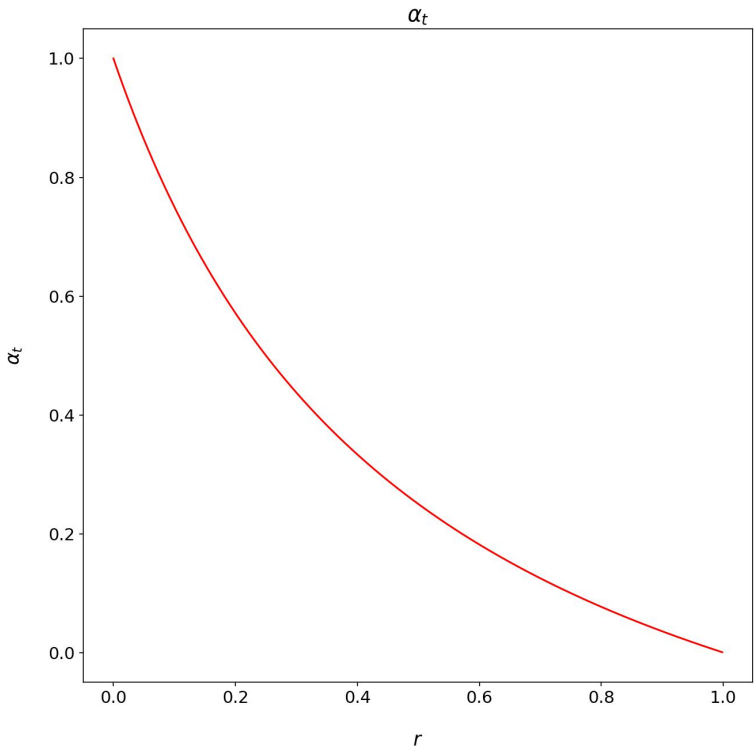


Figure 2: α_t as r varies.

a reserve price r. For each $\alpha \in [0, 0.5)$, the curve shows the revenue the principal can achieve when she optimizes the reserve price r. There is a clear minimum around $\alpha \approx 0.11$.

At $\alpha = 0$, bidders are carrying out the McAfee-McMillan knockout perfectly. The optimal reserve price at $\alpha = 0$ is $r = \frac{1}{\sqrt{3}}$, which is the optimal posted price when the principal faces a single buyer with value $\max\{v_1, v_2\}$ and results in a revenue of $\frac{2}{3\sqrt{3}}$. At first, as α increases, the principal's optimal reserve price declines slowly since it is profitable to increase the probability that she sells to the lower value bidder. Bidders still find it optimal to carry out the noisy knockout.

At the kink where α is approximately 0.11, the principal switches from allowing bidders to collude to inducing bidders to play the BNE. The optimal r is the lowest possible reserve that induces the cartel to play the BNE, and the principal's optimal reserve price jumps up. As α continues to increase, the principal's optimal reserve price declines until it reaches the Myersonian reserve price $r = \frac{1}{2}$. It remains at $r = \frac{1}{2}$ until α reaches 0.5.

Now, consider the principal's problem of choosing a revenue-maximizing mechanism when bidders can implement any viable collusive arrangement.

The principal could at least secure herself a revenue of $\frac{2}{3\sqrt{3}}$ by setting the McAfee-McMillan knockout plus a posted price of $r = \frac{1}{\sqrt{3}}$ as her mechanism. With this mechanism, the principal brings the surplus-maximizing collusive arrangement in-house, rather than allowing the bidders to implement it–or, rather, possibly fail to implement it–on their own. As McAfee and McMillan show, there is no collusive arrangement that would strictly increase bidders' surplus above the surplus of the principal's BNE.[4] It is possible to show that any collusive arrangement bidders could form must replicate the outcome of the BNE with probability 1 in order to generate as much bidders surplus as the BNE does. As a result, the principal secures the revenue of the BNE. By implementing efficient collusion on behalf of the bidders, this mechanism prevents the even worse outcome that bidders implement a compromised, dysfunctional version of the McAfee-McMillan knockout and destroy total surplus by misallocating the good.

In fact, the combination of MM knockout and FPA auction with $r = \frac{1}{\sqrt{3}}$ turns

[4]Such a collusive arrangement would contradict the McAfee and McMillan result, since the composition of the MM knockout auction and this additional collusive arrangement would raise bidder surplus above just the MM knockout auction alone.

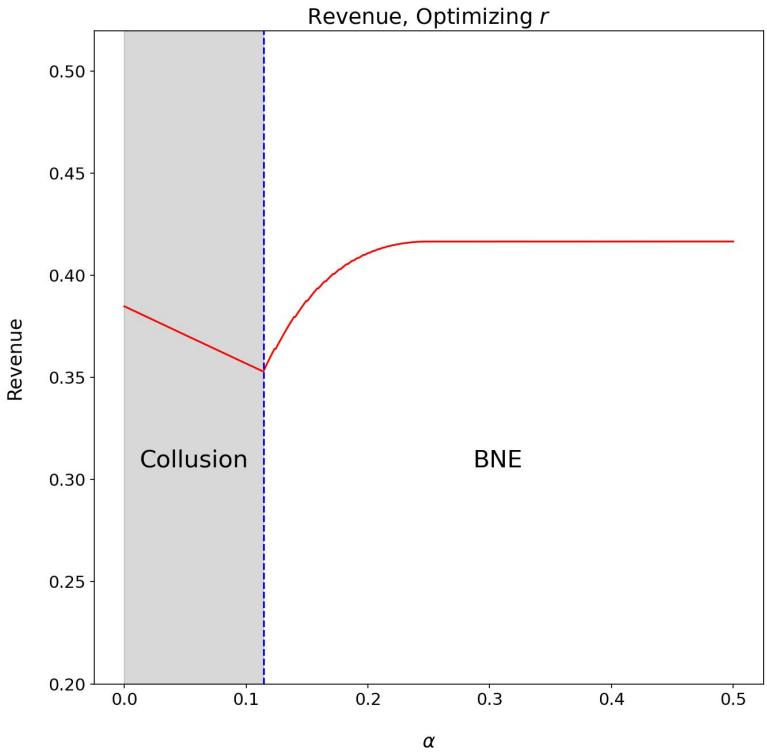


Figure 3: Revenue from a FPA with optimized reserve price r at various levels of α.

out to be the principal's optimal mechanism in general. To see this, it is necessary to observe that the principal's mechanism does not only affect collusion by defining the outside option that may tempt bidders away from colluding. The principal's mechanism also shapes collusion by setting a schedule of prices and quantities at which the cartel can acquire the good. Collusion can neutralize the competition between bidders. In the McAfee-McMillan knockout auction, a single bidder enters the principal's auction; because he faces no competitor, he can treat the principal's mechanism as a posted price and acquire the good by paying exactly the reserve price. In general, after accounting for the surplus the BNE generates for bidders, the exact form or intensity of competition that the principal's mechanism provokes is unimportant. The principal's mechanism can then be reduced to pairs of probabilities of acquiring the good and the associated total costs that the cartel must pay. This idea will be developed formally in the section that follows and in the analysis.

4 Optimal Mechanism

In this section, I present and interpret the solutions to the principal's problem. I provide a brief proof sketch that demonstrates how surplus maximization is a useful instrumental goal for the revenue-maximizing principal. Finally, I explicitly construct an optimal mechanism that solves the principal's problem.

4.1 Main Result

In this subsection, I present the solution of the principal's problem. Then, I interpret the optimal mechanism as resulting from the combination of a posted price and bringing the efficient collusive arrangement in-house.

If the principal simply posted a price, she would choose a price p^* that solves the following revenue maximization problem:

$$\max_{p\in[0,1]} \quad p\Big(1-\prod_i F_i(p)\Big) \tag{opt-price}$$

Certain qualitative features that depend on p^* ensure that a mechanism solves the principal's problem:

Theorem 1. *Any feasible mechanism s.t.*

- *in equilibrium, the good is allocated to the highest value bidder if* $\max_i v_i > p^*$
- *in equilibrium, the good is not allocated if* $\max_i v_i < p^*$,
- *and the principal collects* p^* *in total revenue whenever the good is allocated to a bidder and* 0 *otherwise,*

solves the principal's problem and attains the value opt-price. Such a mechanism exists.

Mechanisms that satisfy the conditions given in Theorem 1 allocate the good efficiently in equilibrium whenever the good is sold. While these mechanisms screen for which bidders have higher valuations, the principal does not use the screening to collect a higher revenue. Instead, she caps the total revenue collected whenever the good is sold to be p^*.

An optimal mechanism that satisfies the conditions given in Theorem 1 can be interpreted as fixing a posted price and bringing the surplus-maximizing collusive arrangement against it "in-house." To define this, let (q^{p^*}, t^{p^*}) be the direct mechanism that corresponds to the BNE of posting price p^*. A surplus-maximizing collusive arrangement against this mechanism solves the following problem:

$$\max_{(\kappa, q^c, t^c)} \quad \mathbb{E}_v \left[\sum_i q_i^c(v) v_i - \sum_{\tilde{v} \in \text{supp}(\kappa(\cdot|v))} \sum_i \kappa(\tilde{v}|v) t_i^{p^*}(\tilde{v}) \right] \tag{EC}$$

$$\text{s.t.} \quad \text{C-F}, \ \text{C-EAIR}, \ \text{C-EABB}$$

An *efficient knockout auction* is a solution to EC that is also (1) ex post budget balanced so that $\sum_i t_i^c(v) = 0$ for all v, (2) individually rational so that interim payoffs for each type of colluding bidder is weakly positive, and (3) incentive compatible so that colluding bidders are incentivized to reveal their values:

$$\mathbb{E}_{v_{-i}} \left[q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) t_i^{p^*}(a) - t_i^c(v) \right] \geq \tag{C-IC}$$

$$\mathbb{E}_{v_{-i}} \left[q_i^c(\tilde{v}_i, v_{-i}) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|\tilde{v}_i, v_{-i}))} \kappa(a|\tilde{v}_i, v_{-i}) t_i^{p^*}(a) - t_i^c(\tilde{v}_i, v_{-i}) \right] \quad \forall v_i, \tilde{v}_i, i$$

The principal can *bring an efficient knockout auction in-house* by defining a new game (A, q, t), which combines the principal's original game and the efficient collusive outcome. In this new game, each bidders can report their values as part of their possible actions, i.e., $V_i \subseteq A_i$ for each i. Reporting a profile of values v corresponds to implementing the collusive outcome for that profile v:

$$\begin{aligned} q_i(v) &= q_i^c(v) && \forall v, \forall i \\ t_i(v) &= \sum_{\tilde{v} \in \text{supp}(\kappa(\cdot|v))} \kappa(\tilde{v}|v) t_i^{p^*}(\tilde{v}) + t_i^e(v) && \forall v, \forall i \end{aligned}$$

Besides reporting a value, bidders can play their opt-out action so that $A_i \setminus V_i = \{a_i^0\}$. If any bidder plays their opt-out action at action profile a, $q_i(a) = 0$ and $t_i(a) = 0$ for all bidders i.[5]

Let σ^h denote the truthful (i.e., honest) strategy profile where each bidder v_i reports their true value to the principal. Strategy profile σ^h in (A, q, t) produces the same outcome in terms of ex post allocations and transfers as the collusive outcome (κ, q^c, t^c) defined for the original mechanism. As a result, σ^h is a Bayes-Nash equilibrium of (A, q, t). To see this, observe that since (κ, q^c, t^c) is individually rational, no bidder type would strictly prefer to play their opt-out action over reporting their type truthfully. Since (κ, q^c, t^c) is incentive compatible, no bidder type would prefer to misrepresent their value than report truthfully.

Theorem 2. *There exists an efficient knockout auction. Furthermore, for any efficient knockout auction, bringing it in-house is an optimal mechanism for the principal.*

The existence of an efficient knockout auction follows from the proof of Theorem 1, which constructs an optimal mechanism.[6] The full proof is in Appendix C.

While the proof of Theorem 2 presents a particular efficient knockout auction to demonstrate existence, there are many efficient knockout auctions.

One is identified by Mailath and Zemsky (1991) (1991). While Mailath and Zemsky consider collusion against a second-price auction with a fixed reserve price r, the collusive arrangement in their Theorem 3 is also well-defined as a collusive arrange-

[5] There are many ways to include opt-out actions. This is just one. Any definition of game at the action profiles with opt-out actions that results in a cost function $c(Q) = p^*Q$ with domain $[0, 1]$ works. See next subsection for definition and discussion of cost function.

[6] This mechanism is presented in Section 4.3.

ment against a posted price of r. This is because colluding bidders following their scheme only ever pay the principal r.

The Mailath and Zemsky collusive outcome, reformulated for the optimal posted price p^*, is given by:

$$\kappa^{MZ} = \begin{cases} (p^*, 0, ..., 0) & \text{if } \max_i v_i \geq p^* \\ (0, 0, ..., 0) & \text{if } \max_i v_i < p^* \end{cases}$$

$$q_i^{MZ}(v) = \begin{cases} \frac{1}{|\{j : v_j = \max_k v_k\}|} & \text{if } v_i = \max_k v_k \geq p^* \\ 0 & \text{if } v_i < \max_k v_k \end{cases}$$

$$t_i^{MZ}(v) = \begin{cases} \xi_i(v) - \mathbb{I}\{i = 1\}p^* + \frac{1}{|\{j : v_j = \max_k v_k\}|}p^* & \text{if } v_i = \max_k v_k \geq p^* \\ \xi_i(v) - \mathbb{I}\{i = 1\}p^* & \text{if } v_i < \max_k v_k \text{ and } \max_k v_k \geq p^* \\ \xi_i(v) & \text{if } \max_k v_k < p^* \end{cases}$$

where

$$\xi_i(v) = -\int_0^{v_i} \max\{\tilde{v}_i - p^*, 0\} d\left(\prod_{j \neq i} F_j(\tilde{v}_i)\right) + \frac{1}{N-1} \sum_{j \neq i} \int_0^{v_j} \max\{\tilde{v}_j - p^*, 0\} d\left(\prod_{k \neq j} F_k(\tilde{v}_j)\right) + \sigma_i$$

and σ_i are bidder specific constants that are parameters.

This collusive arrangement is incentive compatible and ex post budget balanced. Mailath and Zemsky (1991) (1991) show that it is possible to choose $\{\sigma_i\}_i$ so that it is also individually rational. Because the good is allocated to the bidder with the highest value whenever that value exceeds p^*, $(\kappa^{MZ}, q^{MZ}, t^{MZ})$ also solves EC, so bringing it in-house is an optimal mechanism:

Corollary 1. *Bringing $(\kappa^{MZ}, q^{MZ}, t^{MZ})$ in-house solves the principal's problem.*

When bidders are symmetric, the principal can bring a particularly interpretable collusive arrangement in-house to form an optimal mechanism. McAfee and McMillan (1992) (1992) show that a first-price sealed-bid pre-auction knockout can implement an incentive compatible collusive arrangement. In this pre-auction knockout, bidders compete for the right to enter the principal's mechanism alone. Ties are broken uniformly at random. If the highest bid in this knockout exceeds r, the winner of the pre-auction knockout enters the principal's mechanism and buys the good at the price r. The winner of the knockout then pays the losers an equal share of the difference

between his first-stage bid and r. If no one bids above r, no bidders are allowed to enter the principal's mechanism, and the game ends.

McAfee and McMillan (1992) present a symmetric Bayes-Nash equilibrium of this game, β, where bidders bid in the knockout according to a strictly increasing function of their values. In this equilibrium, the good is allocated to the highest value bidder whenever $\max_j v_j \geq r$, and the principal collects exactly r in revenue whenever the good is sold. Since colluding bidders following the McAfee-McMillan knockout only ever pay the reserve price r when they acquire the good from the principal, this collusive arrangement is also well-defined against a posted price of p^*. Furthermore, this collusive arrangement solves EC.

The following describes the collusive outcome that results from bidders running the McAfee-McMillan knockout against (q^{p^*}, t^{p^*}):

$$\kappa^{MM}(v) = \begin{cases} (p^*, 0, ..., 0) & \text{if } \max_i v_i \geq p^* \\ (0, 0, ..., 0) & \text{if } \max_i v_i < p^* \end{cases}$$

$$q_i^{MM}(v) = \begin{cases} \frac{1}{|\{j : v_j = \max_k v_k\}|} & \text{if } v_i = \max_k v_k \geq p^* \\ 0 & \text{if } v_i < \max_k v_k \end{cases}$$

$$t_i^{MM}(v) = \begin{cases} \frac{1}{|\{j : v_j = \max_k v_k\}|}\big(\beta_i(v_i)\big) & \\ -\big(1 - \frac{1}{|\{j : v_j = \max_k v_k\}|}\big)\frac{1}{N-1}\big(\beta_i(v_i) - p^*\big) - \mathbb{I}\{i = 1\}p^* & \text{if } v_i = \max_k v_k \geq p^* \\ -\frac{1}{N-1}\big(\beta_i(\max_k v_k) - p^*\big) - \mathbb{I}\{i = 1\}p^* & \text{if } v_i < \max_k v_k \text{ and } \max_k v_k \geq p^* \\ 0 & \text{if } \max_k v_k < p^* \end{cases}$$

While McAfee and McMillan did not emphasize these properties, this pre-auction knockout is also individually rational and ex post budget balanced, so like the Mailath and Zemsky collusive arrangement, the McAfee-McMillan pre-auction knockout can be brought in-house to form an optimal mechanism.

Corollary 2. *(Jagadeesan et al. (2025) 2025) Suppose bidders are symmetric so that $F_i = F$ for all i. Bringing $(\kappa^{MM}, q^{MM}, t^{MM})$ in-house solves the principal's problem.* [7]

[7] This optimal mechanism was first discussed formally in the May 2025 draft of Jagadeesan et al.

4.2 Proof Sketch

In this subsection, I sketch the proof of Theorem 1 and discuss the main ideas within it.

The proof demonstrates how the principal's mechanism affects collusion through two channels. The first is a schedule of prices and quantities at which an adversarial cartel would purchase the good from the principal. The second is the surplus corresponding to playing non-collusively, which serves as a tempting outside option for colluding bidders. I will show how these channels can be formalized and how the principal's problem can be solved by doing a nested optimization over the two channels.

Before formalizing the two channels, I establish two preliminary results. First, the space of mechanisms that the principal optimizes over can be reduced to just mechanisms that are *essentially direct*. An *essentially direct* mechanism (A, q, t, σ) is such that the set of actions for each i, A_i, is the space of i's values and i's opt-out action, $V_i \cup a_i^0$, and σ^h is the profile of truthful strategies, i.e., every type reports honestly. It is without loss of optimality for the principal to restrict attention to mechanisms that are essentially direct:

Lemma 1. *For every mechanism (A, q, t, σ), there is an essentially direct mechanism $\left(\cup_i \left(V_i \cup \hat{a}_i^0\right), \hat{q}, \hat{t}, \sigma^h\right)$ such that $minRev(A, q, t, \sigma) \leq minRev\left(\cup_i \left(V_i \cup \hat{a}_i^0\right), \hat{q}, \hat{t}, \sigma^h\right)$.*

The equivalent essentially direct mechanism induces a smaller space of possible collusive outcomes compared to (A, q, t, σ). As a result, the worst-case revenue in the essentially direct mechanism is weakly greater.

From here, I assume the principal sets an essentially direct mechanism. I suppress the dependence of mechanisms on the equilibrium σ and action space A as a result.

Second, bidders' C-EAIR constraints can be combined into a single constraint. Fix an arbitrary mechanism (q, t) that may be set by the principal. Since each bidder's expected surplus from collusion must exceed their surplus from playing the truth-telling BNE, the joint surplus of all bidders colluding must exceed the joint

(2025) (2025). The author was aware of this result and informally discussed it in the introduction of the April 18th, 2025 draft of this paper. Jagadeesan et al. (2025) (2025) also show that this particular mechanism is robust to collusion among subsets of bidders and a larger set of possible collusive arrangements.

non-collusive surplus:

$$\sum_i \mathbb{E}_{v_{-i}}[\tilde{q}_i(v)v_i - \tilde{t}_i(v) - \sum_{\hat{v}\in \text{supp}(\kappa(\cdot|v))} \kappa(\hat{v}|v)t_i(\hat{v})] \geq \sum_i U_i^n \qquad \text{(C-JEAIR)}$$

The first channel through which the principal's mechanism affects collusion is parameterized by the schedule of prices and associated probabilities of getting the good the cartel faces. It captures the cost-minimizing way a cartel can acquire the good from the principal. As a result, I call this schedule a *cost function*. This object is first studied in Pavlov (2008) (2008) and Che and Kim (2009) (2009) and is called a menu in Pavlov (2008) (2008).

The cost function is specified in the space of *aggregate outcomes*. An aggregate outcome is the total probability that some bidder gets the good and the total cost that bidders pay to the principal. The set of aggregate outcomes that can be induced by playing an action profile in the principal's mechanism is the following:

$$P_{(q,t)} = \{(\sum_i q_i(a), \sum_i t_i(a)) : a \in A\},$$

Because the cartel can possibly coordinate to randomize which action profiles are played, any point in the convex hull of $P_{(q,t)}$ can be induced as an aggregate outcome. Furthermore, the cartel may be able to dispose of the good, so any point in the following set can be induced as an aggregate outcome:

$$\hat{P}_{(q,t)} = \left\{(Q,T) : \exists(\tilde{Q},\tilde{T}) \in conv(P_{(q,t)}) \text{ s.t. } Q \leq \tilde{Q}, T = \tilde{T}, Q \geq 0\right\}$$

The cost function induced by (q,t), $c_{(q,t)}$ is a mapping from domain $[0, \sup_a \sum_i q_i(a))$ or $[0, \sup_a \sum_i q_i(a)]$ to $\mathbb{R} \cup -\infty$ such that $c_{(q,t)}(Q) = \inf\left\{T : (Q,T) \in \hat{P}_{(q,t)}\right\}$ Figure 4 demonstrates the construction of the cost function when the principal sets a first-price auction with a "strict" reserve price $r = 0.5$ so that the good is sold only if the highest bid strictly exceeds r.

Every mechanism (q,t) the principal sets induces an associated cost function $c_{(q,t)}$. All cost functions have the following properties:

Lemma 2. *For any (q,t), the induced cost function c is (1) continuous on $[0,\bar{q})$ where $\bar{q} := \sup_{a\in A} \sum_i q_i(a)$, (2) convex, and (3) increasing. Additionally, (4) $c(0) \leq 0$.*

That $c(0) \leq 0$ is a consequence of the opt-out actions in the principal's mechanism.

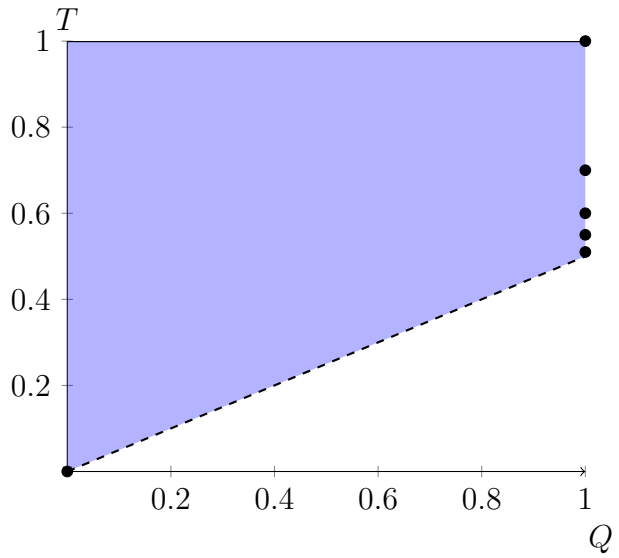


Figure 4: Feasible aggregate outcomes for FPA with a strict reserve price of 0.5 are represented by the blue region. The lower boundary of the set $c(Q) = 0.5Q$ is the cost function and is not included.

The cartel may coordinate all agents on playing their opt-out actions in order to guarantee no one pays anything when no one gets the good. Including the opt-out actions allows the cartel to walk away from the mechanism.[8] From here, I call any function $c : [0, \bar{q}] \to \mathbb{R}$ or $c : [0, \bar{q}) \to \mathbb{R}$ where $\bar{q} \leq 1$ a *cost function* if it satisfies the properties given in Lemma 2.

A mechanism (q, t) that induces an unbounded cost function has $\text{minRev}(q, t) = -\infty$.[9] From here, I restrict attention just to mechanisms with bounded cost functions.

Cost functions can be ranked. Consider two mechanisms, (q, t) and $(\hat{q}, \hat{t})$. Let $\hat{\bar{q}}$ be the supremum of the domain of $c_{(\hat{q},\hat{t})}$. Cost functions are ranked so that $c_{(\hat{q},\hat{t})} \geq c_{(q,t)}$ if for every Q in the domain of $c_{(\hat{q},\hat{t})}$, $c_{(\hat{q},\hat{t})}(Q) \geq c_{(q,t)}(Q)$. The adversarial cartel acquiring the good for any probability Q in mechanism $(\hat{q}, \hat{t})$ can acquire the good

[8] In contrast, in Che and Kim (2006) (2006) where the principal is not required to include opt-out actions, the optimal collusion-proof mechanism is constructed so that every action profile results in the same total revenue for the principal. In this sense, the optimal mechanism in Che and Kim (2006) (2006) traps the cartel and forces it to participate.

[9] Suppose (q, t) has an unbounded cost function. Then, there exists $Q \in [0, 1]$ s.t. $c(Q) < -\sum_i U_i^n - K$ for any $K > 0$. As a result, there exists $\kappa \in \Delta A$ such that $\sum_{a \in \text{supp}(\kappa)} \sum_i t_i(a)$ is within $\varepsilon > 0$ of $c(Q)$. Then, for sufficiently small ε, a collusive outcome where the cartel to carries κ for all v and makes constant payment T_i to each cartel member i where $T_i > U_i$ is possible. This collusive outcome results in revenue $-\sum_i U_i^n - K$, so worst-case revenue is unbounded.

for a weakly lower or arbitrarily close cost in mechanism (q,t).

When two mechanisms (q,t) and $(\hat{q},\hat{t})$ induce ranked cost functions and there is a strict difference in the total bidder surplus each mechanism generates, it may be possible to determine the principal's preferences between the two mechanisms without calculating the worst-case collusive arrangement:

Lemma 3. *Fix mechanisms (q,t) and $(\hat{q},\hat{t})$. Let $\sum_i U_i^n$ be the total bidder surplus generated in the truth-telling BNE of (q,t). Similarly, let $\sum_i \hat{U}_i^n$ be the total bidder surplus generated in $(\hat{q},\hat{t})$. Suppose $c_{(\hat{q},\hat{t})} \geq c_{(q,t)}$. If $\sum_i \hat{U}_i^n > \sum_i U_i^n$, then $minRev(\hat{q},\hat{t}) \geq minRev(q,t)$.*

For each mechanism (q,t) that induces a cost function c, Lemma 3 says that any mechanism $(\hat{q},\hat{t})$ that induces a cost function $\hat{c} \geq c$ and results in strictly greater bidder surplus also induces a weakly higher worst-case revenue. In this way, surplus maximization is a useful instrumental goal for the revenue-maximizing principal.

In fact, the principal's problem can be solved by a nested optimization, which involves a constrained surplus maximization problem as the inner problem. The inner problem is parameterized by a bounded, continuous cost function with compact domain. I call such a cost function *well-behaved*. It turns out that it is without loss of optimality for the principal to restrict attention just to the mechanisms that solve the inner problem for some well-behaved cost function.

The outer problem then optimizes over well-behaved cost functions. An optimal mechanism is the mechanism that solves the inner problem given the optimal choice of cost function.

In the inner problem, the principal maximizes the bidder surplus, optimizing over IC, IR direct mechanisms that induce cost functions above continuous function $c : [0,\bar{q}] \to \mathbb{R}$:

$$\begin{aligned}
\max_{q:V\to[0,1]^N, t:V\to\mathbb{R}} \quad & \mathbb{E}_v[\sum_i q_i(v)v_i - t_i(v)] && (c\text{-Surplus Max})\\
\text{s.t.} \quad & \sum_i q_i(v) \leq \bar{q} \quad \forall v \\
& \mathbb{E}_{v_{-i}}[q_i(v)v_i - t_i(v)] \geq \mathbb{E}_{v_{-i}}[q_i(\tilde{v}_i, v_{-i})v_i - t_i(\tilde{v}_i, v_{-i})] \quad \forall v_i, \tilde{v}_i, i \\
& \mathbb{E}_{v_{-i}}[q_i(v)v_i - t_i(v)] \geq 0 \quad \forall i \\
& \sum_i t_i(v) \geq c(\sum_i q_i(v)) \quad \forall v
\end{aligned}$$

By choosing a direct mechanism, the principal chooses the truth-telling BNE of her essentially direct mechanism. c-Surplus Max also drops the requirement that the mechanism include opt-out actions. That the direct mechanism induced by the BNE satisfies IR is necessary. It is also sufficient for the direct mechanism to be the BNE of an essentially direct mechanism since opt-out actions can be included in a way that ensures truth-telling is still an equilibrium. This is explored in detail in B.5.3.

It is not obvious ex ante that mechanisms that solve c-Surplus Max are without loss for the principal to optimize over. Lemma 3 does not tell us the principal's preferences between mechanisms that result in the same bidder surplus and ranked cost functions, so it is not clear what the principal's preferences would be among the (possibly multiple) mechanisms that solve c-Surplus Max for a given c. Furthermore, c-Surplus Max is formulated only for well-behaved cost functions. It is not clear that the principal should prefer the solution of c-Surplus Max for some continuous c to a mechanism that induces a cost function $\hat{c}$ that is discontinuous or has a domain that does not include its supremum, $\bar{q}$.

However, the properties of the mechanisms that solve the inner problem enable me to close these gaps. It turns out that for every possible c, all solutions of c-Surplus Max induce the same worst-case revenue, so no additional argument is needed to pick out the principal's preferred mechanism among the solutions of c-Surplus Max. Furthermore, given a mechanism (q, t) that induces a bounded cost function $\hat{c}$, $\hat{c}$ can be mapped to a well-behaved cost function c such that $\hat{c} \geq c$. Additionally, the solution of c-Surplus Max parameterized by c results in a weakly greater worst-case revenue than (q, t).

The solutions of c-Surplus Max also have several qualities that result in an easy outer optimization. First, the solutions of c-Surplus Max also solve a relaxed problem that drops the incentive compatibility (IC) and individual rationality (IR) constraints. As a result, these mechanisms also solve pointwise surplus maximization problems and allocate the good efficiently among the bidders. Second, because the solutions of c-Surplus Max solve the relaxed problem without IC or IR constraints, no collusive arrangement against these mechanisms can result in strictly greater surplus for the bidders. As a result, any collusive arrangement that results in the same surplus for the bidders must replicate the outcome of the mechanism. Thus, the worst-case revenue of these mechanisms is the revenue that results from the non-collusive BNE.

The principal's outer optimization over cost functions is akin to a single-agent

screening problem. The "agent" the outer problem faces is the pointwise surplus-maximization problem. The pointwise problem "buys" the good at the probability and price that maximizes bidder surplus given the value $\max_i v_i$.

Given Myerson (1981) (1981), the solution of the outer problem is to post a price. The optimal price is p^*. This corresponds to choosing a linear cost function with slope p^* and domain $[0, 1]$. One optimal mechanism is the essentially direct mechanism constructed from the solution of c-Surplus Max given this cost function. This mechanism has worst-case revenue opt-price. Finally, any mechanism with the properties given in Theorem 1 achieves worst-case revenue opt-price and so solves the principal's problem.

4.3 One Optimal Mechanism

In this subsection, I present the explicit formula for an optimal mechanism and describe its construction.

Given Theorem 1, an optimal mechanism (q^{pk}, t^{pk}) has the following ex post allocation function:

$$q_i^{pk}(v) = \begin{cases} 1 & \text{if } v_i = \max_j v_j \geq p^* \\ 0 & \text{otherwise} \end{cases} \quad \forall v$$

Additionally, t^{pk} must satisfy the following restriction on the transfers:

$$\sum_i t_i^{pk}(v) = \begin{cases} p^* & \text{if } v_i = \max_j v_j \geq p^* \\ 0 & \text{otherwise} \end{cases} \quad \forall v$$

Given that truth-telling is a BNE, the total surplus of all bidders' lowest types, denoted $\sum_i U_i(0)$, can be calculated using just interim allocation and the ex ante expected total transfer. It is possible to show that $\sum_i U_i(0)$ is weakly greater than 0, so it is possible to split the total among the bidders so that each bidder's lowest type receives at least 0. Given these $\{U_i(0)\}_i$, the interim transfers of each i, $\bar{t}_i^{pk}(\cdot)$, are pinned down by the envelope theorem and the interim allocation functions.

To define the ex post transfer function, define constants γ_i for each i:

$$\gamma_i = \frac{\mathbb{E}_{v_i}[\bar{t}_i^{pk}(v_i)]}{Pr(\max v \geq p)p}$$

γ_i is in $[0,1]$ for each i, and $\sum_i \gamma_i = 1$. One optimal transfer function is partially given by the following:

$$t_i^{pk}(v) = \gamma_i p^* \mathbb{I}\{\max v \geq p^*\} + \bar{t}_i^{pk}(v_i) - \gamma_i \mathbb{E}_{v_{-i}}[p^* \mathbb{I}\{\max v \geq p^*\}]$$
$$- \frac{1}{N-1}\Big(\sum_{j \neq i} \bar{t}_j^{pk}(v_j) - \gamma_j \mathbb{E}_{v_{-j}}[p^* \mathbb{I}\{\max v \geq p^*\}]\Big)$$

Since $U_i(0) \geq 0$ for each i, opt-out actions can easily be added to the direct mechanism make this an essentially direct mechanism. Let $q_j^{pk}(a_i^0, a_{-i}) = t_j^{pk}(a_i^0, a_{-i}) = 0$ for all i, j, and $a_{-i} \in \cup_{j \neq i} \left(V_j \cup a_j^0\right)$. Since each type of each bidder receives a payoff that is at least 0, deviating to playing the opt-out action is not tempting. Truth-telling is indeed a BNE of this essentially direct mechanism.

5 Extensions

In this section, I vary the assumptions of the baseline model to test the robustness of Theorem 1 to alternative assumptions. In the first subsection, I vary the assumptions on the collusive outcomes that are possible. In the second subsection, I restrict the principal to setting mechanisms that do not make payments to the bidders.

5.1 Additional Assumptions on Collusion

In this subsection, I add additional restrictions on the collusive outcomes the principal considers possible. I investigate whether the optimal mechanisms identified in Theorem 1 are still optimal if the principal knows that collusive outcomes are constrained by more frictions than C-F, C-EABB, and C-EAIR.

The first additional constraint I consider is an incentive compatibility constraint. A cartel may be unable to pool truth-telling incentives across instances of play. Bidders may have to be motivated to reveal their private information in the short-term,

so collusive outcomes may be restricted by a truth-telling constraint:

$$\mathbb{E}_{v_{-i}}\left[q_i^c(v)v_i - \sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)t_i(a) - t_i^c(v)\right] \geq \tag{C-IC}$$
$$\mathbb{E}_{v_{-i}}\left[q_i^c(\tilde{v}_i, v_{-i})v_i - \sum_{a\in\text{supp}(\kappa(\cdot|\tilde{v}_i,v_{-i}))}\kappa(a|\tilde{v}_i, v_{-i})t_i(a) - t_i^c(\tilde{v}_i, v_{-i})\right] \quad \forall v_i, \tilde{v}_i, i$$

Cartels as conceptualized by McAfee and McMillan (1992) (1992), Mailath and Zemsky (1991) (1991) and Graham and Marshall (1987) (1987) were restricted by interim incentive constraints.

I also consider the possibility that bidders can reallocate the good but cannot throw it away:

$$\sum_i q_i^c(v) = \sum_{a\in\text{supp}(\kappa(\cdot|v))}\sum_i \kappa(\hat{a}|v)q_i^c(\hat{a}) \quad \forall v \tag{C-F'}$$

For example, in a procurement setting, the good being auctioned off corresponds to a contract to produce something for the principal. Bidders' values can be interpreted as their private costs of producing the good. The bids are the prices the producers offer to make the good at. As discussed in Section 2, if the cartel can mandate that bidders produce in an inefficient manner that raises their marginal costs, this can be interpreted as throwing away the good or throwing away the surplus from the contract. A less powerful cartel may be unable to monitor how bidders produce the good, so cartels may have to allocate the good to a bidder whenever they get it.

The last additional constraint I consider is that colluding bidders maximize their surplus. This constraint can be interpreted as a result of the principal's confidence that the other constraints are the only constraints that the cartel faces. As a result, the cartel maximizes joint surplus subject to the other constraints.

Since a mechanism may induce a set of possible aggregate outcomes that is not closed, bidders may be unable to achieve a maximum joint surplus. To get around this technical issue, the principal calculates the supremum of bidders' joint surplus

that can arise for each (A, q, t, σ):

$$S(A,q,t,\sigma) := \sup_{(\kappa,q^c,t^c)} \mathbb{E}_v \left[\sum_i q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right]$$
$$\text{s.t.} \quad \text{C-F'}, \text{C-EABB}, \text{C-EAIR}, \text{C-IC}$$

The principal evaluates mechanism (A, q, t, σ) according to the following:

$$\text{minRevR}(A,q,t,\sigma) := \liminf_{\varepsilon \to 0} \Bigg\{ \mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] : (\kappa, q^c, t^c) \text{ satisfies C-F', C-EABB, C-EAIR, C-IC, and } \left| \mathbb{E}_v \left[\sum_i q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] - S(A,q,t,\sigma) \right| < \varepsilon \Bigg\}$$

As before, the principal's problem is to maximize the revenue she can guarantee:

$$\max_{(A,q,t,\sigma)} \quad \text{minRevR}(A,q,t,\sigma) \qquad \text{(P-R)}$$
$$\text{s.t. } (A,q,t,\sigma) \text{ is feasible}$$

Under these additional assumptions on collusive outcomes, the mechanisms identified as optimal under the baseline model remain optimal:

Theorem 3. *Any feasible mechanism s.t.*

- *in equilibrium, the good is allocated efficiently among the bidders whenever* $\max_i v_i > p^*$ *and is not allocated to a bidder when* $\max_i v_i < p^*$,
- *the principal collects exactly* p^* *in revenue whenever the good is allocated to a bidder and* 0 *otherwise.*

solves P-R and attains value opt-price.

The proof proceeds along the same lines as the proof of the main theorem. For every feasible mechanism, I construct a *restricted cost function* in a process similar to the construction of the cost function. I then show that for every feasible mechanism (A, q, t, σ), the principal prefers a mechanism that solves a surplus maximization

problem c-Surplus Max for a related restricted cost function. As a result, it is without loss of optimality for the principal to restrict attention to mechanisms that solve c-Surplus Max for a restricted cost function. Then, the principal optimizes over cost functions as in the proof of Theorem 1. The details are in Appendix F.

If bidders are assumed to maximize their surplus, then the principal can simply post a price of p^* as her mechanism. Intuitively, the principal does not have to include efficient collusion in her mechanism if bidders are assumed to carry it out:

Corollary 3. *Posting a price of p^* solves P-R. Specifically, let $A_i = \{buy, don't\ buy\}$ for each i.*

$$q_i^p(a) = \begin{cases} \frac{1}{|\{j: a_j = \ buy\}|} & \text{if } a_i = \ buy \\ 0 & \text{if } a_i = \ don't\ buy \end{cases} \quad \forall a, \forall i$$

$$t_i^p(a) = \begin{cases} \frac{1}{|\{j: a_j = \ buy\}|} p^* & \text{if } a_i = \ buy \\ 0 & \text{if } a_i = \ don't\ buy \end{cases} \quad \forall a, \forall i$$

$$\sigma_i(v_i) = \begin{cases} buy & \text{if } v_i \geq p^* \\ don't\ buy & \text{if } v_i < p^* \end{cases} \quad \forall v_i, \forall i$$

(A, q^p, t^p, σ) solves P-R.

In contrast, under the baseline model, posting a price p^* is not an optimal mechanism. By simply posting a price, the principal has limited ability to screen the values of the bidders. If multiple bidders play buy, the principal cannot determine who has the highest value and so cannot allocate the good efficiently. Because posting a price does not efficiently allocate the good, the bidders can improve their surplus by reallocating the good as part of their collusive arrangement. By boosting their surplus via reallocation, bidders can buy the good with probability lower than $1 - \prod_i F_i(p^*)$ and still satisfy C-EAIR. As a result, the worst-case revenue of posting a price p^* in the baseline model is lower than opt-price.

I now consider alternative models where collusive outcomes constrained by a subset of the assumptions I have introduced. I parameterize these alternative models with a set $\mathcal{A}$. $\mathcal{A}$ contains the assumptions that constrain collusive outcomes and must contain at a minimum the constraints of the baseline model:

$$\{\text{C-F, C-EABB, C-EAIR}\} \subseteq \mathcal{A} \subseteq \{\text{C-F, C-F', C-EABB, C-EAIR, C-IC, surplus maximization}\}$$

The $\mathcal{A}$ that corresponds to Theorem 3 is {C-F', C-EABB, C-EAIR, C-IC, surplus maximization}. Since C-F is implied by C-F', this $\mathcal{A}$ corresponds to the set on the right of the above display equation. Thus, the alternative models I consider are "in between" the baseline model and the model of Theorem 3.

Given a set of assumptions $\mathcal{A}$, I consider an analogous exercise where the principal maximizes the worst-case revenue that could arise given collusive outcomes are constrained $\mathcal{A}$. For any set $\mathcal{A}$ that does not include surplus maximization, the worst-case revenue of a feasible mechanism (A, q, t, σ) is given by

$$\text{minRev}\mathcal{A}(A, q, t, \sigma) := \inf_{(\kappa, q^c, t^c)} \mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right]$$
$$\text{s.t. } \mathcal{A}$$

If $\mathcal{A}$ also contains the assumption that colluding bidders maximize their surplus, then as before, it may be impossible for collusive outcomes to exactly achieve the supremal bidder surplus possible. As before, the principal calculates the supremum:

$$S_{\mathcal{A}}(A, q, t, \sigma) := \sup_{(\kappa, q^c, t^c)} \mathbb{E}_v \left[\sum_i q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right]$$
$$\text{s.t. } \mathcal{A} \setminus \text{surplus maximization}$$

The principal calculates the worst-case revenue of a feasible mechanism by

$$\text{minRev}\mathcal{A}(A, q, t, \sigma) := \text{liminf}_{\varepsilon \to 0} \Big\{ \mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] : (\kappa, q^c, t^c) \text{ satisfies}$$
$$\mathcal{A} \setminus \text{surplus maximization and}$$
$$\left| \mathbb{E}_v \left[\sum_i q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] - S_{\mathcal{A}}(A, q, t, \sigma) \right| < \varepsilon \Big\}$$

The principal's problem given $\mathcal{A}$ is to maximize worst-case revenue:

$$\max_{(A,q,t,\sigma)} \text{minRev}\mathcal{A}(A,q,t,\sigma) \qquad \text{(P-}\mathcal{A}\text{)}$$
$$\text{s.t. } (A,q,t,\sigma) \text{ is feasible}$$

Under models in between the baseline model and that of Theorem 3, the optimal mechanisms described by Theorem 1 are still optimal:

Theorem 4. *For sets of assumptions $\mathcal{A}$ such that C-F, C-EABB, C-EAIR $\in \mathcal{A}$, any mechanism that satisfies the conditions of Theorem 1 solves P-$\mathcal{A}$. Furthermore, the problem has value opt-price.*

The detailed proof is in Appendix F.

5.2 Non-Negative Transfers

In this subsection, I assume that the principal cannot make payments to the bidders. Under this assumption, I identify the qualitative features of the optimal mechanism in a restricted class of mechanisms. These qualitative features demonstrate that the optimal mechanisms identified in Theorem 1 are not feasible if the principal cannot make payments to the bidders.

The optimal mechanism constructed in Section 4.3 and the McAfee-McMillan knockout combined with a posted price involve the principal making payments to bidders, but it is unclear whether all mechanisms that satisfy Theorem 1 would require this. Making payments to the bidders is a feature rarely seen in practice, so for reasons outside the model, the principal may be unable or unwilling to make them.

For tractability, I focus on the special case where the principal is restricted to setting a mechanism (A,q,t,σ) that induces a linear cost function $c(Q) = pQ$ with domain $[0,1]$. This mechanism must also includes the graph of c in $\hat{P}_{(q,t)}$. I call these mechanisms *closed and linear*.

Closed and linear mechanisms have a fixed minimum price per unit of probability, which, for example, could be induced by a deterministic reserve price. This class of mechanisms includes common mechanisms such as a posted price, a first-price auction with a reserve price, and a second-price auction with a reserve price. Mechanisms that satisfy the conditions in Theorem 1 are also closed and linear, where p^* is the slope of the induced cost function.

Furthermore, I assume that bidders are symmetric so that $F_i = F_j$ for all bidders i and j.

For every closed and linear feasible mechanism (A, q, t, σ), the principal evaluates it according to $\text{minRev}(A, q, t, \sigma)$ as defined in Section 2. The principal's problem is

$$\begin{aligned} \max_{(A,q,t,\sigma)} \ & \text{minRev}(A, q, t, \sigma) \\ \text{s.t.} \ & (A, q, t, \sigma) \text{ is feasible, closed, and linear} \\ & t \geq 0 \end{aligned} \tag{P-NN}$$

As in the baseline model, the principal's optimal mechanism is without loss of optimality an essentially direct mechanism. Lemma 1 goes through once more. From here, I restrict attention to essentially direct mechanisms for notational ease. As before, I suppress the dependence on the space of actions and the principal's chosen BNE.

Also, as in the baseline model, the principal's optimal mechanism is the solution of a constrained surplus maximization problem without loss of optimality.

Lemma 4. *Let (q, t) be a feasible, closed, and linear mechanism that induces cost function $c(Q) = pq$. Let (q^*, t^*) solve the following problem:*

$$\begin{aligned} \max_{q:V\to[0,1]^N, t:V\to\mathbb{R}_+} \ & \mathbb{E}_v[\sum_i q_i(v)v_i - t_i(v)] \\ s.t. \ & \sum_i q_i(v) \leq \bar{q} \quad \forall v \\ & \mathbb{E}_{v_{-i}}[q_i(v)v_i - t_i(v)] \geq \mathbb{E}_{v_{-i}}[q_i(\tilde{v}_i, v_{-i})v_i - t_i(\tilde{v}_i, v_{-i})] \quad \forall v_i, \tilde{v}_i, i \\ & \mathbb{E}_{v_{-i}}[q_i(v)v_i - t_i(v)] \geq 0 \quad \forall i \\ & \sum_i t_i(v) \geq c(\sum_i q_i(v)) \quad \forall v \end{aligned} \tag{NN-c-SM}$$

Then, $\text{minRev}(q^*, t^*) \geq \text{minRef}(q, t)$.

The key feature of the optimal mechanism is its ex post transfer function. Consider mechanisms that have transfer functions of the following form:

$$t_i^{\hat{v}_t}(v) = \begin{cases} p & \text{if } v_i = \max_j v_j \geq \hat{v}_t \\ 0 & \text{if o/w} \end{cases}$$

where v_t is exogenous and in $[0,1]$. In these mechanisms, if there is a bidder with a value above the threshold value v_t, the bidder with the highest realized value pays p. This ex post transfer function induces an interim transfer function:

$$\bar{t}_i^{\hat{v}_t}(v_i) = \begin{cases} 0 & \text{if } v_i < \hat{v}_t \\ F^{N-1}(v_i)p & \text{if } v_i \geq \hat{v}_t \end{cases}$$

Given incentive compatibility, the interim allocation function is given by the payoff formula, which allows us to write the following differential equation to describe $\bar{q}_i^{\hat{v}_t}(v_i)$ for $v_i \geq \hat{v}_t$:

$$\frac{d\bar{q}_i^{\hat{v}_t}}{dv_i} = \frac{p}{v_i} \underbrace{(N-1)F^{N-2}(v_i)f(v_i)}_{=:\tilde{f}(v_i)}$$

where $\tilde{f}(v_i)$ is defined as the derivative of $F^{N-1}(v_i)$. From here, I also define $\tilde{F}(v_i) := F^{N-1}(v_i)$. $\bar{q}_i^{\hat{v}_t}$ has a positive derivative when $v_i \geq \hat{v}_t$. If there exists an ex post allocation function $q\hat{v}_t$ that induces $\bar{q}_i^{\hat{v}_t}$ as the interim allocation for each i, (q,t) would satisfy incentive compatibility.

Note that for any value of $\hat{v}_t$, the lowest type of each agent must be given 0 surplus since $\bar{t}_i^{\hat{v}_t}(0) = 0$ and no surplus can be given to a type with value 0 by giving them the good. As a result, $\bar{q}_i^{\hat{v}_t}(\hat{v}_t)$ must satisfy:

$$\bar{q}_i^{\hat{v}_t}(\hat{v}_t)v_t - \bar{t}_i^{\hat{v}_t}(\hat{v}_t) = 0$$

So, if $v_t \neq 0$, we have that $\bar{q}_i^{\hat{v}_t}(\hat{v}_t) = \frac{\bar{t}_i^{\hat{v}_t}(\hat{v}_t)}{\hat{v}_t}$.

Fixing p, this family of mechanisms is parameterized by $\hat{v}_t$. It turns out that there is a smallest $\hat{v}_t$ at which it is possible to find an ex post allocation function that induces the interim allocation functions specified:

Lemma 5. *Fix a $p \in (0,1)$. There exists a minimum $\hat{v}_t$ for which there exists an ex post allocation function that induces the interim allocation function $\bar{q}^{\hat{v}_t}$. Denote this minimum by v_t. Additionally, if $v_t > 0$, then*

$$\frac{1 - F(v_t)^N}{N} = \frac{F^{N-1}(v_t)(1 - F(v_t))p}{v_t} + \int_{v_t}^{1} \frac{p}{\tilde{v}} \tilde{f}(\tilde{v})(1 - F(\tilde{v}))d\tilde{v}$$

Under some conditions, the principal's optimal mechanism among closed, linear mechanisms with non-negative transfers is the mechanism with transfers given by t^{v_t} and interim allocation $\bar{q}^{v_t}$, where v_t is as defined in the lemma above.

Proposition 1. *Suppose for all $p \in (0,1)$, v_t, as described in Lemma 5, is interior and $\frac{p\int_{v_t}^1 \hat{v}dF_i(\hat{v})}{f_i(v_t)(v_t)(p-v_t)+p(1-F_i(v_t))} \geq \mathbb{E}_{v\sim F_i}[v]$. Furthermore, suppose $\frac{1-F_i(v_i)}{f_i(v_i)}$ is decreasing and differentiable in v_i for each bidder i and that the following condition holds*

$$\frac{2}{v}\frac{1-F_i(v)}{f_i(v)} - \frac{d}{dv}\left(\frac{1-F_i(v)}{f_i(v)}\right) \leq \frac{\mathbb{E}[\hat{v}] + \mathbb{E}[\hat{v}|\hat{v}\geq v] - v}{\mathbb{E}[\hat{v}|\hat{v}\geq v] - \mathbb{E}[\hat{v}]} \quad \forall v \in (0,1) \qquad \text{(BIH)}$$

Then, the principal's optimal closed and linear mechanism (q^{nnl}, t^{nnl}) is such that:

$$t_i^{nnl}(v) = \begin{cases} p^{nnl} & \text{if } v_i = \max_j v_j \geq v_t^{nnl} \\ 0 & \text{if } o/w \end{cases} \quad \forall i, v$$

and

$$\bar{q}_i^{nnl}(v_t^{nnl}) = \frac{F^{N-1}(v_t^{nnl})p^{nnl}}{v_t^{nnl}}$$
$$\bar{q}_i^{nnl}(v_i) = \int_{v_t^{nnl}}^{v_i} \frac{p^{nnl}}{\tilde{v}}(N-1)F^{N-2}(\tilde{v})f(\tilde{v})d\tilde{v} + \bar{q}_i(v_t^{nnl}) \quad \forall v_i \in [v_t^{nnl}, 1]$$

where $p^{nnl} \in (0,1)$.

The proof is in Appendix G.0.2 and uses weak duality to establish the claim. Condition BIH is satisfied by a variety of distributions including the uniform distribution and power distributions $F_i(v_i) = v_i^\alpha$ with $\alpha > 1$. The binding IC and IR constraints turn out to be the local upward IC constraints for all types above a threshold v_t^{IR}, which is weakly below v_t, and a range of IR constraints for types between 0 and v_t^{IR}.

6 Related Literature

The extent to which collusion might threaten revenue in auctions was first studied in settings where the principal's auction format is fixed. Graham and Marshall (1987) (1987) and Mailath and Zemsky (1991) (1991) study surplus-maximizing collusion against a second-price auction with a fixed reserve price, while McAfee and McMillan

(1992) (1992) study surplus-maximizing collusion against a first-price auction with fixed reserve. All three papers focus on collusive arrangements that must be incentive compatible, but ultimately, colluding bidders are able to allocate the good efficiently among themselves.

How the principal should design the entire auction in response to collusion has largely been studied in models that fully specify how collusion can function and forms. This strand of the literature was initiated by Laffont and Martimort (1997) (1997) and Laffont and Martimort (2000) (2000) who consider collusion in contracting settings. In these papers and related papers such as Che and Kim (2009) (2009) and Pavlov (2008) (2008), the constraints on collusive arrangements are explicitly modeled. For example, bidders decide whether to participate in collusion after they learn their values, so collusive arrangements must satisfy an interim participation constraint. Colluding bidders are then assumed to implement interim Pareto efficient arrangements in response.

A notable exception is Che and Kim (2006) (2006), which allows for a very large set of possible collusive behaviors, larger than the set I consider as possible, and shows that the principal can secure her payoff in the presence of any collusive arrangement. While their model minimally constrains collusion and in this sense is agnostic about what collusive arrangements arise, their very strong positive conclusion depends on cartels being unable to exit the principal's mechanism. Follow-up studies like Che and Kim (2009) (2009) and Pavlov (2008) (2008) require the principal to include opt-out actions in her mechanism to remedy this. I also require opt-out actions as well.

The most closely related papers in this literature are Che and Kim (2009) (2009) and Pavlov (2008) (2008). Given that collusive arrangements in these models must satisfy an interim participation constraint and must be incentive compatible, the principal can focus without loss of optimality on "collusion-proof" mechanisms. These mechanisms are interim Pareto efficient, so colluders cannot improve upon them and must replicate the non-collusive BNE. Che and Kim (2009) (2009) and Pavlov (2008) (2008) show that under certain distributional assumptions, the principal can still achieve his optimal payoff assuming no collusion in the presence of collusion.

My paper differs in the basic constraints on collusion assumed. First, in my setting, focusing on collusion-proof mechanisms is not without loss since collusive arrangements may fail individual rationality and incentive compatibility. Second, in Che and Kim (2009) (2009) and Pavlov (2008) (2008), bidders choose whether

to participate in collusion based on their interim payoffs, so collusion is implicitly framed as a short-lived arrangement. Additionally, I take the worst-case over possible collusive outcomes as in the robust mechanism design literature.

Broadly, the robust mechanism design literature has revisited existing mechanism design problems with weaker assumptions about features of the environment, e.g., the information structure between bidders or the solution concept. Carroll (2019) (2019) provides an excellent survey. A related paper that also takes a worst-case approach to bidder interactions in auctions is Carroll and Segal (2019) (2019), which considers the optimal mechanism for a revenue-maximizing principal who is uncertain about the exact process through which bidders could resell the good after the auction. The principal in this paper is concerned about a much larger set of possible bidder behaviors, which includes Carroll and Segal's resale procedures.

Finally, Jagadeesan et al. (2025) (2025) contemporaneously establish similar results but with a different starting point and different solution methods. Like this paper, they study the design of a revenue-maximizing auction when bidders can form a collusive arrangement to play the principal's mechanism. The first theorem of Jagadeesan et al. (2025) (2025) corresponds to Corollary 3 of this paper, which studies the optimal mechanism when colluding bidders are assumed to maximize their surplus. Jagadeesan, Segal, and Tordjman further consider how to design the mechanism when arbitrary subsets of bidders can collude and when a much larger set of collusive arrangements can arise. This set contains all collusive arrangements I consider in the main model. The proof strategy in this paper centers on the principal's proximate goal to maximize bidder surplus in order to forestall collusion. This auxiliary problem does not arise in Jagadeesan et al. (2025) (2025).

7 Conclusion

In this paper, I study how a principal should design an auction when bidders may be colluding in rich, varied ways. Under a variety of assumptions on possibly collusive outcomes, the principal's optimal mechanism to post a price and bring the efficient collusive arrangement in-house. Then, I consider a principal who cannot set any mechanism. When the principal is constrained to set a mechanism where she cannot pay the bidders, the optimal mechanism compromises efficient allocation in contrast to the optimal mechanism under the baseline model.

Two key qualitative insights emerge from the results. First, the principal should maximize the bidder surplus generated by non-collusive play in order to stave off collusion by making it less attractive. Second, the principal should bring efficient collusion in-house. This leaves no scope for bidders to deviate from the non-collusive outcome and so prevents bidders from implementing inefficient collusive outcomes which lower total surplus and may threaten the principal's revenue. These insights seem portable beyond this context.

A Example Equilibrium

For $\alpha = 0$, McAfee and McMillan (1992) provide a symmetric BNE where bidders follow a strictly increasing bidding function.

Consider $\alpha \in (0, \frac{1}{2})$. There is a symmetric BNE of the malfunctioning McAfee-McMillan knockout where bidders bid according to the following function:

$$\beta(v_i) = \begin{cases} 0 \text{ if } v_i < r \\ c_1(\alpha - 2\alpha v_i + v_i)^{-(2-4\alpha)/(1-2\alpha)} + \frac{(2\alpha-1)v_i(-6\alpha r+(6\alpha-3)rv_i+(4\alpha-2)v_i^2-3\alpha v_i}{6(\alpha-2\alpha v_i+v_i)^2} \text{ if } v_i \geq r \end{cases}$$

where $c_1 = \frac{r(4\alpha^2r^2-6\alpha^2r+6\alpha^2-4\alpha r^2+3\alpha r+r^2}{6}$. Observe that $\beta(r) = r$.

This bidding function is strictly increasing for $v_i \geq r$, which can be verified by looking at the first and second derivatives of the bidding function over $(r, 1)$.

$$\begin{aligned} \frac{d\beta}{dv_i}_{v_i \in (r,1)} &= \frac{1}{3(\alpha + v_i - 2\alpha v_i)^3}\Big((2\alpha-1)(r-v_i)(4\alpha^2r^2 + 4\alpha^2rv_i - 6\alpha^2r + 4\alpha^2v_i^2 - 6\alpha^2v_i \\ &\qquad + 3\alpha^2 - 4\alpha r^2 - 4\alpha r v_i + 3\alpha r - 4\alpha v_i^2 + 3\alpha v_i + r^2 + rv_i + v_i^2\Big) \\ \frac{d^2\beta}{dv_i^2} &= \frac{(1-2\alpha)(\alpha + r - 2\alpha r)^3}{(\alpha + v_i - 2\alpha v_i)^4} \end{aligned}$$

Since $\alpha < \frac{1}{2}$ and $r \in [0, 1]$, $1 - 2\alpha > 0$ and $\alpha + r - 2\alpha r > 0$. As a result, $\frac{d^2\beta}{dv_i^2} > 0$. Finally, observe that $\lim_{v_i \to r} \frac{d\beta}{dv_i} = 0$, so for all $v_i > r$, the $\frac{d\beta}{dv_i} > 0$.

To verify that β is a BNE, it is sufficient to show that v_i s.t. $v_i > r$ would find bidding $\beta(v_i)$ optimal among all bids $[r, \beta(1)]$. Given that all bidders are following the strategy β, bidding below r results in 0 payoff while bidding just above r results in strictly positive expected payoff. Furthermore, bidding strictly above $\beta(1)$ is strictly

dominated by bidding $\beta(1)$. By bidding $\beta(1)$, v_i guarantees he wins when the cartel does not malfunction, just as he does when he bids strictly above $\beta(1)$, but v_i pays strictly less in this event. v_i's bid does not impact the payments he receives from the other bidder when the cartel malfunctions, so bidding $\beta(1)$ dominates bidding strictly above.

Since β is strictly increasing, the bidder's problem is single-peaked in their bid b_i, and the FOC is sufficient to solve the problem. To see this, consider the bidder's problem:

$$\begin{aligned}\max_{b_i \in [r,\beta(1)]} \quad & (1-\alpha) Pr\big(b_i \geq \beta(v_j)\big)(v_i - b_i) + (1-\alpha) Pr(\beta(v_j) > b_i)\big(\mathbb{E}[\beta(v_j) - r | \beta(v_j) > b_i]\big) \\ & + \alpha Pr\big(b_i \geq \beta(v_j)\big)\Big(Pr\big(\beta(v_j) > r | b_i \geq \beta(v_j)\big)\mathbb{E}[\beta(v_j) - r | b_i \geq \beta(v_j) > r]\Big) \\ & + \alpha Pr\big(b_i < \beta(v_j)\big)(v_i - b_i)\end{aligned}$$

The derivative of the objective with respect to b_i is

$$(1-2\alpha)\left(\frac{1}{\beta'\left(\beta^{-1}\left(b_i\right)\right)}\left(v_i - 2b_i + r\right) - v_i\right) - \alpha$$

Since $\beta' > 0$ for $\tilde{v} \in (r, 1)$, the term in the large parentheses is positive below some threshold b_i^t and negative above that threshold. As a result, the bidder's objective is single-peaked. Furthermore, for bidders $v_i \in [r, 1]$, $b_i = \beta(v_i)$ solves the FOC, which sets the derivative equal to 0.

Observe that since β is a BNE, $\hat{\beta}$ given as follows

$$\hat{\beta}(v_i) = \begin{cases} v_i \text{ if } v_i < r \\ c_1(\alpha - 2\alpha v_i + v_i)^{-(2-4\alpha)/(1-2\alpha)} + \frac{(2\alpha-1)v_i(-6\alpha r + (6\alpha-3)r v_i + (4\alpha-2)v_i^2 - 3\alpha v_i}{6(\alpha - 2\alpha v_i + v_i)^2} \text{ if } v_i \geq r \end{cases}$$

is also a BNE.

This completes the demonstration that there exists a symmetric BNE of the malfunctioning McAfee-McMillan where bidders follow a strictly increasing bidding function.

B Theorem 1

B.1 Lemma 1

Define $\hat{A}_i = V_i \cup \hat{a}_i^0$ for each i. At each action profile $\hat{a} \in \hat{A}$, some bidders may be playing their opt-out actions. The following function identifies those bidders for every $\hat{a}$:

$$OO(\hat{a}) := \{i \in N : \hat{a}_i = \hat{a}_i^0\}$$

Define a similar (and notationally indistinguishable) function for action profiles $a \in A$. Define the allocation and transfer functions of the essentially direct mechanism, $(\hat{q}, \hat{t})$, as:

$$\hat{q}_j(\hat{a}) = \mathbb{E}_{(a_{k'} \sim \sigma_{k'}(v_{k'}))_{k' \notin OO(\hat{a})}} \left[q_j \left((a_k^0)_{k \in OO(\hat{a})}, (a_{k'})_{k' \notin OO(\hat{a})} \right) \right] \quad \forall \hat{a} \in \hat{A}$$
$$\hat{t}_j(\hat{a}) = \mathbb{E}_{(a_{k'} \sim \sigma_{k'}(v_{k'}))_{k' \notin OO(\hat{a})}} \left[t_j \left(\left(a_k^0\right)_{k \in OO(\hat{a})}, (a_{k'})_{k' \notin OO(\hat{a})} \right) \right] \quad \forall \hat{a} \in \hat{A}$$

For each action profile $\hat{a}$, $\left(\hat{q}_j(\hat{a}), \hat{t}_j(\hat{a})\right)$ is the expected allocation and transfer that would result for bidder j if all bidders $k' \notin OO(\hat{a})$ play according to their equilibrium strategies as $\hat{a}_{k'} = v_{k'}$ and all bidders $k \in OO(\hat{a})$ play their opt-out actions a_k^0 in the original mechanism.

The truthful strategy, σ^h, in $(\hat{A}, \hat{q}, \hat{t})$ where, for all v_i and i, v_i takes action v_i with probability 1, i.e., $\sigma_i^h(v_i|v_i) = 1$, is a BNE in the $(\hat{A}, \hat{q}, \hat{t})$ game because σ is a BNE in the (A, q, t) game. To see this, observe that any profitable deviation $\hat{\sigma}_i(v_i) \in \Delta\hat{A}_i$ for type v_i has a corresponding strategy in (A, q, t), $\sigma_i'(v_i)$:

$$\sigma_i'(a_i|v_i) = \int_{\tilde{v}_i} \hat{\sigma}_i(\tilde{v}_i|v_i) \cdot \sigma_i(a_i|\tilde{v}_i) d\tilde{v}_i \quad \forall a_i \in A_i \text{ s.t. } a_i \neq a_i^0$$
$$\sigma_i'(a_i^0|v_i) = \int_{\tilde{v}_i} \hat{\sigma}_i(\tilde{v}_i|v_i) \cdot \sigma_i(a_i^0|\tilde{v}_i)[a_i^0] d\tilde{v}_i + \hat{\sigma}_i(\hat{a}_i^0|v_i)$$

and $\sigma_i(a_i|v_i)$ denotes the probability of v_i playing a_i in BNE σ and $\hat{\sigma}_i(\tilde{v}_i|v_i)$ is the probability that v_i plays $\tilde{v}_i$. The payoff of deviating to $\sigma_i'(v_i)$ from $\sigma_i(v_i)$ in BNE σ of (A, q, t) is the same as the payoff of deviating to $\hat{\sigma}_i(v_i)$ from $\sigma_i^h(v_i)$ in BNE σ^h of $(\hat{A}, \hat{q}, \hat{t})$. Therefore, σ^h is a BNE of $(\hat{A}, \hat{q}, \hat{t})$ because σ is a BNE of (A, q, t). This is the standard Revelation Principle logic. Furthermore, the ex ante equilibrium payoff

for each bidder i under σ is equal to his payoff under σ^h.

$minRev(\hat{A}, \hat{q}, \hat{t}, \sigma^h) \geq minRev(A, q, t, \sigma)$ holds because every collusive scheme against $(\hat{A}, \hat{q}, \hat{t})$, $(\hat{\kappa}, \hat{q}^c, \hat{t}^c)$, can be mapped to a corresponding collusive scheme against (A, q, t), (κ, q^c, t^c), that results in the same revenue for the Principal. Since the set of revenue levels that may be induced by bidders colluding against $(\hat{A}, \hat{q}, \hat{t})$ is a subset of the set of revenue levels that may be induced by bidders colluding against (A, q, t), $minRev(\hat{A}, \hat{q}, \hat{t}, \sigma^h) \geq minRev(A, q, t, \sigma)$.

To complete the argument, fix a collusive scheme $(\hat{\kappa}, \hat{q}^c, \hat{t}^c)$. Define the corresponding (κ, q^c, t^c) as follows:

$$\kappa(a|v) = \int_{\hat{a} \in \hat{A}} \mathbb{I}\{OO(a) = OO(\hat{a})\} \prod_{i \notin OO(\hat{a})} \sigma_i(a_i|\hat{a}_i) \hat{\kappa}(d\hat{a}|v) \quad \forall v, \forall a \in A$$

$$q^c(v) = \hat{q}^c(v) \quad \forall v \qquad t^c(v) = \hat{t}^c(v) \quad \forall v$$

B.2 Consolidation of C-EAIR: Lemma 6

The bidder specific C-EAIR constraints can be consolidated into a joint participation constraint, C-JEAIR.

Lemma 6. *Given the principal sets essentially direct mechanism* (q, t)*,*

$$\inf_{\kappa, q^c, t^c} \mathbb{E}_v[\sum_{a \in A} \kappa(a|v) \sum_i t_i(a)] \qquad \text{(J)}$$
$$s.t.\ \ C\text{-}F,\ \ C\text{-}EABB,\ \ C\text{-}JEAIR$$

has the same value as

$$\inf_{\kappa, q^c, t^c} \mathbb{E}_v[\sum_{a \in A} \kappa(a|v) \sum_i t_i(a)] \qquad \text{(I)}$$
$$s.t.\ \ C\text{-}F,\ \ C\text{-}EABB,\ \ C\text{-}EAIR$$

Proof. Since C-JEAIR is a necessary condition implied by C-EAIR, every collusive outcome (κ, q^c, t^c) feasible for I-R also is feasible for J-R. As a result, it is clear that J-R $\leq$ I-R. To see that J-R $\geq$ I-R, for every collusive outcome (κ, q^c, t^c) feasible for J-R, we construct a collusive outcome $(\kappa, q^c, \hat{t}^c)$ that is feasible for I-R.

Define U_i^c to be the ex ante surplus i gets from (κ, q^c, t^c), i.e., $U_i^c := \mathbb{E}_v[q_i^c(v)v_i -$

$t_i^c(v) - \sum_{a \in A} \kappa(a|v) t_i(a)]$. By C-JEAIR, $\sum_i U_i^c \geq \sum_i U_i^n$, so there exist lump sum transfers $\{T_i\}_i$ such that $U_i^c + T_i \geq U_i^n$ and $\sum_i T_i = 0$. Define $\hat{t}_i^c(v) = t_i^c(v) + T_i$. $(\kappa, q^c, \hat{t}^c)$ is feasible for I-R. □

B.3 Lemma 2

I start with (2), convexity. First, note that $\hat{P}_{(q,t)}$ is convex. Let $q, q' \in [0, \bar{q}]$, and let $\alpha \in [0, 1]$. For all $\varepsilon > 0$, $\exists (q, T_\varepsilon)$ and $(q', T'_\varepsilon) \in \hat{P}_{(q,t)}$ where $|T_\varepsilon - c(q)| < \varepsilon$ and $|T'_\varepsilon - c(q')| < \varepsilon$. The convex combination of (q, T_ε) and (q', T'_ε) is in $\hat{P}_{(q,t)}$ as well. As a result, $c(\alpha q + (1-\alpha) q') \leq \alpha c(q) + (1-\alpha) c(q') + \varepsilon$. Since ε is arbitrary, we have that $c(\alpha q + (1-\alpha) q') \leq \alpha c(q) + (1-\alpha) c(q')$.

For (3), that c is increasing, fix two points $(Q, c(Q))$, $(Q', c(Q'))$ with $Q' > Q$. Suppose for a contradiction that $c(Q') < c(Q)$. Let ε be such that $\varepsilon \in (0, |c(Q') - c(Q)|)$. There exists $(Q', T) \in \hat{P}_{(q,t)}$ such that $|T - c(Q')| < \varepsilon$ and so $T < c(Q)$. Since colluding bidders can dispose of the good, $(\tilde{Q}, T) \in \hat{P}_{(q,t)}$ for all $\tilde{Q} \leq Q'$. Specifically, $(Q, T) \in \hat{P}_{(q,t)}$. This implies $c(Q) \leq T$. This is a contradiction.

For (1), that c is continuous on $[0, \bar{q})$, we get that c is continuous on $(0, \bar{q})$ from (2). We show that $c(0) = \lim_{q \to 0^-} c(q)$. Because c is increasing, $c(0) \leq c_-(0)$. Fix $\varepsilon > 0$ and small $\delta > 0$. By definition, there exists $(0, T_\delta) \in \hat{P}_{(q,t)}$ s.t. $|T_\delta - c(0)| < \delta$, and similarly, there exists $(\varepsilon, T'_\delta) \in \hat{P}_{(q,t)}$ such that $|T'_\delta - c(\varepsilon)| < \delta$. Since $\hat{P}_{(q,t)}$ is convex, the following inequality holds for all $\alpha \in [0, 1]$:

$$c(\alpha\varepsilon + (1-\alpha)0) < \alpha c(\varepsilon) + (1-\alpha)c(0) + \delta$$

Taking the limit as $\alpha \to 0$ and $\delta \to 0$, $\lim_{q \to 0^-} c(q) \leq c(0)$. So, $c(0) = \lim_{q \to 0^-} c(q)$.

Finally, for (4), $c(0) \leq 0$ holds since every bidder i playing a_i^0 results in no one getting the good and no one paying or being paid. As a result, $(0, 0) \in \hat{P}_{(q,t)}$, and $c(0) \leq 0$. □

B.4 Lemma 3

Suppose $\text{minRev}(\hat{q}, \hat{t}) < \text{minRev}(q, t)$. Then, there exists a collusive outcome (κ, q^c, t^c) that is feasible against $(\hat{q}, \hat{t})$ that results in revenue $\mathbb{E}_v \left[\sum_{a \in A} \kappa(a|v) \left(\sum_i \hat{t}_i(a)\right)\right] < \text{minRev}(q, t)$. We will show that this leads to a contradiction by constructing a collusive outcome feasible for (q, t) with similar revenue.

Throughout the argument, we will work with the total transfers that colluders pay at each v, i.e., the transfers inclusive of side payments to other bidders and payments to the principal. Fixing a collusive arrangement (κ, q^c, t^c) against mechanism $(\hat{q}, \hat{t})$, the total amount i pays to the principal and the collusive arrangement when the true profile of values is v is denoted by $\hat{t}_i^{total}(v) := \sum_{a \in A} \kappa(a|v) \left(\sum_i \hat{t}_i(a)\right) + t_i^c(v)$. Specifying κ and $\hat{t}^{total}$ is equivalent to specifying the side payments t^c. Given κ and $\hat{t}^{total}$, $t_i^c(v)$ is implicitly defined as $\hat{t}_i^{total}(v) - \sum_{a \in A} \kappa(a|v) \sum_i \hat{t}_i(a)$.

The EABB constraint on collusion can be expressed in terms of $\hat{t}^{total}$ as $\mathbb{E}_v \left[\sum_i \hat{t}_i^{total}(v)\right] = \mathbb{E}_v \left[\sum_{a \in A} \kappa(a|v) \sum_i \hat{t}_i(a)\right]$.

To start, we "scale down" (κ, q^c, t^c) by mixing executing (κ, q^c, t^c) with coordinating all bidders to play their opt-out actions. Fix $\alpha \in (0, 1)$. We define, for all v,

$$q_i^{scale}(v) = \alpha q_i^c(v) \qquad t_i^{scale}(v) = \alpha \hat{t}_i^{total}(v)$$
$$\kappa^{scale}(a|v) = \alpha \kappa(a|v) \quad \text{if } a \neq a^0 \qquad \kappa^{scale}(a|v) = \alpha \kappa(a|v) + 1 - \alpha \quad \text{if } a = a^0$$

where $a^0 = (a_1^0, ..., a_n^0)$. $(\kappa^{scale}, q^{scale}, t^{scale})$ is a collusive outcome that induces aggregate probabilities of acquiring the good, Q, that are strictly less than $\hat{\bar{q}}$, the maximum of the domain of the cost function induced by $(\hat{q}, \hat{t})$, $\hat{c}$. This maneuver is necessary because it is possible there is no mixture over action profiles in (q, t) that would exactly produce an aggregate probability of $\hat{\bar{q}}$; instead, it may only be possible to get arbitrarily close to the aggregate probability of $\hat{\bar{q}}$ from below.

$(\kappa^{scale}, q^{scale}, t^{scale})$ satisfies feasibility:

$$\sum_i q_i^{scale}(v) = \sum_i \alpha q_i^c(v) \leq \alpha \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \left(\sum_i \hat{q}_i(a)\right) = \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa^{scale}(a|v) \left(\sum_i q_i^{scale}(a)\right)$$

Similarly, because the original collusive scheme (κ, q^c, t^c) is EABB, the scaled collusive scheme is also EABB. The scaled collusive scheme achieves joint surplus $\alpha \sum_i \hat{U}_i^n$. For α sufficiently close to 1, we can get $\sum_i U_i^n < \alpha \sum_i \hat{U}_i^n < \sum_i \hat{U}_i^n$.

Now, let's define a corresponding collusive outcome feasible for (q, t), $(\tilde{\kappa}, \tilde{q}^c, \tilde{t}^c)$. For each $v \in V$, define $Q(v) := \alpha \sum_a \kappa(a|v) \left(\sum_i q_i(a)\right)$. $Q(v)$ is the expected probability with which colluding bidders acquire the good given their values are v and they collude against $(\hat{q}, \hat{t})$ to produce $(\kappa^{scale}, q^{scale}, t^{scale})$.

Fix $\delta > 0$. For each v, it is possible to find a $(Q(v), T) \in \hat{P}_{(q,t)}$ such that $T <$

$c\left(Q\left(v\right)\right)+\delta$. Since $(Q(v),T)\in\hat{P}_{(q,t)}$, define $\tilde{\kappa}(\cdot|v)\in\Delta A$ so that $\mathbb{E}_{a\sim\tilde{\kappa}(\cdot|v)}\left[\sum_i q_i(a)\right]\geq Q\left(v\right)$ and $\mathbb{E}_{a\sim\tilde{\kappa}(\cdot|v)}\left[\sum_i t_i(a)\right]=T$.

Let $\tilde{q}_i^c(v)=q_i^{scale}(v)$ for all i and v.

To construct the transfers, define the following constant, which captures the difference in total cost to the cartel of running $(\kappa^{scale},q^{scale},t^{scale})$ against $(\hat{q},\hat{t})$ and running EABB $(\tilde{\kappa},\tilde{q}^c,\tilde{t}^c)$ against (q,t).

$$\begin{aligned}D&:=\mathbb{E}_v\left[\sum_i t_i^{scale}(v)\right]-\mathbb{E}_v\left[\sum_{a\in\text{supp}(\tilde{\kappa}(\cdot|v))}\tilde{\kappa}(a|v)\left(\sum_i t_i(a)\right)\right]\\&\geq\mathbb{E}_v\left[\hat{c}\left(Q(v)\right)\right]-\mathbb{E}_v\left[c\left(Q(v)\right)\right]-\delta\\&\geq-\delta\end{aligned}$$

D is not necessarily positive.

The transfers $\tilde{t}^c(\cdot)$ split D across bidders as lump-sum transfers:

$$\tilde{t}_i^c(v):=t_i^{scale}(v)-\frac{D}{N}$$

Clearly, $(\tilde{\kappa},\tilde{q}^c,\tilde{t}^c)$ inherits C-F and C-EABB from (κ,q^c,t^c). To verify C-EABB, let U_i^c be the ex ante surplus bidder i gets from (κ,q^c,t^c). Similarly, let $\tilde{U}_i^c$ be the ex ante surplus bidder i gets from $(\tilde{\kappa},\tilde{q}^c,\tilde{t}^c)$.

$$\sum_i\tilde{U}_i^c=\sum_i U_i^c+D\geq\sum_i U_i^c-\delta\geq\alpha\sum_i\hat{U}_i^n-\delta$$

As a result, for δ small enough, $\sum_i\tilde{U}_i^c\geq\sum_i U_i^n$.

Taking the limit as δ approaches 0, we get

$$\mathbb{E}_v\left[c\left(Q\left(v\right)\right)\right]\geq\text{minRev}(q,t)$$

Fixing α close enough to 1, we have that:

$$\begin{aligned}\text{minRev}(q,t) &> \mathbb{E}\left[\sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)\left(\sum_i \hat{t}_i(a)\right)\right]\\ &\geq \mathbb{E}\left[\sum_{a\in\text{supp}(\kappa^{scale}(\cdot|v))}\kappa^{scale}(a|v)\left(\sum_i \hat{t}_i(a)\right)\right]\\ &\geq \mathbb{E}\left[\hat{c}\left(Q\left(v\right)\right)\right] \geq \mathbb{E}_v\left[c\left(Q\left(v\right)\right)\right]\end{aligned}$$

Contradiction. □

B.5 Inner Problem: Surplus Maximization

B.5.1 Relaxed Problem

To solve c-Surplus Max, I solve first a relaxed problem:

$$\begin{aligned}\max_{q:V\to[0,1]^N, t:V\to\mathbb{R}} \quad & \mathbb{E}_v[\sum_i q_i(v)v_i - t_i(v)] \\ \text{s.t.} \quad & \sum_i q_i(v) \leq \bar{q} \quad \forall v \\ & \sum_i t_i(v) \geq c(\sum_i q_i(v)) \quad \forall v\end{aligned} \tag{relax-cSM}$$

relax-cSM relaxes c-Surplus Max by removing the IC and IR constraints.

It is possible to solve relax-cSM by solving the pointwise problems:

$$\begin{aligned}\max_{q\in\mathbb{R}^N_+, t\in\mathbb{R}^N} \quad & \sum_i q_i \cdot v_i - \sum_i t_i \\ \text{s.t.} \quad & \sum_i q_i \leq \bar{q} \\ & \sum_i t_i \geq c(\sum_i q_i)\end{aligned} \tag{v-Problem}$$

It is clear that the $(q^*(v), t^*(v))$ that solves v-Problem for a fixed v makes the second constraint bind so that $\sum_i t^*(v) = c\left(\sum_i q_i^*(v)\right)$. Furthermore, to maximize surplus, the good should go to the bidder with the highest value, so given that i is the bidder with the unique highest realized value, $q_i^*(v) = \sum_j q_j^*(v)$. After imposing

these two observations, solving the pointwise problem is tantamount to choosing the total probability that the good is given to the bidders, $\sum_j q_j^* =: Q$, for each v:

$$\max_{Q \in [0,\bar{q}]} \quad Q \cdot \max_i v_i - c(Q) \tag{reduced-vProb}$$

The objective is continuous since c-Surplus Max is formulated only for continuous c, so the problem is well-defined for all $\max_i v_i \in [0,1]$. Since the objective is a concave function, we have that conditions for local optimality are sufficient for global optimality. Let $Q^*(\cdot)$ be a selection of maximizers for reduced-vProb; it is a function of $\max_i v_i$.

Lemma 7. *Suppose cost function $c : [0,\bar{q}] \to \mathbb{R}$ is continuous at $\bar{q}$.*

1. *Any selection of maximizers $Q^*(\cdot)$ is non-decreasing.*
2. *The value of the problem is continuous in $\max_i v_i$.*
3. *$c\left(Q^*\left(\max_i v_i\right)\right)$ is increasing.*
4. *The selection of maximizers $Q^*\left(\max_i v_i\right) := \max argmax_{Q \in [0,\bar{q}]} Q \max_i v_i - c(Q)$ is well-defined and right continuous on $(0,1)$.*

Proof. For (1), fix v and $\tilde{v}$ such that $\max_i v_i > \max_i \tilde{v}_i$. Let $Q^*(\cdot)$ be any selection of maximizers. Suppose $Q^*(\max_i v_i) < Q^*(\max_i \tilde{v}_i)$. Since c is convex, it is differentiable almost everywhere. Fix $\tilde{Q} \in \left(Q^*\left(\max_i v_i\right), Q^*\left(\max_i \tilde{v}_i\right)\right)$ such that c is differentiable at $\tilde{Q}$. We have that

$$c'_+\left(Q^*\left(\max_i v_i\right)\right) \leq c'_+\left(\tilde{Q}\right) = c'\left(\tilde{Q}\right) = c'_-\left(\tilde{Q}\right) \leq c'_-\left(Q^*\left(\max_i \tilde{v}_i\right)\right)$$

Because $Q^*(\max_i v_i)$ solves reduced-vProb for $\max_i v_i$, we have that $c'_-\left(Q^*\left(\max_i v_i\right)\right) \leq \max_i v_i \leq c'_+\left(Q^*\left(\max_i v_i\right)\right)$. Analogous inequalities hold for $\max_i \tilde{v}_i$. By these optimality conditions, we have then that $\max_i v_i \leq \max_i \tilde{v}_i$. Contradiction. So $Q^*(\max_i v_i) \geq Q^*(\max_i \tilde{v}_i)$.

Next, Berge's theorem implies the correspondence of maximizers is compact-valued and upper hemicontinuous as well as (2).

For (3), I prove this claim in a way that does not rely on $c(\cdot)$ being increasing. This will be useful in the extensions section. First, observe that if $\bar{q} = 0$, then $c\left(Q^*((v))\right)$ is constant and so is increasing.

Suppose $\bar{q} > 0$ now. Since $c'_+(0)$ is well-defined, we have that $0 \leq c'_+(Q^*(0))$. Since c is convex, it is differentiable a.e.

$$c(Q) = \int_{Q^*(0)}^{Q} c'(\tilde{Q}) d\tilde{Q} \quad \forall Q \geq Q^*(0)$$

$c'(\tilde{Q}) \geq 0$ for all $\tilde{Q} \geq Q^*(0)$ since $c'(\tilde{Q}) \geq c'_+(Q^*(0)) \geq 0$. Since $Q^*(\cdot)$ is increasing, $c(Q^*(\cdot))$ is increasing.

For (4), observe that the set of maximizers for any instance of the problem is compact, so the selection given in the statement is well-defined. We show that the $Q^*(\cdot)$ specified is also right-continuous on $(0,1)$. Fix any decreasing sequence of real numbers $\{v_n\}_n$ with $v_n \in [0,1]$ for all n. Let v denote its limit point. $\{Q^*(v_n)\}_n$ is decreasing and is bounded below by $Q^*(v)$ due to (1), so it is convergent and $Q^*(v) \leq \lim_{n\to\infty} Q^*(v_n)$. The correspondence of maximizers is upper hemicontinuous, so $Q^*(v) \geq \lim_{n\to\infty} Q^*(v_n)$. This implies $Q^*(v) = \lim_{n\to\infty} Q^*(v_n)$. □

Now, we fix a particular selection of maximizers of reduced-vProb, $Q^*(\cdot)$, to construct a solution to relax-cSM, (q^*, t^*). For each $\max_i v_i \in [0,1]$, define $Q^*(\max_i v_i) := \max \operatorname{argmax}_{Q\in[0,\bar{q}]} Q \max_i v_i - c(Q)$. Define the following allocation function for all i and v:

$$q_i^*(v) = \begin{cases} \frac{1}{|\{v_j : v_j = \max_k v_k\}|} Q^*(\max_j v_j) & \text{if } v_i = \max_j v_j \\ 0 & \text{otherwise} \end{cases}$$

Given any transfer function t^* such that $\sum_i t_i^*(v) = c(Q^*(\max_j v_j))$ for all v, (q^*, t^*) solves relax-cSM.

B.5.2 Solution to c-Surplus Max

A particular solution of relax-cSM also solves c-Surplus Max. Not every ex post transfer function t^* will satisfies the IC and IR constraints of c-Surplus Max, but it is possible to construct a particular transfer function that does.

Proposition 2. *Fix a continuous cost function $c : [0,\bar{q}] \to \mathbb{R}$. There exists t^* such that (q^*, t^*) solves c-Surplus Max and relax-cSM.*

Proof. $q^*(\cdot)$ induces weakly increasing interim allocation functions. As a bidder's value increases, (1) the probability that they hold the highest realized value increases,

and (2) the total probability at which the good will be given to the bidder increases. So, if we can provide a transfer function that satisfies the envelope theorem derived formula for interim transfers and satisfies the ex post restrictions from relax-cSM on total cost paid a.e., we have demonstrated that c-Surplus Max = relax-cSM.

First, we demonstrate that if we could find a transfer function t^* so that (q^*, t^*) is IC, we could also satisfy IR. To do this, notice that given t^* satisfies IC, we can calculate the total surplus given to each bidder's lowest types via the payoff formula:

$$\sum_i U_i(0) = \mathbb{E}_v\left[\sum_i q_i^*(v)v_i - \sum_i t_i^*(v)\right] - \sum_i \mathbb{E}_{v_i}\left[\int_0^{v_i} \bar{q}_i^*(\tilde{v})d\tilde{v}\right]$$

where $\sum_i t_i^*(v) = c(\sum_i q_i^*(v))$ for all v.

Let $\tilde{q}^* : [0,1] \to [0,1]$ be the mapping from the maximum realized value to the optimal probability at which to acquire the good. Lemma 7 establishes that $\tilde{q}^*$ is increasing and is right-continuous on $(0,1)$, so it is differentiable almost everywhere with jump discontinuities only at countable locations. Let V_d be the set of discontinuities in $\tilde{q}^*$.

Since $\tilde{q}^*$ is increasing, we can identify a threshold v_t at which q^* starts being strictly positive. Formally, define:

$$v_t := \sup\{v \in [0,1] : Q^*(v) = 0\}$$

For $v > v_t$, we can define $\tilde{c}(v) := c(\tilde{q}^*(v))/\tilde{q}^*(v)$, the optimal cost per unit of probability that colluders will pay in total when the maximum value is v. Note that since c is continuous on $[0, \bar{q}]$, $\tilde{c}$ is discontinuous at $v \in [0,1]$ if and only if $\tilde{q}$ is discontinuous at v.

Consider $\sum_i U_i(0)$:

$$\begin{aligned}\sum_i U_i(0) &= \sum_i \int_{v_t}^{1} \Big[\bar{q}_i^*(v)(v_i - \tilde{c}(v_i)) - \int_0^{v_i} \bar{q}_i^*(\tilde{v})d\tilde{v}\Big] dF_i(v_i) + \int_0^{v_t} -c(0)dF_i(v_i) \\ &= \sum_i \underbrace{\int_{v_t}^{1} \Big[(\prod_{j\neq i} F_j(v_i)\tilde{q}_i^*(v_i))(v_i - \tilde{c}(v_i)) - \int_{v_t}^{v_i} \prod_{j\neq i} F_j(\tilde{v})\tilde{q}_i^*(\tilde{v})d\tilde{v}\Big] dF_i(v_i)}_{=:S_i} \\ &\quad + \underbrace{\int_0^{v_t} -c(0)dF_i(v_i)}_{\geq 0}\end{aligned}$$

The second term is weakly positive since $c(0) \leq 0$, so we concentrate on assigning a sign to S_i.

Integrate the last term by parts:

$$\begin{aligned}\int_{v_t}^{v_i} \prod_{j\neq i} F_j(v_i)\tilde{q}_i^*(v_i)d\tilde{v} = \prod_{j\neq i} F_j(v_i)\tilde{q}^*(v_i)v_i - \sum_{v\in V_d\cap(v_t,v_i)} [\prod_{j\neq i} F_j(s)\tilde{q}^*(s)s]_{v-}^{v+} \\ - \int_{v_t}^{v_i} \tilde{v}d\big(\prod_{j\neq i} F_j(\tilde{v})\tilde{q}^*(\tilde{v})\big)\end{aligned}$$

$$\begin{aligned}S_i &= \int_{v_t}^{1} \Big[(\prod_{j\neq i} F_j(v_i)\tilde{q}_i^*(v_i))(v_i - \tilde{c}(v_i)) - \prod_{j\neq i} F_j(v_i)\tilde{q}^*(v_i)v_i + \sum_{v\in V_d\cap(v_t,v_i)} [\prod_{j\neq i} F_j(s)\tilde{q}^*(s)s]_{v-}^{v+} \\ &\quad + \int_{v_t}^{v_i} \tilde{v}d\big(\prod_{j\neq i} F_j(\tilde{v})\tilde{q}^*(\tilde{v})\big)\Big] dF_i(v_i) \\ &= \int_{v_t}^{1} \Big[\int_{v_t}^{v_i} \tilde{v} - \tilde{c}(v_i)d\big(\prod_{j\neq i} F_j(\tilde{v})\tilde{q}^*(\tilde{v})\big) + \sum_{v\in V_d\cap(v_t,v_i)} [\prod_{j\neq i} F_j(s)\tilde{q}^*(s)(s - \tilde{c}(v_i))]_{v-}^{v+}\Big] dF_i(v_i) \\ &= \int_{v_t}^{1} (1 - F_i(v_i))(v_i - \tilde{c}(v_i))\left(\frac{d\big(\prod_{j\neq i} F_j(\tilde{v})\tilde{q}^*(\tilde{v})\big)}{d\tilde{v}}\right)\Big|_{\tilde{v}=v_i} dv_i \\ &\quad + \sum_{v\in V_d} (1 - F_i(v))[\prod_{j\neq i} F_j(s)\tilde{q}^*(v)(v - \tilde{c}(v))]_{v-}^{v+} \geq 0\end{aligned}$$

where $\frac{d\left(\prod_{j\neq i} F_j(\tilde{v})\tilde{q}^*(\tilde{v})\right)}{d\tilde{v}}$ is defined almost everywhere and weakly positive because $\prod_{j\neq i} F_j(\tilde{v})\tilde{q}^*(\tilde{v})$ is increasing. Furthermore, $v - \tilde{c}(v) \geq 0$ for all v since surplus maximization implies that $\tilde{q}(v)v - c(\tilde{q}(v)) \geq 0$.

If $\mathbb{E}_v[c(\sum_i q_i^*(v))] \neq 0$, we can take any division of the surplus to the lowest types such that for all i, $U_i(0) \geq 0$. To define t^*, define constants κ_i so that

$$\kappa_i = \frac{\mathbb{E}_{v_i}[\bar{t}_i^*(v_i)]}{\mathbb{E}_v[c(\sum_i q_i(v))]}$$

where $\bar{t}_i^*(v_i)$ is given by the payoff formula:

$$\bar{t}_i^*(v_i) = \bar{q}^*(v_i)v_i - U_i(0) - \int_0^{v_i} \bar{q}^*(\tilde{v})d\tilde{v}$$

The following transfers satisfy the constraints $\sum_i t_i^*(v) = c(\sum_i q_i^*(v))$ for all v and match the interim transfers required for IC:

$$\begin{aligned} t_i(v) = \kappa_i c(\sum_k q_k^*(v)) + \bar{t}_i^*(v_i) - \kappa_i \mathbb{E}_{v_{-i}}[c(\sum_k q_k^*(v))] \\ - \frac{1}{N-1}\big(\sum_{j\neq i} \bar{t}_j^*(v_j) - \kappa_j \mathbb{E}_{v_{-j}}[c(\sum_k q_k^*(v))]\big) \end{aligned}$$

If $\mathbb{E}_v[f(\sum_i q_i^*(v))] = 0$, consider the division of surplus between the lowest types induced by restricting each bidder's expected transfers to be 0, i.e.,

$$\mathbb{E}_{v_i}[\bar{t}_i^*(v_i)] = 0 \quad \forall i$$

This restriction pins down the surplus to the lowest type for each bidder through the payoff formula:

$$U_i(0) = \mathbb{E}_{v_i}[\bar{q}_i^*(v_i)v_i - \int_0^{v_i} \bar{q}_i^*(\tilde{v})d\tilde{v}] \geq 0$$

where the inequality holds because $\bar{q}^*(\cdot)$ is weakly increasing in v_i. The following transfer rule satisfies the desired interim transfers and $\sum_i t_i^*(v) = c(\sum_k q_k^*(v))$ for all v:

$$\begin{aligned} t_i^*(v) = \frac{1}{N} c(\sum_k q_k^*(v)) + \bar{t}_i^*(v_i) - \frac{1}{N}\mathbb{E}_{v_{-i}}[c(\sum_k q_k^*(v))] \\ - \frac{1}{N-1}\bigg(\sum_{j\neq i} \bar{t}_j^*(v_j) - \frac{1}{N}\mathbb{E}_{v_{-j}} c(\sum_k q_k^*(v))]\bigg) \end{aligned}$$

□

I have shown that there is an IC, IR mechanism with (1) ex post allocation function q^* and (2) ex post transfer function t^* s.t. $\sum_i t_i^*(v) = c\left(\sum_i q_i^*(v)\right)$ for all v. As a result, relax-cSM has the same value as c-Surplus Max. Furthermore, *any* such IC, IR mechanism would solve c-Surplus Max; t^* does not have to be the construction given in the proof.

B.5.3 Making the Solution Feasible

Still fixing a well-behaved cost function c, let (q^*, t^*) solve c-Surplus Max.

First, this direct mechanism can easily be made into an essentially direct mechanism, $(\tilde{q}^*, \tilde{t}^*)$, by adding an opt-out action for each bidder, a_i^0. For action profiles where $a = v \in V$, $\tilde{q}^*(v) = q^*(v)$, and $\tilde{t}^*(v) = t^*(v)$. For action profiles a where at least one bidder plays their opt-out action, then define $\tilde{q}_j^*(a) = \tilde{t}_j^*(a) = 0$ for all bidders j. Since playing the opt-out action guarantees that a bidder gets 0 payoff, deviating from the truth-telling strategy is not profitable since the direct mechanism is IR.

From here, I abuse notation by denoting the essentially direct mechanism as defined above by (q^*, t^*). Furthermore, I say an essentially direct mechanism (q^*, t^*) *solves c-Surplus Max* if its truth-telling equilibrium solves c-Surplus Max and its allocations and transfers at action profiles where at least one bidder plays their opt-out action are as defined above.

B.5.4 Worst-Case Revenue of the Solution

The worst-case revenue of the solution (q^*, t^*) is the value of the BNE:

Proposition 3. *Let $c : [0, \bar{q}] \to \mathbb{R}$ be a continuous cost function. Let (q^*, t^*) be the essentially direct mechanism defined in B.5.3 whose truth-telling equilibrium solves c-Surplus Max.*

$$minRev(q^*, t^*) = \mathbb{E}_v\left[\sum_i t_i^*(v)\right] = \mathbb{E}_v\left[c\left(\sum_i q_i^*(v)\right)\right]$$

Proof. I show that $\text{minRev}(q^*, t^*) = \mathbb{E}_v\left[c\left(\sum_i q_i^*(v)\right)\right]$ by showing that every collusive arrangement against (q^*, t^*) must result in revenue $\mathbb{E}_v\left[c\left(\sum_i q_i^*(v)\right)\right]$.

At a high level, the argument proceeds by showing that the surplus generated by the BNE of (q^*, t^*), $\sum_i U_i^n$, is so high that any collusive outcome must essentially replicate the outcome of the BNE in order to produce at least $\sum_i U_i^n$ in surplus. In general, at almost all v, any collusive outcome must also allocate the good efficiently, acquire the good with total probability $\sum_i q_i^*(v)$, and so must pay the principal at least $c\left(\sum_i q_i^*(v)\right) = \sum_i t_i^*(v)$.

To formalize this, I first show that with probability 1, there is a unique solution to reduced-vProb. Observe that with probability 1, there is a single bidder with the maximum realized value. Next, let

$$V_m := \{v \in [0,1] : \textit{reduced} - \textit{vProb} \text{ has at least two solutions}\}$$

With probability 1, the solution to reduced-vProb is unique:

Lemma 8. *V_m has measure 0.*

Proof. Fix a $\max_i v_i \in V_m$. Let Q and Q' be two distinct solutions to reduced-vProb with $Q < Q'$. The concavity of the objective implies that any convex combination of Q and Q' also solves the problem, so $[Q, Q']$ is a subset of the solutions to reduced-vProb. The local optimality conditions imply that $c'(\tilde{Q}) = \max_i v_i$ for all $\tilde{Q} \in (Q, Q')$. Denote the interval (Q, Q') by $I(\max_i v_i)$. Every $\max_i v_i \in V_m$ can be mapped in this way to a non-degenerate $I(\max_i v_i)$. These intervals are pairwise disjoint, i.e. $I(\max_i v_i) \cap I(\max_i \tilde{v}_i) = \emptyset$, since $Q \in I(\max_i v_i) \cap I(\max_i \tilde{v}_i)$ implies $\max_i v_i = c'(Q) = \max_i \tilde{v}_i$. We can map each $\max_i v_i \in V_{max,m}$ to a distinct rational $q \in I(\max_i v_i) \cap \mathbb{Q}$. Thus, $V_{max,m}$ is countable and so has measure 0. □

Fix a collusive scheme possible against (q^*, t^*), (κ, q^c, t^c). The special feature of (q^*, t^*) is that it maximizes joint total surplus for each v given a cost function $c(\cdot)$:

$$\begin{aligned}\sum_i \left(q_i^*(v)v_i - t_i^*(v)\right) &= Q^*(\max_i v_i)\max_i v_i - c(Q^*(\max_i v_i)) \\ &\geq \sum_i q_i^c(v)v_i - \sum_a \kappa(a|v)\sum_i t_i^*(a)\end{aligned}$$

In the second line, I observe that the collusive scheme's effective contribution to joint surplus is weakly below the joint surplus generated by the BNE at each v. Since any collusive scheme (κ, q^c, t^c) against (q^*, t^*) must raise at least $\mathbb{E}_v\left[\sum_i q_i^*(v)v_i - \sum_i t_i^*(v)\right]$

in surplus for the bidders, the above inequality must bind with probability 1. With probability 1, there is a unique bidder with the highest realized value, and by Lemma 8, we have that $\sum_i q_i^c(v) = Q^*(\max_i v_i)$ with probability 1. Let

$$A := \{v \in V : \sum_i (q_i^*(v)v_i - t_i^*(v)) = \sum_i q_i^c(v)v_i - \sum_a \kappa(a|v) \sum_i t_i^*(a)$$
$$\text{and reduced-vProb has a unique solution}\}$$

Note that A occurs with probability 1. For each $v \in A$, we have that $\sum_a \kappa(a|v) \sum_i t_i^*(a) = c(Q^*(\max_i v_i))$. So we have that $\mathbb{E}_v \left[\sum_a \kappa(a|v) \sum_i t_i^*(a)\right] = \mathbb{E}_v \left[c\left(\sum_i Q^*(\max_i v_i)\right)\right]$. □

B.5.5 WLOO Reduction of Mechanisms

To show that it is possible to restrict attention to essentially direct mechanisms that solve c-Surplus Max for a well-behaved cost function, we need to be able to map each feasible mechanism (q, t) to a well-behaved cost function c and show that (q^*, t^*) that solves c-Surplus Max for c is weakly preferred to (q, t).

The following lemma maps $\hat{c}$, the cost function induced by an arbitrary feasible (q, t), to a well-behaved cost function:

Lemma 9. *Let $\hat{c}$ be the cost function induced by (q, t). Let $\bar{q}$ be the supremum of the domain of $\hat{c}$. Define $c : [0, \bar{q}] \to R$ so that*

$$c(Q) = \begin{cases} \hat{c}(Q) & \text{if } Q \in [0, \bar{q}) \\ \lim_{Q \to \bar{q}^-} \hat{c}(Q) & \text{if } Q = \bar{q} \end{cases}$$

Then, c is also a cost function.

Proof. To verify that c is a cost function, I check the properties given in Lemma 2. Observe that (1) and (4) are inherited from $\hat{c}$. To check (2), I check that the following inequality holds for all $\alpha \in (0, 1)$, $Q \in [0, \bar{q})$:

$$\alpha c(\bar{q}) + (1 - \alpha)c(Q) \geq c\left(\alpha\bar{q} + (1 - \alpha)\, Q\right)$$

Note that $\alpha\bar{q} + (1-\alpha)Q \in (0, \bar{q})$. For small $\varepsilon > 0$, the following inequality holds

$$\alpha\hat{c}(\bar{q} - \varepsilon) + (1-\alpha)\hat{c}(Q) \geq \alpha\hat{c}\left(\alpha(\bar{q} - \varepsilon) + (1-\alpha)Q\right)$$

Taking the limit as $\varepsilon \to 0$ and using the continuity of $\hat{c}$ and the definition of c, we have that

$$\alpha c(\bar{q}) + (1-\alpha)\hat{c}(Q) \geq \alpha\hat{c}\left(\alpha\bar{q} + (1-\alpha)Q\right)$$

It is also clear that $c(\bar{q}) \geq c(Q)$ for all $Q < \bar{q}$ since $\hat{c}(\cdot)$ is increasing. □

The next proposition allows us to focus on mechanisms that solve c-Surplus Max for a continuous c without loss of optimality:

Proposition 4. *Suppose (q,t) induces cost function $\hat{c}$ and $\bar{q} = \sup_{a\in A}\sum_i q_i(a)$. Define $c : [0,\bar{q}] \to \mathbb{R}$ so that*

$$c(Q) = \begin{cases} \hat{c}(Q) & \text{if } Q \in [0,\bar{q}) \\ \lim_{Q\to\bar{q}^-}\hat{c}(Q) & \text{if } Q = \bar{q} \end{cases}$$

Let (q^*,t^*) solve c-Surplus Max for continuous cost function c. Then, $minRev(q,t) \leq minRev(q^*,t^*)$.

Proof. From Proposition 3, $\text{minRev}(q^*,t^*) = \mathbb{E}_v\left[c\left(\sum_i q_i^*(v)\right)\right]$. To show that $\text{minRev}(q,t) \leq \text{minRev}(q^*,t^*)$, let $\sum_i U_i^n$ ($\sum_i U_i^*$) be the surplus from the truth-telling BNE of (q,t) ((q^*,t^*)). Since (q^*,t^*) solves c-Surplus Max and $c \leq \hat{c}$, we have that $\sum_i U_i^n \leq \sum_i U_i^*$.

Suppose $\sum_i U_i^n < \sum_i U_i^*$. Suppose towards a contradiction that $\text{minRev}(q,t) > \text{minRev}(q^*,t^*)$. Then, there exists collusive outcome (κ, q^c, t^c) possible against (q^*,t^*) such that $Rev(\kappa,q^c,t^c) < \text{minRev}(q,t)$. But, then, it is possible to construct a collusive outcome $(\tilde{\kappa},\tilde{q}^c,\tilde{t}^c)$ against (q,t) such that $Rev(\tilde{\kappa},\tilde{q}^c,\tilde{t}^c) < \text{minRev}(q,t)$, which is a contradiction.

To construct this collusive outcome, fix $\alpha \in [0,1)$ very close to 1 so that $\sum_i U_i^n < \alpha\sum_i U_i^*$. The construction that follows mirrors arguments in the proof of Lemma 3, and as there, transfers in collusive outcomes will refer to the total transfer paid by a bidder to the principal and to the collusive arrangement. See the proof of Lemma 3 for further discussion.

As an intermediate step, mix between executing (κ, q^c, t^c) and all bidders playing $a^0 = (a_1^0, ..., a_n^0)$ in a "scaled down" (κ, q^c, t^c). The construction of this intermediate collusive outcome is the same as the construction of $(\kappa^{scale}, q^{scale}, t^{scale})$ in the proof of Lemma 3. I omit the construction here to save space. Call this intermediate collusive outcome also $(\kappa^{scale}, q^{scale}, t^{scale})$. As in the proof of Lemma 3, let $Q(v) = \sum_{a \in \text{supp}(\kappa^{scale}(\cdot|v))} \kappa^{scale}(a|v) \sum_i q_i^*(a)$ be the aggregate probability with which the cartel acquires the good when values are v. Note that $Q(v) \in [0, \bar{q})$ for all $v \in V$.

Now, fix $\delta > 0$ such that $\sum_i U_i^n < \alpha \sum_i U_i^* - \delta$. For each v, there exists $\tilde{\kappa}(\cdot|v) \in \Delta A$ s.t. $\sum_{a \in \text{supp}(\tilde{\kappa}(\cdot|v))} \tilde{\kappa}(a|v) \sum_i q_i(a) \geq Q(v)$ and $\sum_{a \in \text{supp}(\tilde{\kappa}(\cdot|v))} \tilde{\kappa}(a|v) \sum_i t_i(a) < \hat{c}\left(Q\left(v\right)\right) + \delta$.

Define a constant:

$$\begin{aligned} D &:= \mathbb{E}_v \left[\sum_i t_i^{scale}(v) \right] - \mathbb{E}_v \left[\sum_{a \in \text{supp}(\tilde{\kappa}(\cdot|v))} \tilde{\kappa}(a|v) \left(\sum_i t_i(a) \right) \right] \\ &\geq \mathbb{E}_v \left[c\left(Q(v)\right] - \mathbb{E}_v \left[c\left(Q(c)\right] - \delta \geq v \right. \right. \end{aligned}$$

From there, we can define $\tilde{q}_i^c(v) = q_i^{scale}(v)$ and $\tilde{t}_i^c(v) = t_i^{scale}(v) - \frac{D}{N}$.

Let $\sum_i U_i^c$ be the bidder surplus from executing (κ, q^c, t^c) against (q^*, t^*). Let $\sum_i \tilde{U}_i^c$ be the bidder surplus from executing $(\tilde{\kappa}, \tilde{q}^c, \tilde{t}^c)$ against (q, t).

$$\begin{aligned} \sum_i \tilde{U}_i^c &= \alpha \sum_i U_i^c + D \geq \alpha \sum_i U_i^* - \delta \\ &\geq \sum_i U_i^n \end{aligned}$$

So, $(\tilde{\kappa}, \tilde{q}^c, \tilde{t}^c)$ is a collusive outcome possible against (q, t). Furthermore, taking the limit as $\delta \to 0$, we get the following inequalities: $\mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i^*(v) \right] \geq \mathbb{E}_v \left[c(Q(v)) \right] \geq \text{minRev}(q, t)$. This contradicts the premise that $\text{minRev}(q, t) > \text{minRev}(q^*, t^*)$.

Suppose $\sum_i U_i^n = \sum_i U_i^*$. Fix any collusive outcome (κ, q^c, t^c) against (q, t). For this collusive outcome to be possible, it must produce as much surplus as $\sum_i U_i^*$.

Since (q^*, t^*) solves relax-cSM and was constructed by combining the pointwise

solutions of reduced-vProb, the following inequality holds for all v:

$$\sum_i q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \leq \sum_i q_i^c(v) \max_j v_j - \hat{c}\left(\sum_i q_i^c(v)\right)$$
$$\leq Q^*(\max_i v_i) \max_i v_i - c\left(Q^*(\max_i v_i)\right)$$

since $\hat{c} \geq c$.

Since collusive arrangements must satisfy C-EABB, the expectation of the LHS is the expected surplus generated by the collusive outcome. As a result, the inequality must bind for almost all v since the collusive outcome must generate at least as much surplus as $\sum_i U_i^*$, which is the expectation of the term on the second line.

If $\hat{c}(\bar{q}) > c(\bar{q})$, then the set $\{v \in V : Q^*(\max_i v_i) = \bar{q}\}$ must have probability 0. So, with probability 1, $Q^*(\max_i v_i) < \bar{q}$. Furthermore, for the inequality to bind for almost all v, Lemma 8 implies that for almost all v, $Q^*(\max_i v_i) = \sum_i q_i^c(v)$ and $\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) = \hat{c}\left(Q^*(\max_i v_i)\right)$. As a result, the revenue from the collusive arrangement must equal the revenue from the BNE of (q^*, t^*):

$$\mathbb{E}_v\left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a)\right] = \mathbb{E}_v\left[c\left(Q^*(\max_i v_i)\right)\right]$$

Since this can be argued for all (κ, q^c, t^c) possible against (q, t), $\text{minRev}(q, t) = \text{minRev}(q^*, t^*)$, and the claim in the proposition holds. □

B.6 Outer Problem: Optimization of Cost Function

Since essentially direct mechanism (q^*, t^*) that solves c-Surplus Max allocates the good to the bidder with the highest realized value, we can also write the principal's problem as

$$\max_{\bar{q} \in [0,1], c:[0,\bar{q}] \to \mathbb{R}} \quad \mathbb{E}_v\left[c\left(Q^*(\max_i v_i)\right)\right]$$
$$\text{s.t. } Q^*(\max_i v_i) \text{ solves reduced-vProb}$$
$$c \text{ is a well-behaved cost function}$$

This problem can further be relaxed to the following problem:

$$\max_{\bar{q}\in[0,1],c:[0,\bar{q}]\to\mathbb{R}} \quad \mathbb{E}_v\left[c\left(Q^*(\max_i v_i)\right)\right]$$
$$\text{s.t. } Q^*(\max_i v_i) \text{ solves reduced-vProb}$$
$$\text{reduced-vProb} \geq 0 \quad \forall \max_i v_i$$

where the requirement that c be a well-behaved cost function has been replaced by the necessary condition that the value of reduced-vProb be weakly greater than 0 for each v. This condition is necessary because $c(0) \leq 0$, so as long as c is a cost function, the value of reduced-vProb must be weakly greater than 0 for all $\max_i v_i$.

Using the revelation principle, this relaxed problem is equivalent to designing a revenue-maximizing direct mechanism to sell the good to a single bidder subject to the usual incentive compatibility and individual rationality conditions. IC comes from the requirement that $Q^*(\max_i v_i)$ solves reduced-vProb, and IR comes from the requirement that the interim payoff of each $\max_i v_i$ is at least 0.

Using Myerson (1981) (1981), the optimal mechanism to set as a revenue-maximizing principal facing a single bidder is to post a price. Since the single bidder's value is $\max_i v_i$, the optimal price solves opt-price. Let p^* be the optimal price. This mechanism can be implemented indirectly with a cost function $c(Q) = p^* \cdot Q$ with domain $[0, 1]$. The optimal mechanism is solves c-Surplus Max with $c(Q) = p^* \cdot Q$ and $\bar{q} = 1$.

Returning to the statement of Theorem 1, let (A, q, t, σ) be a mechanism that satisfies the theorem. Because the principal collects p^*q in revenue at any action profile where the good is allocated with probability q, the principal's mechanism induces cost function $c(Q) = p^* \cdot Q$. The domain of the cost function is $[0, 1]$ because the good is allocated efficiently whenever $\max_i v_i > p^*$, which happens with positive probability. The BNE σ produces bidder surplus equal to the value of c-Surplus Max for $c(Q) = p^* \cdot Q$, $\bar{q} = 1$; to see this, observe that it solves v-Problem for almost all v by allocating the good efficiently whenever $\max_i v_i > p^*$ and bidders only pay p^* in total when the good is allocated. The essentially direct mechanism $(\tilde{q}, \tilde{t})$ which corresponds to this mechanism solves the principal's problem given the arguments above.

To see that the original mechanism also solves the principal's problem, note that an argument similar to that used to show Proposition 3 can be used to show that

$\text{minRev}(A, q, t, \sigma) = \text{minRev}(\tilde{q}, \tilde{t})$. Basically, we are concerned that there is a collusive arrangement possible against (A, q, t, σ) that would lower the revenue below that of the BNE, which is $\text{minRev}(\tilde{q}, \tilde{t})$. The intuition is the same as for Proposition 3; because (A, q, t, σ) generates non-collusive bidder surplus equal to c-Surplus Max, every collusive arrangement against (A, q, t, σ) must replicate the same outcome as the BNE.

Fix an arbitrary collusive outcome possible against (A, q, t, σ). Observe that

$$\sum_i \left(\tilde{q}_i(v)v_i - \tilde{t}_i(v)\right) \geq \sum_i q_i^c(v)v_i - \sum_a \kappa(a|v) \sum_i t_i(a)$$

The above inequality must bind with probability 1. With probability 1, there is a unique bidder with the highest realized value, and by Lemma 8, we have that $\sum_i q_i^c(v) = \sum_i \tilde{q}_i(v)$ with probability 1. Observe that $\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \geq p^* \sum_i q_i^c(v)$ for all v. This inequality must bind with probability 1 also. So

$$\mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] = \mathbb{E}_v \left[\sum_i \tilde{t}_i(v) \right]$$

Since this holds for all collusive outcomes of (A, q, t, σ), we have that $\text{minRev}(A, q, t, \sigma) = \text{minRev}(\tilde{q}, \tilde{t})$. □

C Theorem 2

I build the desired solution to EC by using the constructed mechanism that solves c-Surplus Max for $c(Q) = p^* \cdot Q$ with $\bar{q} = 1$. Let (q^*, t^*) denote the direct mechanism that solves c-Surplus Max for this cost function. See the proof of Proposition 2 for the exact definition of (q^*, t^*).

$$q_i^e(v) := q_i^*(v) \qquad \forall v, \forall i$$

$$\kappa^e(\max_i v_i, 0, ..., 0|v) := 1 \qquad \forall v$$

$$t_i^e(v) := \begin{cases} t_i^*(v) & \text{if } \max_i v_i < p^* \\ t_i^*(v) - p^* & \text{if } \max_i v_i \geq p^* \text{ and } i = 1 \\ t_i^*(v) & \text{if } \max_i v_i \geq p^* \text{ and } i \neq i \end{cases} \qquad \forall v$$

(κ^e, q^e, t^e) is ex post budget balanced, incentive compatible, and individually rational because the essentially direct mechanism (q^*, t^*) solves c-Surplus Max for $c(Q) = p^*q$, so it is indeed an efficient knockout auction.

Next, let (A, q, t, σ^h) be the mechanism produced by bringing any efficient knockout auction (κ^e, q^e, t^e) in-house. The truth-telling equilibrium of (A, q, t, σ^h) solves c-Surplus Max for cost function $c(Q) = p^*Q$ with domain $[0, 1]$. The proof of Proposition 3 can be used to show that the mechanism that brings (κ^e, q^e, t^e) in-house has worst-case revenue opt-price, so (A, q, t, σ^h) also solves the principal's problem. □

D Theorem 3

To prove the theorem, I follow the same strategy as the proof of Theorem 1.

In this setting, Lemma 1 does not necessarily go through.[10]

A result analogous to Lemma 6 holds. Given the principal sets essentially direct mechanism (q, t),

$$\inf_{\kappa, q^c, t^c} \mathbb{E}_v[\sum_{a \in A} \kappa(a|v) \sum_i t_i(a)] \tag{J-R}$$
$$\text{s.t. C-F', C-EABB, C-EAIR, C-IC}$$

has the same value as

$$\inf_{\kappa, q^c, t^c} \mathbb{E}_v[\sum_{a \in A} \kappa(a|v) \sum_i t_i(a)] \tag{I-R}$$
$$\text{s.t. C-F', C-EABB, C-EAIR, C-IC}$$

The proof of Lemma 6 can be used to show this claim.

Even though essentially direct mechanisms are not necessarily without loss, an essentially direct mechanism is still optimal.

[10] At a high level, Lemma 1 holds in the baseline model because for every mechanism that is not essentially direct, the corresponding essentially direct mechanism has a set of possible collusive outcomes that is a subset of the collusive outcomes possible given the original mechanism. This argument works because the worst-case is taken over the set of possible collusive outcomes. Because bidders are assumed to approximate surplus-maximizing, by "removing" actions from the mechanism by constructing a corresponding essentially direct mechanism, the principal may actually lower the supremal surplus colluding bidders might achieve and actually lower the revenue that colluding bidders might achieve.

D.1 Cost Function Properties

First, the class of possible cost functions must be adjusted because colluding bidders can no longer dispose of the good.

For each feasible mechanism (A, q, t, σ), the following set contains all the aggregate outcomes induced by playing an action profile:

$$P^{F'}_{(A,q,t,\sigma)} := \left\{ (Q,T) : \exists a \in A \text{ s.t. } \sum_i q_i(a) = Q, \sum_i t_i(a) = T \right\}$$

From there, it is possible to define the set of all aggregate outcomes that can be induced by mixing over action profiles:

$$\hat{P}^{F'}_{(A,q,t,\sigma)} := conv\left(P^{F'}_{(A,q,t,\sigma)}\right)$$

where $conv\left(P^{F'}_{(A,q,t,\sigma)}\right)$ is the convex hull of $P^{F'}_{(A,q,t,\sigma)}$. The *restricted cost function* of (A, q, t, σ) is defined as

$$c^{F'}_{(A,q,t,\sigma)}(Q) := \inf\left\{ T : (Q,T) \in \hat{P}^{F'}_{(A,q,t,\sigma)} \right\}$$

for all $Q \in \left\{ \tilde{Q} : (\tilde{Q}, \tilde{T}) \in \hat{P}^{F'}_{(A,q,t,\sigma)} \right\}$. Note that $\left\{ \tilde{Q} : (\tilde{Q}, \tilde{T}) \in \hat{P}^{F'}_{(A,q,t,\sigma)} \right\}$ is convex since $\hat{P}^{F'}_{(A,q,t,\sigma)}$ is convex. Let $\bar{q} := \sup\left\{ Q : (Q,T) \in \hat{P}^{F'}_{(A,q,t,\sigma)} \right\}$. Note that $\bar{q} \in [0,1]$.

Since bidders cannot dispose of the good, these restricted cost functions do not have all the properties of cost functions in the baseline model:

Lemma 10. *Fix a feasible mechanism (A, q, t, σ). Let $c^{F'}$ be the induced restricted cost function. $c^{F'}$ is (1) continuous on $(0, \bar{q})$ and (2) convex. Additionally, (3) $c^{F'}(0) \leq 0$.*

Proof. The proof of Lemma 2 can be used to demonstrate the claim. □

Unlike c, $c^{F'}$ is not necessarily increasing and is not continuous at 0. To establish these properties, Lemma 2 utilized colluding bidders' ability to dispose of the good. While $c^{F'}$ is not necessarily continuous at $Q = 0$ and $Q = \bar{q}$, it is upper semicontinuous.

Lemma 11. *Fix a feasible mechanism (A, q, t, σ). Let $c^{F'}$ be the induced cost function, and suppose that the supremum of the domain of $c^{F'}$ is $\bar{q}$. Then, $\lim_{Q\to 0^+} c^{F'}(Q) \leq c^{F'}(0)$. If $c^{F'}$ includes $\bar{q}$ in the domain, then $\lim_{Q\to \bar{q}^-} c^{F'}(Q) \leq c^{F'}(\bar{q})$.*

Proof. Fix $\varepsilon > 0$. By the definition of $c^{F'}$, the following inequality holds for all $\alpha \in [0, 1]$:

$$\alpha c^{F'}(0) + (1-\alpha)c^{F'}(\varepsilon) \geq c^{F'}\left((1-\alpha)\varepsilon\right)$$

Taking the limit as $\alpha \to 0$,

$$c^{F'}(0) \geq \lim_{Q\to 0^+} c^{F'}(0)$$

A similar argument can be used to demonstrate that $\lim_{Q\to \bar{q}^-} c^{F'}(Q) \leq c^{F'}(\bar{q})$. □

From here, I call any function that satisfies the properties given in Lemmas 10 and 11 a *restricted cost function*. From here, I restrict attention to bounded cost functions since mechanisms that induce an unbounded cost function result in $-\infty$ worst-case revenue. See footnote 9 for a discussion. If a restricted cost function c also has a compact domain, is bounded, and is continuous, then I call it a *well-behaved* restricted cost function.

D.2 Surplus Maximization Problem

I construct a class of feasible mechanisms that it will be without loss of optimality for the principal to maximize over. To do this, I again solve a constrained surplus maximization problem c-Surplus Max for well-behaved restricted cost functions.

D.2.1 Relaxed c-Surplus Max

Fix a well-behaved restricted cost function c with domain $[0, \bar{q}]$.

To solve c-Surplus Max, I relax c-Surplus Max to relax-cSM and solve the problem pointwise for each value v. v-Problem is the pointwise problem. reduced-vProb is the problem after observing that the cost constraint $\sum_i t_i \geq c\left(\sum_i q_i\right)$ binds at the optimum, the good should be allocated efficiently, and that the problem can be rewritten with the total probability with which to acquire the good as the sole choice variable.

Let $Q^*(\cdot)$ denote a selection of maximizers for reduced-vProb.

When c-Surplus Max is solved under the baseline model, Lemma 7 gives relevant properties of the solution of reduced-vProb. Note that Lemma 7 holds for restricted cost functions too.

D.2.2 Construction of WLOO Class of Feasible Mechanisms

Now, fix $Q^*(\max_i v_i) := \max \operatorname{argmax}_{Q\in[0,\bar{q}]} Q \max_i v_i - c(Q)$, and define the following allocation function for all i and v:

$$q_i^*(v) = \begin{cases} \frac{1}{|\{v_j : v_j = \max_k v_k\}|} Q^*(\max_j v_j) & \text{if } v_i = \max_j v_j \\ 0 & \text{otherwise} \end{cases}$$

Given any transfer function t^* such that $\sum_i t_i^*(v) = c\left(Q^*\left(\max_j v_j\right)\right)$ for all v, (q^*, t^*) solves relax-cSM.

Since Lemma 7 goes through in this setting, Proposition 2 also goes through in this setting. There exists a direct mechanism (q^*, t^*) that solves c-Surplus Max and relax-cSM for restricted cost function c. Following B.5.3, (q^*, t^*) can be incorporated into an essentially direct mechanism, which is feasible for the principal's problem. Abusing notation, let (q^*, t^*) denote the resulting essentially direct mechanism.

For mechanisms that solve c-Surplus Max for some continuous restricted cost function, the value of the worst-case revenue under this more restricted model of possible collusive outcomes is given by the following result:

Proposition 5. *Let $c : [0, \bar{q}] \to \mathbb{R}$ be a continuous restricted cost function. Let (q^*, t^*) be the essentially direct mechanism that solves c-Surplus Max for c. Then,*

$$\mathit{minRevR}(q^*, t^*) = \mathit{minRev}(q^*, t^*) = \mathbb{E}_v\left[\sum_i t_i^*(v)\right]$$

Proof. First, note that the proof of Proposition 3 can be used to show that $\text{minRev}(q^*, t^*) = \mathbb{E}_v\left[\sum_i t_i^*(v)\right]$ holds here too. The proof of Proposition 3 rests on the fact that (q^*, t^*) is constructed using the solution to reduced-vProb for each $\max_i v_i$ and does not utilize that $c(\cdot)$ is increasing.[11]

[11] Alternatively, it is possible to demonstrate that given $\tilde{c}$, the cost function by (q^*, t^*), (q^*, t^*) solves c-Surplus Max.

Now, consider the following collusive outcome (κ, q^c, t^c):

$$\begin{aligned} q_i^c(v) &= q_i^*(v) && \forall v, i \\ t_i^c(v) &= 0 && \forall v, i \\ \kappa(v|v) &= 1 && \forall v \end{aligned}$$

Note that $S(A,q,t,\sigma) = \mathbb{E}_v\left[\sum_i q_i^*(v)v_i - \sum_i t_i^*(v)\right]$. As a result, (κ, q^c, t^c) satisfies C-F', C-EABB, C-EAIR, C-IC, and $\left|\mathbb{E}_v\left[\sum_i q_i^*(v)v_i - \sum_i t_i^*(v)\right] - S(A,q,t,\sigma)\right| < \varepsilon$ for all $\varepsilon > 0$. It is possible to bound $\text{minRevR}(q^*, t^*)$ from above:

$$\mathbb{E}_v\left[\sum_i t_i^*(v)\right] \geq \text{minRevR}(q^*, t^*)$$

Next, it is possible to bound $\text{minRevR}(q^*, t^*)$ from below using $\text{minRev}(q^*, t^*)$. This is true for all mechanisms (A, q, t, σ), so I capture this in a lemma:

Lemma 12. *For every feasible* (A, q, t, σ)*,*

$$\textit{minRev}(A,q,t,\sigma) \leq \textit{minRevR}(A,q,t,\sigma)$$

Proof. For every $\varepsilon > 0$, the following in equality holds

$$\begin{aligned} \text{minRev}(A,q,t,\sigma) = \inf\left\{\mathbb{E}_v\left[\sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)\sum_i t_i(a)\right] : (\kappa,q^c,t^c) \text{ s.t. C-F, C-EABB, C-EAIR}\right\} \\ \leq \inf\left\{\mathbb{E}_v\left[\sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)\sum_i t_i(a)\right] : (\kappa,q^c,t^c) \text{ s.t. C-F', C-EABB,} \right. \\ \left. \text{C-EAIR, C-IC,} \left|\mathbb{E}_v\left[\sum_i q_i^c(v)v_i - \sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)\sum_i t_i(a)\right]\right| < \varepsilon\right\} \end{aligned}$$

Taking the limit as $\varepsilon \to 0$,

$$\text{minRev}(A,q,t,\sigma) \leq \text{minRevR}(A,q,t,\sigma)$$

□

I conclude $\text{minRevR}(q^*, t^*) = \text{minRev}(q^*, t^*) = \mathbb{E}_v\left[\sum_i t_i^*(v)\right]$. □

Every feasible (A, q, t, σ) with induced restricted cost function $\hat{c}$ can be associated with a well-behaved restricted cost function:

Lemma 13. *Let $\hat{c}$ be the restricted cost function induced by (A, q, t, σ), and let $\bar{q}$ be the supremum of its domain. Define $c : [0, \bar{q}] \to R$ so that*

$$c(Q) = \begin{cases} \hat{c}(Q) & \text{if } Q \neq 0, \bar{q} \\ \lim_{Q \to 0^+} \hat{c}(Q) & \text{if } Q = 0 \\ \lim_{Q \to \bar{q}^-} \hat{c}(Q) & \text{if } Q = \bar{q} \end{cases}$$

Then, c is also a restricted cost function.

Proof. To verify that c is a cost function, I check the properties listed in Lemmas 10 and 11. Clearly, the properties in Lemma 11 are satisfied by c. For Lemma 10, (1) is inherited from $\hat{c}$. Since $\hat{c}(0) \geq \lim_{Q \to 0^+} c(Q)$ and $0 \geq \hat{c}(0)$, $c(0) \leq 0$, so (3) holds for c.

Finally, to check convexity (2), I check that the following inequalities hold for all $\alpha \in (0, 1)$, $Q \in (0, \bar{q})$:

$$\begin{aligned} \alpha c(\bar{q}) + (1 - \alpha) c(Q) &\geq c\left(\alpha \bar{q} + (1 - \alpha)\, Q\right) \\ \alpha c(0) + (1 - \alpha) c(Q) &\geq c\left((1 - \alpha)\, Q\right) \\ \alpha c(\bar{q}) + (1 - \alpha) c(0) &\geq c\left(\alpha \bar{q}\right) \end{aligned}$$

If $\bar{q} = 0$, then these inequalities clearly hold.

Otherwise, for the first inequality, fix a small $\varepsilon > 0$. Note that $\alpha \bar{q} + (1 - \alpha) Q \in (0, \bar{q})$. The following inequality holds

$$\alpha \hat{c}(\bar{q} - \varepsilon) + (1 - \alpha) \hat{c}(Q) \geq \alpha \hat{c}\left(\alpha(\bar{q} - \varepsilon) + (1 - \alpha) Q\right)$$

Taking the limit as $\varepsilon \to 0$ and using the continuity of $\hat{c}$ and the definition of c, we have that

$$\alpha c(\bar{q}) + (1 - \alpha) \hat{c}(Q) \geq \alpha \hat{c}\left(\alpha \bar{q} + (1 - \alpha) Q\right)$$

The second inequality can also be shown for arbitrary $Q \in (0, \bar{q})$ using a similar argument. I omit this argument.

The third inequality can be shown by observing that

$$\alpha\hat{c}(\bar{q}-\varepsilon)+(1-\alpha)\hat{c}(\varepsilon)\geq\hat{c}\left(\alpha\left(\bar{q}-\varepsilon\right)+(1-\alpha)\varepsilon\right)$$

holds for small $\varepsilon>0$. Observe that $\alpha\bar{q}\in(0,\bar{q})$ since $\alpha\in(0,1)$. As a result, we can again take the limit as $\varepsilon\to 0$ to get that $\alpha c(\bar{q})+(1-\alpha)c(0)\geq c(\alpha\bar{q})$. □

The next result establishes that every mechanism (A,q,t,σ) can be mapped to an essentially direct mechanism that solves c-Surplus Max for a well-behaved restricted cost function:

Proposition 6. *Let (A,q,t,σ) be a feasible mechanism. Suppose $\hat{c}$ is the restricted cost function induced by (A,q,t,σ), and let $\bar{q}$ be the supremum of the domain of $\hat{c}$. Define*

$$c(Q)=\begin{cases}\hat{c}(Q) & \text{if } Q\neq 0,\bar{q}\\ \lim_{Q\to 0^{+}}\hat{c}(Q) & \text{if } Q=0\\ \lim_{Q\to\bar{q}^{-}}\hat{c}(Q) & \text{if } Q=\bar{q}\end{cases}$$

Let (q^{*},t^{*}) be the essentially direct mechanism that solves c-Surplus Max for restricted cost function c.

Then, $\text{minRevR}(q^{*},t^{*})\geq\text{minRevR}(A,q,t,\sigma)$.

Proof. The arguments in the proof of Proposition 4 can be used to establish the result. I divide the argument into two cases. Let $\sum_i U_i^n$ be the bidder surplus generated by (A,q,t,σ) and $\sum_i U_i^*$ be the bidder surplus generated by (q^{*},t^{*}).

In the first case, $\sum_i U_i^n<\sum_i U_i^*$. Throughout this case, we will use t^c to refer to the total transfers that bidders have to pay; see the proof of Lemma 3 for a discussion of how specifying total transfers are sufficient for side payments. Note that the bidder surplus generated by (q^{*},t^{*}) is $S(q^{*},t^{*})$. Since $\hat{c}\geq c$, $S(A,q,t,\sigma)\leq S(q^{*},t^{*})$. Suppose that $\text{minRevR}(A,q,t,\sigma)-\text{minRevR}(q^{*},t^{*})>\beta>0$. Then, for all $\varepsilon>0$, there exists a collusive outcome $(\kappa_\varepsilon,q_\varepsilon^c,t_\varepsilon^c)$ that satisfies C-F', C-EABB, C-EAIR, C-IC, and $\left|\mathbb{E}_v\left[\sum_i q_i^c(v)v_i-\sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)\sum_i t_i^*(a)\right]-S(q^{*},t^{*})\right|<\varepsilon$ and has revenue at most $\text{minRevR}(q^{*},t^{*})+\beta/2$. Furthermore, $\left|\mathbb{E}_v\left[\sum_i q_i^c(v)v_i-\sum_{a\in\text{supp}(\kappa(\cdot|v))}\kappa(a|v)\sum_i t_i^*(a)\right]-S(A,q,t,\sigma)\right|<\varepsilon$.

If $\bar{q} = 0$, then the $\text{minRevR}(A, q, t, \sigma) = \hat{c}(0) = c(0)$ since the cartel can acquire a payment arbitrarily close to $\hat{c}(0)$ and split the it in lump-sum constant payments among the bidders in a IC way to ensure that C-EAIR is satisfied. C-F' is trivially satisfied.

Otherwise, as in the proof of Proposition 4, we will construct a "scaled" version of $(\kappa_\varepsilon, q^c_\varepsilon, t^c_\varepsilon)$ to ensure that the collusive outcome does not require the cartel to acquire the good with aggregate probabilities 0 or $\bar{q}$. Then, we will translate the scaled collusive outcome to a collusive outcome possible against (A, q, t, σ).

To construct the scaled version of $(\kappa_\varepsilon, q^c_\varepsilon, t^c_\varepsilon)$, fix $\alpha \in (0, 1)$ and $\delta > 0$. There exists $(\delta, c(\delta) + \gamma)$ for $\gamma \geq 0$ in $conv\left(P^{F'}_{(q^*,t^*)}\right)$. Let $\kappa^l(\cdot) \in A$ denote the distribution over action profiles that corresponds to $(\delta, \hat{c}(\delta) + \gamma)$. Define the scale collusive outcome

$$\begin{aligned}
q^{scale}_{\varepsilon,i}(v) &= \alpha q^c_{\varepsilon,i}(v) + (1-\alpha)\frac{\delta}{N} && \forall v, \forall i \\
t^{scale}_{\varepsilon,i}(v) &= \alpha t^c_{\varepsilon,i}(v) + (1-\alpha)\frac{c(\delta)+\gamma}{N} && \forall v, \forall i \\
\kappa^{scale}_{\varepsilon}(v) &= \alpha \kappa^c_{\varepsilon}(v) + (1-\alpha)\kappa^l && \forall v
\end{aligned}$$

Following the proof of Proposition 4, define collusive outcome $(\tilde{\kappa}_\varepsilon, \tilde{q}^c_\varepsilon, \tilde{t}^c_\varepsilon)$ possible for (A, q, t, σ) using $(\kappa^{scale}_\varepsilon, q^{scale}_\varepsilon, t^{scale}_\varepsilon)$ for $(\kappa^{scale}, q^{scale}, t^{scale})$. This $(\tilde{\kappa}_\varepsilon, \tilde{q}^c_\varepsilon, \tilde{t}^c_\varepsilon)$ can be constructed so that

$$\begin{aligned}
\mathbb{E}_v\left[\sum_{a\in\text{supp}(\tilde{\kappa}(\cdot|v))} \tilde{\kappa}_\varepsilon(a|v)\sum_i t_i(a)\right] &< \mathbb{E}_v\left[\sum_{a\in\text{supp}(\kappa^{scale}_\varepsilon(\cdot|v))} \kappa^{scale}_\varepsilon(a|v)\sum_i t^*_i(a)\right] + \beta/4 \\
&\leq \alpha\left(\text{minRevR}(q^*, t^*)\right) + (1-\alpha)\left(c\left(\delta\right) + \gamma\right) + 3\beta/4
\end{aligned}$$

Observe that because $(\kappa^{scale}_\varepsilon, q^{scale}_\varepsilon, t^{scale}_\varepsilon)$ is a convex combination of IC collusive outcomes $(\kappa_\varepsilon, q^c_\varepsilon, t^c_\varepsilon)$ and a constant lottery that allocates the good with probability δ, $(\kappa^{scale}_\varepsilon, q^{scale}_\varepsilon, t^{scale}_\varepsilon)$ also satisfies cartel IC. $(\tilde{\kappa}_\varepsilon, \tilde{q}^c_\varepsilon, \tilde{t}^c_\varepsilon)$ inherits C-IC from $(\kappa^{scale}_\varepsilon, q^{scale}_\varepsilon, t^{scale}_\varepsilon)$. Similarly, because $(\kappa^{scale}_\varepsilon, q^{scale}_\varepsilon, t^{scale}_\varepsilon)$ satisfies C-F', $(\tilde{\kappa}_\varepsilon, \tilde{q}^c_\varepsilon, \tilde{t}^c_\varepsilon)$ satisfies C-F'. The proof of Proposition 4 verifies C-EABB holds for $(\tilde{\kappa}_\varepsilon, \tilde{q}^c_\varepsilon, \tilde{t}^c_\varepsilon)$. Since α can be chosen to be arbitrarily close to 1 and δ can be chosen to be arbitrarily close to 0, $\left|\mathbb{E}_v\left[\sum_i \tilde{q}^c_{\varepsilon,i}(v)v_i - \sum_{a\in\text{supp}(\tilde{\kappa}(\cdot|v))} \tilde{\kappa}_\varepsilon(a|v)\sum_i \tilde{t}_{\varepsilon,i}(a)\right] - S(A, q, t, \sigma)\right| < \varepsilon$ as well, and C-JEAIR is also satisfied.

Finally, for every ε, such a $(\tilde{\kappa}_\varepsilon, \tilde{q}^c_\varepsilon, \tilde{t}^c_\varepsilon)$ can be constructed, so taking the limit as

$\alpha \to 1$, we have

$$\text{minRevR}(A, q, t, \sigma) \leq \text{minRevR}(q^*, t^*) + 3\beta/4$$

Contradiction, so $\text{minRevR}(q^*, t^*) \geq \text{minRevR}(A, q, t, \sigma)$.

In the second case, $\sum_i U_i^n = \sum_i U_i^*$, and the proof utilizes that (q^*, t^*) solves reduced-vProb and that $\hat{c} \geq c$ to establish that any collusive outcome (κ, q^c, t^c) must produce the same revenue as $\text{minRev}(q^*, t^*) = \text{minRevR}(q^*, t^*)$. □

D.3 Outer Problem

Given that mechanisms that solve c-Surplus Max for some well-behaved restricted cost function are WLOO, the principal can solve her problem by maximizing over restricted cost functions and setting the mechanism that solves the induced c-Surplus Max problem. Once again, this establishes that the (q^*, t^*) that solves c-Surplus Max for restricted cost function $c(Q) = p^* \cdot Q$ with domain $[0, 1]$ solves the principal's problem.

To establish that any mechanism as described by Theorem 3 solves the principal's problem, the remainder of the argument follows B.6. For any such mechanism (A, q, t, σ), B.6 shows that $\text{minRev}(A, q, t, \sigma) = \text{minRev}(q^*, t^*)$. Using Lemma 12 and observing that $\mathbb{E}_{a \sim \sigma(\cdot|v), v} \left[\sum_i t_i(a)\right] \geq \text{minRevR}(A, q, t, \sigma)$[12], I conclude that $\text{minRevR}(A, q, t, \sigma) = \text{minRevR}(q^*, t^*)$, so (A, q, t, σ) also solves the principal's problem.

E Corollary 3

To prove this corollary, I show that $\text{minRevR}(A, q^p, t^p, \sigma) =$ opt-price. The mechanism (A, q^p, t^p, σ) defines a restricted cost function $c(Q) = p^* Q$ with domain $[0, 1]$. Let (q^*, t^*) be the essentially direct mechanism that solves c-Surplus Max for this restrict cost function. I define a collusive outcome (κ, q^c, t^c) of (A, q^p, t^p, σ) that achieves the

[12] See the proof of Proposition 5 for a similar argument.

value of relax-cSM for this c:

$$q_i^c(v) = q_i^*(v) \quad \forall v, \forall i$$

$$t_i^c(v) = \begin{cases} t_i^*(v) - p^* & \text{if } i = 1 \text{ and } \max_j v_j \geq p^* \\ t_i^*(v) & \text{if } i = 1 \text{ and } \max_j v_j < p^* \\ t_i^*(v) & \text{if } i \neq 1 \end{cases}$$

$$\kappa(\text{ buy, don't buy,...don't buy}|v) = 1 \quad \forall v \text{ s.t. } \max_j v_j \geq p^*$$

$$\kappa(\text{don't buy, don't buy, ..., don't buy}|v) = 1 \quad \forall v \text{s.t. } \max_j v_j < p^*$$

As a result, $S(A, q^p, t^p, \sigma) =$ relax-cSM.

Fix a sequence of $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$. For each ε_n, let $(\kappa_{\varepsilon_n}, q^c_{\varepsilon_n}, t^c_{\varepsilon_n})$ be a collusive outcome that satisfies C-EABB, C-F', C-IC, and C-EAIR and produces joint bidder surplus within ε_n of $S(A, q^p, t^p, \sigma)$, and let the revenue from each $(\kappa_\varepsilon, q^c_\varepsilon, t^c_\varepsilon)$, $\mathbb{E}_v\left[p^* \sum_i q^c_{\varepsilon_n,i}(v)\right]$, converge to minRevR$(A, q^p, t^p, \sigma)$. Thus, the probability that the cartel buys the good from the principal converges to minRevR$(A, q^p, t^p, \sigma)/p^*$. The surplus generated by the collusive outcomes can be bounded above:

$$\begin{aligned} \mathbb{E}_v\left[\sum_i q^c_{\varepsilon_n,i}(v)v_i - \sum_a \kappa_{\varepsilon_n}(a|v)\sum_i t_{\varepsilon_n,i}(a)\right] &= \mathbb{E}_v\left[\sum_i q^c_{\varepsilon_n,i}(v)v_i\right] - \mathbb{E}_v\left[\sum_i q^c_{\varepsilon_n,i}(v)p^*\right] \\ &\leq \mathbb{E}_v\left[\sum_i q^c_{\varepsilon_n,i}(v)\left(\max_j v_j - p^*\right)\right] \\ &\leq \mathbb{E}_v\left[\sum_i q^c_i(v)\max_j v_j\right] - \mathbb{E}_v\left[\sum_i q^c_i(v)p^*\right] \end{aligned}$$

If minRevR$(A, q^p, t^p, \sigma)/p^* < 1 - \prod_i F_i(p^*)$, then bidder surplus in the sequence of $\{(\kappa_{\varepsilon_n}, q^c_{\varepsilon_n}, t^c_{\varepsilon_n})\}$ is bounded away from $S(A, q, t, \sigma)$ since in the limit, the last inequality would hold strictly. This is a contradiction. So minRevR$(A, q^p, t^p, \sigma) \geq (1 - \prod_i F_i(p^*))p^* =$opt-price.

Since the cartel can execute (κ, q^c, t^c) to achieve $S(A, q^p, t^p, \sigma)$, minRevR$(A, q^p, t^p, \sigma) = (1 - \prod_i F_i(p^*))p^*$. □

F Theorem 4

First, suppose $\mathcal{A}$ does not include surplus maximization. Then, for all feasible mechanisms (A, q, t, σ), $\text{minRev}\mathcal{A}(A, q, t, \sigma)$ can be bounded:

$$\text{minRevR}(A, q, t, \sigma) \geq \text{minRev}\mathcal{A}(A, q, t, \sigma) \geq \text{minRev}(A, q, t, \sigma)$$

The lower bound follows because minRev an infimum over a larger set of collusive arrangements. To see the upper bound, note that, for each $\varepsilon > 0$, the following inequalities hold:

$$\inf_{(\kappa, q^c, t^c)} \left\{ \mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] : (\kappa, q^c, t^c) \text{ satisfies } \mathcal{A} \right\} \leq$$

$$\inf \left\{ \mathbb{E}_v \left[\sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] : (\kappa, q^c, t^c) \text{ s.t. C-F', C-EABB,} \right.$$

$$\left. \text{C-EAIR, C-IC,} \left| \mathbb{E}_v \left[\sum_i q_i^c(v) v_i - \sum_{a \in \text{supp}(\kappa(\cdot|v))} \kappa(a|v) \sum_i t_i(a) \right] \right| < \varepsilon \right\}$$

After taking the limit as $\varepsilon \to 0$, the upper bound follows.

Now, note that P-R upper bounds the value of P-$\mathcal{A}$. Fix any mechanism (A, q, t, σ) that satisfies Theorem 1. This mechanism solves both P-R and P, which both have value opt-price. Thus, $\text{minRev}\mathcal{A}(A, q, t, \sigma) =$ opt-price as well. (A, q, t, σ) solves P-$\mathcal{A}$.

Suppose $\mathcal{A}$ includes surplus maximization. If C-F' is in $\mathcal{A}$, then the arguments in the proof of Theorem 3 can be repeated to show that $\text{minRev}\mathcal{A}(q^*, t^*) = \text{minRev}(q^*, t^*)$ for (q^*, t^*) that solves c-Surplus Max for continuous restricted cost function c with compact domain and then that restricting attention to such mechanisms that solve c-Surplus Max is without loss. If C-F' is not in $\mathcal{A}$, then the arguments similar to the proof of Theorem 1 can be repeated to show that $\text{minRev}\mathcal{A}(q^*, t^*) = \text{minRev}(q^*, t^*)$ for (q^*, t^*) that solves c-Surplus Max for continuous cost function c with compact domain and then that restricting attention to such mechanisms that solve c-Surplus Max is without loss. Then, the outer problem optimization again delivers that a linear cost function with slope p^* and domain $[0, 1]$ is optimal.

□

G Non-Negative Transfers, Linear Cost Functions

G.0.1 Lemma 5

We directly check the Border inequalities (Border (1991) 1991):

$$\int_{v_i}^{1} \bar{q}(\tilde{v}) dF(\tilde{v}) \leq \frac{1 - F(v_i)^N}{N} \quad \forall v_i \in [0,1]$$

to verify the existence of an ex post allocation that corresponds to the interim allocation function. Note that the expression on the RHS is increasing as v_i decreases since $F(v_i)$ decreases as v_i increases. So it is sufficient to check the above inequality for $v_i \in [\hat{v}_t, 1]$ since satisfying the inequality at $v_i = \hat{v}_t$ implies it is satisfied for all $v_i < \hat{v}_t$.

Consider the LHS, dropping the i subscripts due to the symmetry:

$$\begin{aligned}
\int_v^1 \bar{q}(\tilde{v}) dF(\tilde{v}) &= \int_v^1 \Big(\int_{\hat{v}_t}^{\tilde{v}} \frac{d\bar{q}(\hat{v})}{d\hat{v}} d\hat{v} + \bar{q}(\hat{v}_t) \Big) dF(\tilde{v}) \\
&= \Big(\int_{\hat{v}_t}^{\tilde{v}} \frac{p}{\hat{v}} \tilde{f}(\hat{v}) d\hat{v} + \bar{q}(\hat{v}_t) \Big) (F(\tilde{v}) - F(v)) \big|_{\tilde{v}=v}^{1} - \int_v^1 \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (F(\tilde{v}) - F(v)) d\tilde{v} \\
&= \left(\int_{\hat{v}_t}^{1} \frac{p}{\hat{v}} \tilde{f}(\hat{v}) d\hat{v} + \bar{q}(\hat{v}_t) \right) (1 - F(v)) - \int_v^1 \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (F(\tilde{v}) - F(v)) d\tilde{v} \\
&= \bar{q}(\hat{v}_t)(1 - F(v)) + \int_{\hat{v}_t}^{v} \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (1 - F(v)) d\tilde{v} + \int_v^1 \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (1 - F(v)) d\tilde{v} \\
&\qquad - \int_v^1 \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (F(\tilde{v}) - F(v)) d\tilde{v} \\
&= \bar{q}(\hat{v}_t)(1 - F(v)) + \int_{\hat{v}_t}^{v} \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (1 - F(v)) d\tilde{v} + \int_v^1 \frac{p}{\tilde{v}} \tilde{f}(\tilde{v}) (1 - F(\tilde{v})) d\tilde{v}
\end{aligned}$$

Lemma 14. *If $\hat{v} > v > 0$, and the Border inequalities are satisfied with $\hat{v}_t = v$, then the Border inequalities are satisfied with $\hat{v}_t = \hat{v}$.*

Proof. To show this, we focus on how the LHS of the Border inequalities changes as $\hat{v}_t$ changes to demonstrate that as $\hat{v}_t$ increases, the LHS decreases. Recall that

$\bar{q}(\hat{v}_t) = \frac{p}{\hat{v}_t} F^{N-1}(\hat{v}_t)$. Let's then calculate the derivative of the LHS wrt $\hat{v}_t$ for $v > \hat{v}_t$:

$$\begin{aligned}
\frac{dLHS_u}{d\hat{v}_t} &= -\frac{p}{\hat{v}_t}\tilde{f}(\hat{v}_t)(1 - F(\hat{v}_t)) + (1 - F(v))\frac{d\bar{q}(\hat{v}_t)}{d\hat{v}_t} \\
&= -\frac{p}{\hat{v}_t}\tilde{f}(\hat{v}_t)(1 - F(\hat{v}_t)) + (1 - F(v))\left(\frac{\tilde{f}(\hat{v}_t)p}{\hat{v}_t} - \frac{F^{N-1}(\hat{v}_t)p}{(\hat{v}_t)^2}\right) \\
&= -\frac{p}{\hat{v}_t}\tilde{f}(\hat{v}_t)(F(v) - F(\hat{v}_t)) - (1 - F(v))\frac{F^{N-1}(\hat{v}_t)p}{(\hat{v}_t)^2} < 0
\end{aligned}$$

As a result, given $\hat{v} > v$ and that the Border inequalities are satisfied with $\hat{v}_t = v$, we have that for all $v \geq \hat{v}$, the change in the LHS going from $\hat{v}_t = v$ to $\hat{v}_t = \hat{v}$ is

$$\int_v^{\hat{v}} \frac{dLHS}{d\hat{v}_t} d\tilde{v} < 0$$

For $\tilde{v} < \hat{v}_t$, the LHS is

$$\bar{q}(\hat{v}_t)(1 - F(\hat{v}_t)) + \int_{\hat{v}_t}^1 \frac{p}{\tilde{v}}\tilde{f}(\tilde{v})(1 - F(\tilde{v}))d\tilde{v}$$

and the derivative of the LHS is

$$\begin{aligned}
\frac{dLHS_l}{d\hat{v}_t} &= -\frac{p}{\hat{v}_t}\tilde{f}(\hat{v}_t)(1 - F(\hat{v}_t)) + (1 - F(\hat{v}_t))\frac{d\bar{q}(\hat{v}_t)}{d\hat{v}_t} - f(\hat{v}_t)\bar{q}(\hat{v}_t) \\
&= -f(\hat{v}_t)\bar{q}(\hat{v}_t) - (1 - F(v))\frac{F^{N-1}(\hat{v}_t)p}{(\hat{v}_t)^2} < 0
\end{aligned}$$

For $\tilde{v} \in [v, \hat{v}]$, the change in the LHS is:

$$\int_v^{\tilde{v}} \frac{dLHS_u}{d\hat{v}_t}(\tau)d\tau + \int_{\tilde{v}}^{\hat{v}} \frac{dLHS_l}{d\hat{v}_t}(\tau)d\tau < 0$$

Since the LHS of the Border inequality decreases for every $\tilde{v}$, the Border inequalities continue to hold at $\hat{v}$. □

Lemma 15. *The Border inequalities are satisfied with $\hat{v}_t = p$.*

Proof. Take $\hat{v}_t = p$. As a result, $\bar{q}(p) = F^{N-1}(p)$. The Border inequalities are

$$F^{N-1}(p)(1 - F(v)) + \int_p^v \frac{p}{\tilde{v}}\tilde{f}(\tilde{v})(1 - F(v))d\tilde{v} + \int_v^1 \frac{p}{\tilde{v}}\tilde{f}(\tilde{v})(1 - F(\tilde{v}))d\tilde{v} \overset{?}{\leq} \frac{1 - F^N(v)}{N}$$

Let's take the derivative of both sides of the inequality with respect to v.

$$\begin{aligned}\frac{dLHS}{dv} &= -F^{N-1}(p)f(v) - f(v)\int_p^v \frac{p}{\tilde{v}}\tilde{f}(\tilde{v})d\tilde{v} \\ &\geq -f(v)\bigg(F^{N-1}(p) + \int_p^v \tilde{f}(\tilde{v})d\tilde{v}\bigg) \\ &= -f(v)\bigg(F^{N-1}(v)\bigg) \\ &= \frac{dRHS}{dv}\end{aligned}$$

So both LHS and RHS decrease as v increases, but the RHS decreases faster at every v. As a result, it suffices to check whether $LHS(1) \leq RHS(1)$ to establish that $LHS(v) \leq RHS(v)$ for all $v \in [\hat{v}_t, 1]$. It is clear that $LHS(1) \leq RHS(1)$ holds since both are equal to 0. So, we conclude that the Border inequalities are satisfied with $\hat{v}_t = p$. □

Observe that the LHS and RHS of the Border inequalities are continuous in $\hat{v}_t$ for $\hat{v}_t \in (0, 1)$, so let $\hat{v}_t$ be the infimal threshold v for which the Border inequalities are satisfied. The Border inequalities are satisfied at this $\hat{v}_t$, so the infimal $\hat{v}_t$ is the minimum $\hat{v}_t$ at which the inequalities hold.

Lemma 16. *If v_t is interior, then the Border constraint binds at v_t.*

Proof. Suppose the Border constraint did not bind at v_t so that

$$\frac{1 - F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) > 0$$

I will show that if $v_t > 0$, there exists $v_t' < v_t$ such that the Border constraints are satisfied.

$$\begin{aligned}\frac{d}{dv_t}\big[\frac{1 - F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})\big] &= -\frac{d\bar{q}(v_t)}{dv_t}(1 - F(v)) + \frac{p}{v_t}\tilde{f}(v_t)(1 - F(v)) \\ &= (1 - F(v))\frac{F^{N-1}(v_t)}{v_t^2}p \geq 0\end{aligned}$$

Further, we can express the difference between the RHS and LHS of the Border

inequality for v as:

$$\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) = \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) + \int_{v_t}^v \frac{d}{d\hat{v}}\left(\frac{1-F(\hat{v})^N}{N} - \int_{\hat{v}}^1 \bar{q}(\tilde{v})dF(\tilde{v})\right)d\hat{v}$$

Let's rewrite the last term:

$$\int_{v_t}^v \frac{d}{d\hat{v}}\left(\frac{1-F(\hat{v})^N}{N} - \int_{\hat{v}}^1 \bar{q}(\tilde{v})dF(\tilde{v})\right)d\hat{v} = \int_{v_t}^v f(\hat{v})\left(\tilde{F}(v_t)(\frac{p}{v_t}-1) + \int_{v_t}^{\hat{v}}(\frac{p}{\tilde{v}}-1)\tilde{f}(\tilde{v})d\tilde{v}\right)d\hat{v}$$

Our goal is to show that $\exists v_t' < v_t$ such that

$$\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) + \int_{v_t}^{v_t'} \frac{d}{dv_t}\big[\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})\big] \geq 0 \quad \forall v \geq v_t$$

We can rewrite the above inequality as:

$$\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq \int_{v_t'}^{v_t} \frac{d}{dv_t}\big[\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})\big] = (1-F(v))\int_{v_t'}^{v_t} \frac{F^{N-1}(\tilde{v}_t)}{\tilde{v}_t^2}pd\tilde{v}_t$$

The RHS is given by

$$\int_{v_t}^v f(\hat{v})\left(\tilde{F}(v_t)(\frac{p}{v_t}-1) + \int_{v_t}^{\hat{v}}(\frac{p}{\tilde{v}}-1)\tilde{f}(\tilde{v})d\tilde{v}\right)d\hat{v} + \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v})$$

This function of $v \in [v_t, 1]$ is single-peaked since the derivative with respect to v is

$$f(v)(\tilde{F}(v_t)(\frac{p}{v_t}-1) + \int_{v_t}^v (\frac{p}{\tilde{v}}-1)\tilde{f}(\tilde{v})d\tilde{v}$$

This derivative is strictly positive at $v = v_t$ since $v_t \leq p$. Furthermore, this derivative is strictly increasing on (v_t, p) and strictly decreasing on $(p, 1)$. If the derivative ever crosses 0, it must do so on the interval $[p, 1]$ since the derivative starts strictly positive. Furthermore, this point is unique, if it exists. Finally, observe that such a point must exist because $\frac{1-F(1)^N}{N} - \int_1^1 \bar{q}(\tilde{v})dF(\tilde{v}) = 0$. If the derivative was ≥ 0 over $[v_t, 1]$, then we would have $\frac{1-F(1)^N}{N} - \int_1^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) > 0$. So, call the unique point at which the derivative crosses 0, v^{th}. $v^{th} < 1$ by this argument.

Finally, we can choose $\varepsilon > 0$ small enough that $1 - \varepsilon > v^{th}$. The derivative is

strictly negative over the interval $[1-\varepsilon, 1]$, so define

$$k := \max_{\hat{v}\in[1-\varepsilon,1]} \tilde{F}(v_t)(\frac{p}{v_t} - 1) + \int_{v_t}^{\hat{v}} (\frac{p}{\tilde{v}} - 1)\tilde{f}(\tilde{v})d\tilde{v} < 0$$

k is well-defined because the derivative is continuous and $[1-\varepsilon, 1]$ is compact. Consider the affine function through $(1, 0)$ with slope k. Say this affine function has value y at $1-\varepsilon$. Now, define $y' := \min\{y, \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v})\} > 0$. Consider the affine function through $(1, 0)$ and (v_t, y'). Denote its slope by $-s$. Note that $s < 0$. This affine function is below $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})$ on the interval $[v_t, v^{th}]$ since $y' < \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v})$ and $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})$ is increasing on $[v_t, v^{th}]$. Next, for all $v \in [v^{th}, 1-\varepsilon]$, $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})$ is decreasing so $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq y \geq y'$ and $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})$ is above the affine function through $(1, 0)$ and (v_t, y'). Finally, the affine function through $(1, 0)$ and (v_t, y') is weakly below the affine function through $(1, 0)$ and $(1-\varepsilon, y)$. The affine function through $(1, 0)$ and $(1-\varepsilon, y)$ is by construction below $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})$ on $[1-\varepsilon, 1]$.

Next, choose $\delta > 0$. Consider $\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v})$ for $v \in [v_t - \delta, v_t]$. On this interval, we have that

$$\begin{aligned}\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) &= \frac{1-F(v)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) \\ &> \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) > 0\end{aligned}$$

We would like to show that $\exists v_t'' < v_t$ such that

$$\begin{aligned}\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq \int_{v_t''}^{v_t} &\mathbb{I}\{\tilde{v}_t \leq v\}\left[\frac{p}{\tilde{v}_t^2}\tilde{f}(\tilde{v}_t)(1-F(v))\right] \\ &+ \mathbb{I}\{\tilde{v}_t > v\}\left[\frac{p}{\tilde{v}_t^2}\tilde{f}(\tilde{v}_t)(1-F(\tilde{v}_t)) + \tilde{F}(\tilde{v}_t)f(\tilde{v}_t)\right]d\tilde{v}_t\end{aligned}$$

To this end, define for all $\tilde{v}_t \in [v_t - \delta, v_t]$

$$\kappa(\tilde{v}_t) := \sup_{v\in[v_t-\delta,v_t]} \mathbb{I}\{\tilde{v}_t \leq v\}\left[\frac{p}{\tilde{v}_t^2}\tilde{f}(\tilde{v}_t)(1-F(v))\right] + \mathbb{I}\{\tilde{v}_t > v\}\left[\frac{p}{\tilde{v}_t^2}\tilde{f}(\tilde{v}_t)(1-F(\tilde{v}_t)) + \tilde{F}(\tilde{v}_t)f(\tilde{v}_t)\right]$$

Notice that $\kappa(\tilde{v}_t) > 0$ since f and $\tilde{f}$ are continuous and so have maximum values

bounded away from 0 on this interval.

$$\int_{v_t''}^{v_t} \kappa(\tilde{v}_t)d\tilde{v}_t \geq \int_{v_t''}^{v_t} \mathbb{I}\{\tilde{v}_t \leq v\}\left[\frac{p}{\tilde{v}_t^2}\tilde{f}(\tilde{v}_t)(1-F(v))\right] + \mathbb{I}\{\tilde{v}_t > v\}\left[\frac{p}{\tilde{v}_t^2}\tilde{f}(\tilde{v}_t)(1-F(\tilde{v}_t)) + \tilde{F}(\tilde{v}_t)f(\tilde{v}_t)\right]d\tilde{v}_t$$

for each $v_t'' \in [v_t - \delta, v_t]$.

$$\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) > \frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq \int_{v_t''}^{v_t} \kappa(\tilde{v}_t)d\tilde{v}_t$$

Since the RHS of this inequality is continuous in v_t'', we can pick a v_t'' sufficiently small to make the above inequality hold.

Define $s' = \min\{s, \max_{v\in[v_t-\delta,v_t]} F(v)^{N-1}f(v)\} > 0$. Finally, observe that we can choose $v_t' < v_t$ so that $v_t - v_t' < \delta$, $v_t' < v_t''$ and

$$\frac{1-F(v_t)^N}{N} - \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq (1-F(v_t))\int_{v_t'}^{v_t} \frac{F^{N-1}(\tilde{v}_t)}{\tilde{v}_t^2}pd\tilde{v}_t$$

and

$$f(v_t)\int_{v_t'}^{v_t} \frac{F^{N-1}(\tilde{v}_t)}{\tilde{v}_t^2}pd\tilde{v}_t \leq s'$$

This is sufficient for

$$\frac{1-F(v)^N}{N} - \int_v^1 \bar{q}(\tilde{v})dF(\tilde{v}) \geq (1-F(v))\int_{v_t'}^{v_t} \frac{F^{N-1}(\tilde{v}_t)}{\tilde{v}_t^2}pd\tilde{v}_t \quad \forall v \in [v_t', 1]$$

We have shown the claim. □

G.0.2 Proposition 1

First, observe that if the principal sets a mechanism that induces a linear cost function with slope $p \leq 0$, then the worst-case revenue is weakly negative. To see this, observe that any $(\rho, \tilde{q}, \tilde{t})$ that does not coordinate colluders to induce aggregate outcomes (Q, T) on the cost function can be improved by coordinating colluders on aggregate outcomes $(Q, c(T) + \varepsilon)$ for arbitrarily small $\varepsilon > 0$. The resulting additional surplus from the principal's payments to bidders can be divided in lump-sum transfers; this new mechanism preserves colluder IC, EABB, EAIR, and feasibility.

If the principal's mechanism induces a cost function with slope $p \geq 1$, then worst-case revenue is again weakly negative. Given $\max_i v_i$ receives the good with probability Q, the total surplus generated is at most $Q \max_i v_i - Q \max_i v_i$ since bidders must together pay at least $Q \max_i v_i$. The joint surplus of the truth-telling BNE must be 0. As a result, each bidder playing their opt-out action is a feasible collusive arrangement, which results in a revenue of 0 for the principal.

To make positive worst-case revenue, the principal must set a mechanism with $p \in (0, 1)$. As long as the mechanism's BNE has a positive surplus, then the principal will make strictly positive worst-case revenue.

The principal's optimal mechanism must solve the NN-c-SM problem for $f(Q) = pQ$. The condition of Proposition 1 implies that each $p \in (0, 1)$ results in an interior v_t. For each $p \in (0, 1)$, I verify that the mechanism given in the statement of the proposition solves the c-Surplus Max problem using weak duality.

To do so, define the following auxiliary function:

$$\begin{aligned}\mathcal{L}(q, t, \lambda, \gamma, \mu, \beta) = &\int_v q_i(v)v_i - t_i(v)dF(v) + \sum_i \int_{v_i} \lambda_i(v_i)\left[- \frac{d\bar{q}_i}{d\tilde{v}_i}(\tilde{v}_i)\tilde{v}_i + \frac{d\bar{t}_i}{d\tilde{v}_i}\right]dv_i \\ &+ \sum_i \int_{\tilde{v}_i} \gamma_i \left[\bar{q}_i(\tilde{v}_i)\tilde{v}_i - \tilde{t}_i(\tilde{v}_i)\right]d\tilde{v}_i \\ &+ \int_v \mu(v)(\sum_i t_i(v) - p\sum_i q_i(v))dv + \int_v \beta(v)(1 - \sum_i q_i(v))dv\end{aligned}$$

Choose the multipliers on the local IC constraint to be

$$\lambda(v_i) = \frac{\int_{v_i}^1 \hat{v} - C dF_i(\hat{v})}{v_i}$$

for the subset $v_i \in [v_t^{IR}, 1]$ with v_t^{IR} given by the solution to:

$$\frac{\int_{v_t^{IR}}^1 \hat{v} dF_i(\hat{v})}{1 - F_i(v_t^{IR})} = C > 0$$

and C is given by

$$C := \frac{p\int_{v_t}^1 \hat{v} dF_i(\hat{v})}{f_i(v_t)(v_t)(p - v_t) + p(1 - F(v_t))}$$

Observe that $v_t^{IR} \leq v_t$ since

$$C = \frac{p\int_{v_t}^1 \hat{v}dF_i(\hat{v})}{f_i(v_t)(v_t)(p-v_t)+p(1-F(v_t))} \leq \frac{\int_{v_t}^1 \hat{v}dF_i(\hat{v})}{(1-F(v_t))}$$

since $p \geq v_t$ and additionally $\frac{\int_x^1 \hat{v}dF_i(\hat{v})}{1-F(x)}$ is increasing in x. Another feature of v_t^{IR} is that $\lambda(v_t^{IR}) = 0$. Furthermore, $C \geq v_t^{IR}$.

With these objects, we study a particular further relaxation, setting the local IC multipliers λ to be 0 for $v_i < v_t^{IR}$ and setting the IR multipliers γ to be 0 for $v_i \geq v_t^{IR}$:

$$\begin{aligned}\mathcal{L}(q,t,\lambda,\gamma,\mu,\beta) = &\int_v q_i(v)v_i - t_i(v)dF(v) + \sum_i \int_{v_t^{IR}}^1 \lambda_i(v_i)\Big[-\frac{d\bar{q}_i}{d\tilde{v}_i}(\tilde{v}_i)\tilde{v}_i + \frac{d\bar{t}_i}{d\tilde{v}_i}\Big]dv_i \\ &+ \sum_i \int_0^{v_t^{IR}} \gamma_i(\tilde{v}_i)\Big[\bar{q}_i(\tilde{v}_i)\tilde{v}_i - \bar{t}_i(\tilde{v}_i)\Big]d\tilde{v}_i \\ &+ \int_v \mu(v)(\sum_i t_i(v) - p\sum_i q_i(v))dv + \int_v \beta(v)(1-\sum_i q_i(v))dv\end{aligned}$$

Integration by parts and then collecting the coefficients of all $q_i(v)$ and $t_i(v)$ variables gives:

$$\begin{aligned}-f(v) - f_{-i}(v_{-i})\lambda_i'(v_i) + \mu(v) &\leq 0 \quad \text{if } v_i \geq v_t^{IR} \\ -f(v) - f_{-i}(v_{-i})\gamma_i(v_i) + \mu(v) &\leq 0 \quad \text{if } v_i < v_t^{IR} \\ f(v)(v_i) + f_{-i}(v_{-i})(\lambda_i(v_i) + \lambda_i'(v_i)v_i) - p\mu(v) - \beta(v)\mathbb{I}\{\max v \geq v_t\} &\leq 0 \quad \text{if } v_i \geq v_t^{IR} \\ f(v)(v_i) + f_{-i}(v_{-i})\gamma_i(v_i)v_i - p\mu(v) - \beta(v)\mathbb{I}\{\max v \geq v_t\} &\leq 0 \quad \text{if } v_i < v_t^{IR}\end{aligned}$$

We choose the remaining multipliers to be:

$$\beta(v) = \begin{cases} f(v)C - p\mu(v) & \forall v \text{ s.t. } \max_j v_j \geq v_t \\ 0 & \forall v \max_j v_j < v_t \end{cases}$$

$$\frac{\mu(v)}{f(v)} = 1 + \frac{\lambda_i'(\max_i v_i)}{f_i(\max_i v_i)} \quad \forall v \text{ s.t. } \max_i v_i \geq v_t$$

$$\frac{\mu(v)}{f(v)} = \frac{C}{v_t^{IR}} \quad \forall v \text{ s.t. } \max_i v_i < v_t^{IR}$$

$$\frac{\mu(v)}{f(v)} = 1 + \frac{\lambda_i'(v_t)}{f_i(v_t)} \quad \forall v \text{ s.t. } \max_i v_i \in [v_t^{IR}, v_t]$$

$$\gamma_i(v_i) = f_i(v_i)\left(\frac{C}{v_t^{IR}} - 1\right) \quad \forall v_i \in [0, v_t^{IR}], \quad \forall i$$

To verify dual feasibility,

1. Suppose $\max_j v_j \geq v_t$.

   - If $v_i \geq v_t^{IR}$, the $t_i(v)$ constraint holds if

   $$1 + \frac{\lambda_i'(\max_j v_j)}{f_i(\max_j v_j)} \leq 1 + \frac{\lambda_i'(v_i)}{f_i(v_i} \tag{1}$$

   which will hold if $1 + \lambda_i'(v_i)/f_i(v_i)$ is decreasing in v_i. The dual constraint corresponding to $q_i(v)$ holds at equality.

   - If $v_i < v_t^{IR}$, then the $q_i(v)$ constraint holds if $C/v_t^{IR} \leq 1 + \gamma_i(v_i)/f_i(v_i) \leq C/v_i$. The dual constraint correspond to $t_i(v)$ holds if

   $$1 + \frac{\lambda_j'(\max_j v_j)}{f_j(\max_j v_j)} \leq 1 + \frac{\gamma_i(v_i)}{f_i(v_i)} = \frac{C}{v_t^{IR}}$$

   Given $1 + \lambda_i'(v_i)/f_i(v_i)$ is decreasing in v_i, a necessary and sufficient condition for the above would be $1 + \frac{\lambda_j'(v_t^{IR})}{f_j(v_t^{IR})} \leq \frac{C}{v_t^{IR}}$, which is always satisfied, given the definition of v_t^{IR}.

2. Suppose $\max_j v_j < v_t$.

   - If $v_i \in [v_t^{IR}, v_t]$, then the $q_i(v)$ constraint is given by

   $$\frac{C}{p} \leq \frac{\mu(v)}{f(v)} = 1 + \frac{\lambda_i'(v_t)}{f_i(v_t)}$$

which holds since C is defined so that $C/p = 1 + \frac{\lambda_i'(v_t)}{f_i(v_t)}$. The $t_i(v)$ constraint is given by

$$1 + \frac{\lambda_i'(v_i)}{f_i(v_i)} \geq \frac{\mu(v)}{f(v)} = 1 + \frac{\lambda_i'(v_t)}{v_i(v_t)}$$

which holds as long as $1 + \frac{\lambda_i'(v_i)}{f_i(v_i)}$ is decreasing in v_i.

- If $v_i < v_t^{IR}$, then the $q_i(v)$ constraint is given by

$$1 + \frac{\gamma_i(v_i)}{f_i(v_i)} \geq \frac{\mu(v)}{f(v)} = \frac{C}{v_t^{IR}}$$

which holds by the choice of $\gamma_i(v_i)$.

The $t_i(v)$ constraint is given by the following

$$1 + \frac{\gamma_i(v_i)}{f_i(v_i)} \leq \frac{p}{v_i}\frac{\mu(v)}{f(v)} = \frac{p}{v_i}\frac{C}{v_t^{IR}}$$

which holds since $v_i < v_t^{IR} \leq p$ and the choice of $\gamma_i(v_i)$.

We must also verify that λ, γ, and β as constructed are all weakly positive. $\lambda(v_i) \geq 0$ holds for all $v_i \geq v_t^{IR}$ because $\int_{v_i}^1 \hat{v} - C dF_i(\hat{v}) \geq 0$. $\gamma_i(v_i) \geq 0$ is equivalent to $C \geq v_t^{IR}$, which holds by construction of v_t^{IR}. $\beta(v) \geq 0$ holds for all v s.t. $\max_i v_i \geq v_t$ if $C/p \geq 1 + \lambda_i'(v_1)/f_i(1) = C$ and $1 + \lambda_i'(v_i)/f_i(v_i)$ is decreasing. The first inequality clearly holds as long as $p \leq 1$.

So, to complete the verification of dual feasibility, it is sufficient to demonstrate that $1 + \frac{\lambda_i'(v_i)}{f_i(v_i)}$ is decreasing in v_i for all $v_i \in [v_t^{IR}, 1]$.

To ensure that $1 + \lambda_i'(v_i)/f_i(v_i)$ is decreasing in v_i, the derivative of the following expression should be decreasing

$$1 + \lambda_i'(\tilde{v})/f_i(\tilde{v}) = \frac{C}{\tilde{v}} - \frac{\int_{\tilde{v}}^1 \hat{v} - C dF_i(\hat{v})}{f_i(\tilde{v})(\tilde{v})^2} = \frac{C}{\tilde{v}} + \frac{1}{\tilde{v}^2}\frac{1 - F_i(\tilde{v})}{f_i(\tilde{v})}\big(C - \mathbb{E}[\hat{v}|\hat{v} \geq \tilde{v}]\big)$$

is decreasing. The derivative is given by

$$\begin{aligned}\frac{d}{dv}\left(\frac{C}{v}-\frac{1}{v^2}\frac{1-F_i(v)}{f_i(v)}(\mathbb{E}[\hat{v}|\hat{v}\geq v]-C)\right) &= -\frac{C}{v^2}+\frac{2}{v^3}\frac{1-F_i(v)}{f_i(v)}\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-C\right)-\frac{1}{v^2}\frac{d}{dv}\left(\frac{1-F_i(v)}{f_i(v)}\right)\\ &\quad -\frac{1}{v^2}\frac{1-F_i(v)}{f_i(v)}\frac{d}{dv}\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]\right)\\ &\leq 0 \quad \forall v\in[v_t^{IR},1]\end{aligned}$$

This is equivalent to

$$\begin{aligned}&\frac{2}{v}\frac{1-F_i(v)}{f_i(v)}\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-C\right)-\frac{d}{dv}\left(\frac{1-F_i(v)}{f_i(v)}\right)\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-C\right)-\frac{1-F_i(v)}{f_i(v)}\frac{d}{dv}\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]\right)\leq C\\ &\iff \frac{2}{v}\frac{1-F_i(v)}{f_i(v)}\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-C\right)-\frac{d}{dv}\left(\frac{1-F_i(v)}{f_i(v)}\right)\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-C\right)-\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-v\right)\leq C\end{aligned}$$

Since $C=\frac{\int_{v_t^{IR}}^1 \hat{v}dF_i(\hat{v})}{1-F_i(v_t^{IR})}$, the following sufficient condition that ensures the inequality holds:

$$\left(\frac{2}{v}\frac{1-F_i(v)}{f_i(v)}-\frac{d}{dv}\left(\frac{1-F_i(v)}{f_i(v)}\right)\right)\left(\mathbb{E}[\hat{v}|\hat{v}\geq v]-\mathbb{E}[\hat{v}]\right)\leq \mathbb{E}[\hat{v}]+\mathbb{E}[\hat{v}|\hat{v}\geq v]-v \quad \forall v\in[0,1]$$

given $\frac{2}{v}\frac{1-F_i(v)}{f_i(v)}-\frac{d}{dv}\left(\frac{1-F_i(v)}{f_i(v)}\right)\geq 0$ holds, which holds as long as $\frac{1-F_i(v)}{f_i(v)}$ is decreasing.

The last step to showing that the mechanism is optimal is to verify complementary slackness. The mechanism given in the proposition statement maximizes $\mathcal{L}$ given λ, γ, β, and μ. Notice that the local IC constraints hold at equality due to the payoff formula. Furthermore, the IR constraints hold at equality for all $v_i < v_t^{IR}$. Next, observe that if we could show that $1=\sum_i q_i(v)$ for almost all v where $\max v\geq v_t$, then we have finished verifying complementary slackness since for almost all v, we would have $\sum_i t_i(v) = p\sum_i q_i(v)$ and $1=\sum_i q_i(v)$. Since v_t is interior for any $p\in(0,1)$, then Lemma 16 tells us that the Border inequality binds at v_t. This implies that

$$Pr(i \text{ gets the good}) = \int_{v_t}^1 \bar{q}(\tilde{v})dF(\tilde{v}) = \frac{1-F(v_t)^N}{N} = \frac{1}{N}Pr(\max v\geq v_t)$$

Since the interim allocation is symmetric, we have that this implies

$$Pr(\text{any bidder gets the good}) = Pr(\max_j v_j \geq v_t).$$

The interim allocation function implies that the good is not allocated to any bidder when $\max v < v_t$, so verifying that the Border constraint binds at v_t is enough to show that $1 = \sum_i q_i(v)$ holds with probability 1 given that $\max_j v_j \geq v_t$.

References

Border, K. C. (1991): "Implementation of reduced form auctions: A geometric approach," *Econometrica: Journal of the Econometric Society*, 1175–1187.

Carroll, G. (2019): "Robustness in mechanism design and contracting," *Annual Review of Economics*, 11, 139–166.

Carroll, G. and I. Segal (2019): "Robustly optimal auctions with unknown resale opportunities," *The Review of Economic Studies*, 86, 1527–1555.

Chassang, S., K. Kawai, J. Nakabayashi, and J. Ortner (2022): "Robust screens for noncompetitive bidding in procurement auctions," *Econometrica*, 90, 315–346.

Chassang, S. and J. Ortner (2019): "Collusion in auctions with constrained bids: Theory and evidence from public procurement," *Journal of Political Economy*, 127, 2269–2300.

Che, Y.-K. and J. Kim (2006): "Robustly collusion-proof implementation," *Econometrica*, 74, 1063–1107.

——— (2009): "Optimal collusion-proof auctions," *Journal of Economic Theory*, 144, 565–603.

Graham, D. A. and R. C. Marshall (1987): "Collusive bidder behavior at single-object second-price and English auctions," *Journal of Political economy*, 95, 1217–1239.

Jackson, M. O. and H. F. Sonnenschein (2007): "Overcoming incentive constraints by linking decisions 1," *Econometrica*, 75, 241–257.

Jagadeesan, R., I. R. Segal, and A. Tordjman (2025): "The Limits of Auctions under Ex-Ante Collusion," .

Kawai, K. and J. Nakabayashi (2022): "Detecting large-scale collusion in procurement auctions," *Journal of Political Economy*, 130, 1364–1411.

Laffont, J.-J. and D. Martimort (1997): "Collusion under asymmetric information," *Econometrica: Journal of the Econometric Society*, 875–911.

——— (2000): "Mechanism design with collusion and correlation," *Econometrica*, 68, 309–342.

Mailath, G. J. and P. Zemsky (1991): "Collusion in second price auctions with heterogeneous bidders," *Games and Economic Behavior*, 3, 467–486.

McAfee, R. P. and J. McMillan (1992): "Bidding rings," *The American Economic Review*, 579–599.

Myerson, R. B. (1981): "Optimal auction design," *Mathematics of operations research*, 6, 58–73.

Pavlov, G. (2008): "Auction design in the presence of collusion," *Theoretical Economics*, 3, 383–429.

Smith, R. A. (1961a): "Part I: The Incredible Electrical Conspiracy," *Fortune*, 63, 132–180.

——— (1961b): "Part II: The Incredible Electrical Conspiracy," *Fortune*, 63, 161–222.