Robust collusion with private information*

David A. Miller
UCSD†

May 6, 2011

Accepted for publication in the *Review of Economic Studies*

Abstract

The game-theoretic literature on collusion has been hard pressed to explain why a cartel should engage in price wars, without resorting to either impatience, symmetry restrictions, inability to communicate, or failure to optimize. This paper introduces a new explanation that relies on none of these assumptions: if the cartel's member firms have private information about their costs, price wars can be optimal in the face of complexity. Specifically, equilibria that are robust to payoff-irrelevant disruptions of the information environment generically cannot attain or approximate efficiency. An optimal robust equilibrium must allocate market shares inefficiently, and may call for price wars under certain conditions. For a two-firm cartel, cost interdependence is a sufficient condition for price wars to arise in an optimal robust equilibrium. That optimal equilibria are inefficient generically applies not only to collusion games, but also to the entire *separable payoff environment* (Chung and Ely 2006)—a class that includes most typical economic models.

Keywords: Collusion, price wars, private information, repeated games, mechanism design, ex post incentive compatibility. *JEL Classifications:* C72, C73, D82, L13.

---

*Previous working versions of this paper were titled "Optimal ex post incentive compatible equilibria in repeated games of private information" and "The dynamic cost of ex post incentive compatibility in repeated games of private information."

†Address: Department of Economics, 9500 Gilman Dr., La Jolla, CA 92093–0508. E-mail: d9miller@ucsd.edu. Home page: http://dss.ucsd.edu/~d9miller. I am deeply grateful to Susan Athey for advice and comments, and am indebted to Bruno Biais and several anonymous referees for excellent critiques and suggestions. I also thank Timothy Bresnahan, Giacomo Calzolari, Harrison Cheng, Gordon Dahl, Peter Hammond, Johannes Hörner, Narayana Kocherlakota, Ivana Komunjer, Jonathan Levin, David Meyer, Moritz Meyer-ter-Vehn, Paul Milgrom, Amalia Miller, Larry Samuelson, Ilya Segal, Andrzej Skrzypacz, Giancarlo Spagnolo, Steven Tadelis, Joel Watson, Mark Wright, and Bill Zame—as well as seminar participants at Stanford, UT Austin, UCSD, Northwestern, Kellogg School of Management, Michigan, Brown, Yale School of Management, Wisconsin, UCLA, USC, UC Riverside, and the 2005 Econometric Society World Congress—for helpful comments. I thank the Stanford Institute for Economic Policy Research for financial support, and Yale University and the Cowles Foundation for Research in Economics for financial support and hospitality.

Contents

1 Introduction **3**
- 1.1 Robustness properties . . . 5
- 1.2 Optimal robust collusion . . . 6

2 Collusion with private information **7**
- 2.1 Collusion under perfect public equilibrium . . . 8
- 2.2 Robust collusion . . . 10
  - 2.2.1 Optimally inefficient allocation . . . 12
  - 2.2.2 Optimal price wars . . . 17
- 2.3 Generalizations . . . 19
  - 2.3.1 No monetary transfers . . . 19
  - 2.3.2 Private strategies . . . 19
  - 2.3.3 Imperfect monitoring . . . 21

3 General results **22**
- 3.1 Bounding equilibrium utility . . . 23
- 3.2 EPIC-rich games . . . 24

4 Conclusion **28**

A Equilibrium **30**

B Proof of Theorem 3 **31**

C Proofs for Section 3 **33**

D Proofs for Section 2 **35**

References **44**

1 Introduction

Firms collude to increase their aggregate profits at the expense of consumers. But collusion is challenging to sustain because each firm has an individual incentive to cheat on the arrangement. Longstanding intuition in industrial organization holds that a "price war"—a period of inefficiently low prices—erupts when collusive efforts falter.[1] However, the game theory literature has found difficulty modeling price wars among patient firms. This paper shows that price wars are a rational cartel response to complexity in the environment, even if the cartel is very patient. Specifically, this paper studies collusion under private information, focusing on equilibria that are robust to payoff-irrelevant disruptions of the information environment. In such equilibria, both inefficient allocation and price wars can be optimal even in the limit as the firms become perfectly patient, due to the burden of providing robust incentives to share their information truthfully.

Models based on full-information repeated games suggest that collusion need never break down on the equilibrium path, and therefore an effective, patient cartel should exhibit neither price wars nor allocative inefficiency. To answer this challenge, Green and Porter (1984) introduce monitoring imperfections. They show that price wars arise after bad outcomes, in the cartel's optimal "strongly symmetric" equilibrium. (In strongly symmetric strategies, the firms play only symmetric stage game action profiles.) The prospect of a price war deters any unobservable deviation, such as secret price cutting or overproduction, that would increase the probability of bad outcomes. Since bad outcomes arise with positive probability even when none of the firms deviates, price wars arise on the equilibrium path, preventing the cartel from attaining monopoly profits.

However, the price wars in the Green and Porter (1984) model are alleviated if the cartel can play asymmetric equilibria, as shown by Fudenberg, Levine, and Maskin (1994). Indeed, the folk theorem applies, so monopoly profits are attainable in the limit as the firms become more patient.[2] Athey and Bagwell (2001, 2008) model collusion with private information, such as information about production costs. Not only does the folk theorem apply in the limit,[3] but also even impatient firms prefer to mis-allocate market shares, rather than engage in price wars. Intuitively, if one firm needs to be punished, another firm can absorb its market share rather than initiating a price war. Broadly speaking, the literature suggests that monopoly profits arise in the limit, and, furthermore, under private information price wars are not optimal even if the cartel is impatient.

[1] The empirical and historical literatures have identified many instances of oligopolistic pricing patterns that appear to involve price wars. See, for example, Porter (1983) on the U.S. railroad industry in 1880–1886, Levenstein (1997) on the bromine industry in 1885–1914, Bresnahan (1987) on the U.S. auto industry in 1954–1956, Slade (1992) on the Vancouver gasoline market in 1983, Busse (2002) on the airline industry in 1985–1992, and Fabra and Toro (2005) on the Spanish electricity market in 1998.

[2] Harrington and Skrzypacz (2007) prove similar results in a model with a different monitoring structure.

[3] Applied to games with private information, the Fudenberg, Levine, and Maskin (1994) folk theorem requires statistically independent costs and private values across firms. Aoyagi (2007) and Miller (2009) provide conditions under which the folk theorem generalizes to collusion with correlated signals and interdependent values.

This paper shows that monopoly profits are not attainable under private information, even in the limit, among equilibria that are robust to payoff-irrelevant disruptions of the information environment. Inefficient allocation is pervasive in such equilibria. In addition, price wars arise, and persist in the limit, in cartels with interdependent costs (such as when private signals about costs reflect both common and idiosyncratic underlying shocks). This is the first game-theoretic explanation for price wars that does not rely on either impatience, a symmetry restriction, inability to communicate, or a failure to optimize.[4]

The robustness criterion I impose is *ex post incentive compatibility* (EPIC) in each period, taking expectations over the future path of play. This criterion was independently introduced by Athey and Miller (2007) and Bergemann and Välimäki (2010, who call it "periodic EPIC"). EPIC is necessary and sufficient for an equilibrium to be robust to the introduction of arbitrary *payoff-irrelevant signals*, which do not enter the firms' ex post profit functions but may be correlated with the contemporaneous payoff-relevant information. In practice, EPIC gives each firm the incentive to share its information without first trying to infer what its collaborators may have learned. For instance, there is no need for them to communicate through a disinterested mediator; they can speak face-to-face and in any order. Similarly, it does not matter whether they spy on each other, communicate privately as well as publicly, or inadvertently learn new payoff-irrelevant information along the way.

More generally, the inefficiency does not arise from features special to collusion models. I show that aggregate utility under EPIC is bounded away from efficiency across a great array of repeated games with private information. Specifically, aggregate payoffs are bounded by the value of implementing a one-shot outcome rule under EPIC and a no-subsidy condition, and this bound is tight under generous conditions. For the separable payoff environment that contains many important

[4]Theoretically, Sannikov and Skrzypacz (2007) generate price wars that worsen as the frequency of adjustment increases, in a model in which information arrives continuously. But for any fixed frequency of adjustment, price wars disappear as firms become more patient. Following Rotemberg and Saloner (1986), there is a substantial literature on price wars in a full-information environment with exogenous public demand shocks, where the effect is driven by the individual rationality constraints, but price wars disappear if the firms are sufficiently patient. Besides Green and Porter (1984), other studies invoking symmetry include Abreu, Pearce, and Stacchetti (1986) and Athey, Bagwell, and Sanchirico (2004). Blume and Heidhues (2006), Hörner and Jamison (2008), Skrzypacz and Hopenhayn (2004) study collusion with private information but without communication. Various intuitive but suboptimal collusive mechanisms have also been studied (e.g., Aoyagi 2003, Fershtman and Pakes 2000, Lopomo, Marshall, and Marx 2005).

Empirically, various studies of collusion in auctions suggest that failure to optimize is rampant. The results in this paper suggest that failure to optimize, such as by naively allocating efficiently, can lead to even more price wars (see Example 1). Kwoka (1997) describes a real estate cartel that sent a designated bidder to obtain properties at public auctions, and then used an ascending first-price "knockout" auction to allocate the property within the cartel. After the designated bidder was compensated, the remaining revenue from the knockout auction was distributed to the cartel members. Bidding your value in such an auction is not incentive compatible; instead, cartel members overbid in the knockout auction, leading the designated bidder to overbid in the public auction. Overbidding in the public auction can be interpreted as a price war. Similar results have been found in detailed case studies by Pesendorfer (2000) in school milk procurement auctions and Asker (2010) in collectible stamp auctions. Lopomo, Marshall, and Marx (2005) survey a range of cases with these characteristics. Outside of repeated auction environments, most studies of collusion and price wars—such as those cited in footnote 1—have focused on industries with relatively undifferentiated products. In such industries, private information is less likely to be a key driver of cartel behavior.

economic models, this bound generically rules out efficiency. These results provide a counterpoint to the folk theorem: If we ask for robustness, then efficiency is no longer attainable, even in the limit.

1.1 Robustness properties

A *perfect public equilibrium* (PPE, studied by most of the sources cited above) is a perfect Bayesian equilibrium in which firms condition their strategies in each period on only the public history and their current private information; they ignore their private histories. I define an *ex post perfect public equilibrium* (EPPPE) as a PPE that satisfies EPIC in every period.

Why are the robustness properties of EPPPE desirable? Suppose that in each period firms observe not only payoff-relevant signals but also payoff-irrelevant signals, and both types of signals may be statistically interdependent in arbitrary ways. (Throughout I assume that signals are independent across time periods, however.) In an ordinary PPE, each firm's incentive constraint for revealing its information truthfully applies to its interim beliefs, knowing its own signals but not those of the other firms. Given its own information, it computes its conditional expectations of other firms' actions given its beliefs about their beliefs, including its beliefs about their beliefs about its beliefs, and so on. These higher order beliefs are influenced by the payoff-irrelevant signals.[5] Hence to construct an ordinary PPE, one in principle must know the joint distribution over all signals, both payoff-relevant and payoff-irrelevant. Since constructing equilibria in the presence of complex higher order beliefs is a complex problem, the literature usually assumes that payoff-irrelevant signals are simply absent. However, this assumption is not robust, as Weinstein and Yildiz (2007) have shown that higher order beliefs can have significant equilibrium consequences.

The class of EPPPEs, in contrast, is the class of PPEs that can be constructed without any knowledge of the distribution over payoff-irrelevant signals (see Bergemann and Morris 2005). With respect to EPPPE, games that differ only in their payoff-irrelevant signals are equivalent. Put another way, a PPE that does not rely on the payoff-irrelevant details of the informational environment must be an EPPPE.

The robustness properties of EPPPE are relevant to collusive interaction in practice. Firms are organizations rather than individuals, and they face internal information aggregation and incentive compatibility issues even in the absence of collusion. A PPE is a fragile equilibrium, which can be disrupted by the slightest deviation from a strict simultaneous communication protocol. Therefore communication between firms must be restricted to the highest levels and subjected to rigid protocols. The firms must not allow any information to leak out prior to communication, even as they face incentives to spy on the information of their cartel partners. In a sufficiently complex information environment, constructing and maintaining an optimal PPE may be prohibitively dif-

[5]More precisely, when we consider the universal type space constructed from the payoff-relevant signals, payoff-irrelevant signals can generate arbitrary higher order beliefs.

ficult. In an EPPPE, on the other hand, each firm can casually share its information in a way that best suits the circumstances, without worrying about information leakage or communication protocols. Its employees can converse with their counterparts in other firms—engineers to engineers, procurement officers to procurement officers—to communicate costs more precisely. The ability to communicate without structure can dramatically ease their collaboration in jurisdictions with anti-trust enforcement, since they may not be able to predict when and how they will be able to exchange information. In short, even if it is feasible to play an optimal PPE, playing an EPPPE instead reduces the unmodeled burdens of maintaining rigid protocols in the face of complexity. These real-world advantages make EPPPE a potentially more appropriate equilibrium concept for the study of collusion.

1.2 Optimal robust collusion

EPPPEs are robust, but Section 2 shows that this robustness comes at a cost: cartel profits under EPPPE are generally bounded away from monopoly profits, and the bound applies uniformly, regardless of patience. The root problem is that if the firms allocate market shares efficiently, then as a group they must generally receive different aggregate continuation profits after different ex post realizations of their signals. For instance, under private values the cartel can run a second-price "knockout" auction (McAfee and McMillan 1992) to determine which firm serves the market, but it is then impossible to fully rebate the winner's payment back to the member firms without disrupting their incentives. So if the firms want to implement efficient allocation, they must burn money—such as by engaging in price wars—in order to provide the necessary incentives.[6]

In the tradeoff between efficient pricing and efficient allocation, efficient allocation turns out to be surprisingly expensive. The cartel is always willing to give up at least a little bit of allocative efficiency to reduce the severity of price wars, as illustrated in Examples 1–2 and proven in Theorem 1. In particular, the cartel can reduce the severity of price wars by ignoring efficiency when both firms' costs are very low. Then the firms' incentive payments for truthful revelation are also constant when their costs are very low. The revenues from constant incentive payments are easier to rebate back to the firms, allowing them to reduce the severity of their price wars. The gain from reducing the severity of price wars more than makes up for the loss in allocative efficiency. Moreover, in a canonical example (Example 3, featuring uniformly and independently distributed costs) this intuition scales up to the whole signal space: the cartel prefers to eliminate price wars entirely, by allocating very inefficiently.

On the other hand, if the firms' costs are interdependent then eliminating price wars would

[6]The need for price wars could be relaxed if the firms could obtain insurance against imbalances in their transfers. Since there is no exogenous authority (other than the antitrust authority who would frown on such an arrangement), there is no presumption that the firms ought to be able to obtain such insurance. If they could indeed accept budget imbalances (such as by self-insuring), these abilities should be modeled within the game. Athey and Miller 2007 take such an approach for a bilateral trading relationship.

require such extremes of inefficiency that the firms are always willing to tolerate some price wars, as illustrated in Example 5 and proven in Theorem 2. Costs will be interdependent if each firm's cost is influenced by both an idiosyncratic shock and a common shock. If their costs are interdependent and the allocation rule is not degenerate, then each firm's incentive payment must depend on comparisons between firms. Such payments are difficult to rebate back to the firms, so price wars are difficult to avoid. In addition, price wars may be optimal for some parameters even when costs are independent, as shown in Example 4. Section 2.3 explains why these results extend to several generalizations of the analysis: eliminating transfers, allowing equilibria in private strategies, and adding monitoring imperfections.

Section 3 expands the analysis to a general class of repeated games with private information. When the players are sufficiently patient, the problem of constructing an optimal EPPPE reduces to the static mechanism design problem of maximizing aggregate utility in the stage game, subject to a no-subsidy condition. Theorem 3 proves this equivalence formally by extending the techniques of Abreu, Pearce, and Stacchetti (1990), Fudenberg, Levine, and Maskin (1994), and Fudenberg and Levine (1994) to this setting. Therefore efficiency is attainable in an EPPPE if and only if there exists an efficient solution to the static mechanism design problem of maximizing aggregate utility subject to EPIC and ex post budget balance. Theorem 5 shows that in the separable payoff environment (Chung and Ely 2006), generically there does not exist such a mechanism.

2 Collusion with private information

Green and Porter (1984) note that standard full-information models of collusion normally yield optimal equilibria in which behavior along the equilibrium path is stationary and allocatively efficient, whereas in practice cartels often exhibit intermittent periods in which they offer inefficiently low prices. Such episodes are called *price wars*, and can arise in robust equilibria of collusion models with private information about their costs.

Consider a simple class of infinitely repeated collusion games with private information, generalizing Athey and Bagwell (2001). A continuum of identical price-taking consumers collectively demands one unit at any price less than or equal to 1, and zero units otherwise. Each firm $i \in \{1, \ldots, N\}$ has a cost function $c_i : \Theta \to [0, 1]$. In each period, the shock $\theta \in \Theta \equiv \Theta_1 \times \cdots \times \Theta_N$ that governs the firms' costs is realized, with each firm i observing only its private signal θ_i. These private signals are distributed according to a joint probability distribution ϕ, possibly with correlation across firms, but identically and independently across periods. In each period, after observing their respective signals, the firms can communicate. After communicating, each firm sets a price $p_i \in \mathbb{R}$, and the consumers buy from the firm with the lowest price. If multiple firms post the lowest price, then they may split the market in any proportions amongst themselves, by mutual agreement (as in Athey and Bagwell). At the end of the period, the firms can send each other monetary transfers. The firms share a common discount factor $\delta < 1$. Aside from the friction of private information and the absence of an external "budget breaker" to provide insurance against budget imbalances, this environment is quite permissive for the firms, giving them their best shot at collusion.

Let $w_i(\theta) \equiv 1 - c_i(\theta)$ be firm i's value of selling to the entire market at a price of 1 when the shock is θ, and let $\mathcal{X} = \left\{\chi \in \mathbb{R}_+^N : \sum_i \chi_i = 1\right\}$ be the set of possible market share allocations. Efficient collusion in this environment means that, in each period, whichever firm (or group of firms, if there is a tie) has the lowest cost should set a price of 1 and receive a market share of 1, while all other firms receive market shares of 0 (such as by setting prices greater than 1).

2.1 Collusion under perfect public equilibrium

Suppose that ϕ is common knowledge at the start of each period, and θ_i is the only new information that each firm i observes before making its announcement to the others. Then average discounted cartel profits under PPE are bounded above by the value of the following mechanism design problem (as shown by Lemma 1, in Appendix B):

$$\begin{aligned} V^{\max} \equiv \max_{\langle x,y\rangle:\Theta\to\mathcal{X}\times\mathbb{R}^N} \; & \mathbb{E}\sum_i \left[w_i(\vartheta)x_i(\vartheta) + y_i(\vartheta)\right] \text{ subject to} \\ & \text{No subsidy: } \sum_i y_i(\theta) \le 0 \text{ for all } \theta, \\ & \text{IIC: } \theta_i \in \arg\max_{\hat\theta_i \in \Theta_i} \mathbb{E}\left[w_i(\theta_i,\vartheta_{-i})x_i(\hat\theta_i,\vartheta_{-i}) + y_i(\hat\theta_i,\vartheta_{-i}) \middle| \theta_i\right] \text{ for all } \theta_i \text{ and all } i, \end{aligned} \tag{1}$$

where ϑ is the random variable of which θ is a particular realization. By the revelation principle, it is without loss of generality to restrict attention to direct revelation mechanisms in which the firms literally report their signals. In the *mechanism* $\langle x, y\rangle$, the *allocation rule* $x : \Theta \to \mathcal{X}$ assigns each firm a market share as a function of the firms' reports, while the *transfer rule* $y : \Theta \to \mathbb{R}^N$ assigns the monetary transfers each firm should receive. The no-subsidy condition embodies the constraint that the cartel cannot bring in money from outside the game. So following any θ for which $\sum_i y_i(\theta) < 0$, the firms should initiate a price war. The *interim incentive compatibility* (IIC) constraint says that each firm i must be willing to report its signal truthfully, given its expectations conditioned on its own private signal θ_i.

If the firms are sufficiently patient then there exists a PPE that actually attains $V^{\max}$. Along the equilibrium path, the cartel simply employs the mechanism that solves Eq. 1 in every period. IIC discourages each firm from lying about its signal, and every other available deviation—such as undercutting the other firms' prices or reneging on equilibrium transfers—is observable. Following an observable deviation, the firms can switch to a punishment path in which they do not communicate, and set their prices as if bidding non-cooperatively in a first-price procurement auction. Such bidding behavior forms a perfect Bayesian equilibrium in the stage game, and yields low profits. If the firms are sufficiently patient then the threat of incurring punishment suffices to discourage all observable deviations.

Under appropriate conditions, when the cartel attains $V^{\max}$, it attains the same profits that would be earned by a monopolist who owned all the firms' production processes. For instance, the cartel attains monopoly profits if the following assumption is satisfied:

Assumption 1 (Monotonicity and Regularity)**.** *There exists $B \in (1, \infty)$ such that, for all θ, all $j \neq i$, and all i*

(i) $w_i(\theta)$ *is continuously differentiable with* $0 \leq \frac{\partial w_j(\theta)}{\partial \theta_i} < \frac{\partial w_i(\theta)}{\partial \theta_i} - \frac{1}{B} < B - \frac{1}{B}$*;*

(ii) $w_i(0, \ldots, 0) = 0$*;*

(iii) ϕ *is a continuously differentiable probability density with* $\frac{1}{B} < \phi(\theta) < B$*.*

Assumption 1 is maintained throughout the remainder of Section 2. The second and third parts of the assumption are regularity conditions that help eliminate distracting anomalies. The substantive part of the assumption is the first part, which ensures that firm i's market share in an efficient allocation is weakly increasing in its own signal θ_i. By Theorem 2 of Chung and Ely (2006), for any such allocation rule there exists a mechanism that implements it under IIC. Furthermore, under fairly broad conditions such an allocation rule can be implemented under IIC without any money burning. For instance, if the firms' signals are statistically independent, then for any efficient allocation rule x and an IIC mechanism $\langle x, y \rangle$, the alternative mechanism $\langle x, \hat{y} \rangle$, where $\hat{y}_i(\theta) = \mathbb{E}\big[y_i(\theta)\big|\theta_i\big] - \frac{1}{N-1}\sum_{j \neq i} \mathbb{E}\big[y_j(\theta)\big|\theta_j\big]$, satisfies IIC without money burning, because the summation term does not depend on firm i's announcement.[7] Hence, under Assumption 1 and independence, a patient, optimizing cartel should not engage in inefficient behavior such as price wars. Even if the cartel could not transfer money, the folk theorem would typically still apply, and any inefficient behavior should typically vanish in the limit as $\delta \to 1$.[8]

However, PPEs in general are not robust to the introduction of payoff-irrelevant signals, since such signals generate complex higher-order beliefs. Specifically, suppose that in each period each firm i observes not only θ_i, but also a *payoff-irrelevant* signal $\omega_i \in \Omega_i$; let $\Omega \equiv \Omega_1 \times \cdots \Omega_N$, and suppose that ψ is the joint distribution on $\Theta \times \Omega$. The signal vector $\omega \in \Omega$ is payoff-irrelevant because it does not enter the objective function. Of course, it is relevant to the incentives that each firm faces, because it affects the firm's beliefs about the other firms' beliefs and payoff-relevant

[7] If instead signals are correlated then firm i can influence the summation term, so a different approach is required. For $N \geq 3$ and Θ finite, d'Aspremont, Crémer, and Gérard-Varet (2004) show that efficiency is IIC-implementable without money burning for generic probability distributions, regardless of the cost functions. For a setting without transfers, Aoyagi (2007) provides more complicated sufficient conditions for the case of $N = 2$ with continuous signal spaces; these conditions apply jointly to the probability distribution ϕ and the cost functions.

[8] See Section 2.3 for further discussion of this point.

signals. In such an environment the value of a PPE is bounded above by the value of a "full" version of Eq. 1 that accounts for these incentive issues:

$$
\begin{aligned}
V^{\text{full}} \equiv \max_{\langle x,y\rangle:\Theta\times\Omega\to\mathcal{X}\times\mathbb{R}^N} \; & \mathbb{E}\sum_i \big[w_i(\vartheta)x_i(\vartheta,\varpi)+y_i(\vartheta,\varpi)\big] \text{ subject to} \\
& \text{No subsidy: } \sum_i y_i(\theta,\omega)\le 0 \text{ for all } \theta \text{ and all } \omega, \\
& \text{Full IIC: } (\theta_i,\omega_i)\in \operatorname*{arg\,max}_{(\hat\theta_i,\hat\omega_i)\in\Theta_i\times\Omega_i} \mathbb{E}\left[\begin{array}{l} w_i(\theta_i,\vartheta_{-i})x_i\big((\hat\theta_i,\hat\omega_i),(\theta_{-i},\varpi_{-i})\big) \\ +\, y_i\big((\hat\theta_i,\hat\omega_i),(\theta_{-i},\varpi_{-i})\big)\end{array}\middle|(\theta_i,\omega_i)\right] \\
& \qquad \text{for all } \theta_i, \text{ all } \omega_i, \text{ and all } i,
\end{aligned}
\tag{2}
$$

where ϖ is the random variable of which ω is a realization. Full IIC is a much more demanding requirement than IIC, as it depends on the full set of payoff-irrelevant details, and beliefs about these details. Realistically, the details of the payoff-irrelevant information environment are likely to be complex—all possible channels of information arrival have to be modeled, along with all possible interactions among them. Full IIC can be reduced to ordinary IIC (i.e., $V^{\text{full}} = V^{\max}$) only if, for each firm i and conditional on any payoff-relevant private signal realization θ_i, its payoff-irrelevant signal ϖ_i is statistically independent of the other firms' payoff-relevant signals ϑ_{-i}. For tractability the prior literature on collusive PPEs implicitly assumes this to be the case.

2.2 Robust collusion

A mechanism satisfies *ex post incentive compatibility* (EPIC) if each firm would still be willing to announce truthfully after observing all the other firms' signals. In principle, this truthful revelation requirement applies to both payoff-relevant and payoff-irrelevant signals. However, if the mechanism does not depend on the payoff-irrelevant signals, then each firm is willing to truthfully announce its payoff-irrelevant signal. So if the allocation rule x does not depend on the payoff-irrelevant signals, an EPIC mechanism can ignore the payoff-irrelevant details of the environment.[9] A PPE that is robust to arbitrary payoff-irrelevant details must satisfy EPIC in every period, taking expectations over the future path of play. I call such an equilibrium an *ex post perfect public equilibrium* (EPPPE). Theorem 3, in Section 3, implies that cartel profits under EPPPE are bounded

[9] Bergemann and Morris (2005) show that a limited converse is also true: if Ω is the universal type space generated by the payoff-relevant signal space Θ and ψ has full support on $\Theta\times\Omega$, then a mechanism satisfies EPIC if (i) it satisfies Full IIC (see Eq. 2) and (ii) its allocation rule does not depend on the payoff-irrelevant signals. This converse holds for quasilinear environments with money-burning, as studied here. It remains an open question how strong a partial converse can be stated if the allocation rule is allowed to depend on payoff-irrelevant signals. Chung and Ely (2007) show that there are reasonable conditions under which the maxmin-optimal auction mechanism satisfying Full IIC must also satisfy EPIC, but their conditions do not address the environment considered here. Regardless, a mechanism in which the allocation rule depends on the payoff-irrelevant signals will generally not attain efficiency.

above by the value of the following mechanism design problem:

$$\begin{aligned} V^* \equiv \max_{\langle x,y\rangle:\Theta\to\mathcal{X}\times\mathbb{R}^N} \; & \mathbb{E}\sum_i \big[w_i(\vartheta)x_i(\vartheta)+y_i(\vartheta)\big] \text{ subject to} \\ & \text{No subsidy: } \sum_i y_i(\theta) \le 0 \text{ for all } \theta, \\ & \text{EPIC: } \theta_i \in \arg\max_{\hat\theta_i\in\Theta_i} w_i(\theta)x_i(\hat\theta_i,\theta_{-i}) + y_i(\hat\theta_i,\theta_{-i}) \text{ for all } \theta \text{ and all } i. \end{aligned} \tag{3}$$

I say that a mechanism $\langle x, y\rangle$ is *optimal* if it solves this problem. The cartel can attain this bound if it is sufficiently patient.[10]

If the firms have private values (such that $w_i(\theta)$ can be written as $w_i(\theta_i)$ for all i), the logic of Groves (1973) mechanisms directly implies that they cannot attain monopoly profits under EPPPE. Under private values, any EPIC mechanism $\langle x, y\rangle$ that allocates efficiently is Groves a mechanism. So the cartel runs a second-price "knockout auction" in which the low bidder earns the right to serve the entire market at a price of 1, and pays the bid of the second lowest bidder. Let $R(\theta; x)$ be the "revenue" from this auction; it is equal to the bid of the second lowest bidder. The cartel would like to give this revenue back to the member firms as much as possible without disrupting incentives, through a "rebate" function $h_i : \Theta_{-i} \to \mathbb{R}$ for each firm i. To attain monopoly profits the cartel would have to fully rebate the revenue back to the firms; i.e., $\sum_i y_i(\theta) = -R(\theta; x) + \sum_i h_i(\theta_{-i}) = 0$ for all θ. This is often called the "ex post budget balance" requirement. Unfortunately for the cartel, any rebate functions that could attain ex post budget balance would disrupt the incentives for the firms to tell the truth about their costs. Hence from time to time the cartel will have unrebatable revenue that must be burned, such as by holding a price war. Furthermore, by Theorem 1, below, the same conclusion holds when values are interdependent (i.e., not private). Since efficient allocation leads to price wars, monopoly profits are unattainable. This is illustrated in the following example.

Example 1 (Private values, efficient allocation)**.** Suppose that $N = 2$, $w_i(\theta) = \theta_i$ for all i, and θ is distributed according to the uniform distribution on $[0, 1]^2$.

In a mechanism that allocates efficiently, transfers that minimize the expected cost of price wars subject to EPIC are (see Lemma 5, in Appendix C):

$$y_i(\theta) = -x_i(\theta)\theta_{-i} + \min\big\{\theta_{-i}, \tfrac{1}{2}\big\} - \tfrac{1}{4}. \tag{4}$$

[10] On the equilibrium path, it simply employs the mechanism that solves Eq. 3. Off-the-equilibrium-path behavior, however, is more subtle than under PPE. The firms cannot simply play the no-communication, no-transfers stage game Bayesian-Nash equilibrium, since it violates EPIC. However, the firms can replicate the same expected profits under EPIC by treating the consumers as the auctioneer in an appropriate generalized Groves mechanism (e.g., a second-price procurement auction in the special case of private values), where the winner's "payment" accrues to the consumers in the form of a price discount.

This mechanism is illustrated in Figure 1. The first term in Eq. 4, $-x_i(\theta)\theta_{-i}$, is firm i's payment in a second-price auction. The remaining terms do not depend on θ_i, so this is a Groves (1973) mechanism. These remaining terms attempt to rebate the revenues of the second-price auction—sometimes fully, sometimes partially—to the firms as much as possible without disrupting incentive compatibility. When $\theta_i > \frac{1}{2} > \theta_{-i}$, the winner receives a rebate of its auction payment minus $\frac{1}{4}$, while the loser receives a rebate of $\frac{1}{4}$, so there is no price war. In contrast, price wars arise when both firms' values are greater than $\frac{1}{2}$: both the winner and the loser receive rebates of $\frac{1}{4}$, which add up to less than the winner's auction payment. Similarly, price wars arise when both firms' values are less than $\frac{1}{2}$: the winner receives a rebate of its auction payment minus $\frac{1}{4}$, while the loser receives a rebate of less than $\frac{1}{4}$. In these regions the revenues depend on both θ_1 and θ_2, and therefore cannot be fully rebated back to the firms without disrupting their incentives.

When the firms initiate a price war, its severity depends on how much money must be burned. For example, when $\theta = (1,1)$, they must burn $\left|\sum_i y_i(1,1)\right| = \frac{1}{2}$, which they can accomplish by setting the price to $\frac{1}{2}$ rather than 1 for the current period. That is, they can burn all the required money immediately by selling to the consumers at a discount. The value of this mechanism is $\frac{7}{12}$, reflecting the value of efficient allocation, $\frac{2}{3}$, minus the expected cost of price wars, $\frac{1}{12}$.

2.2.1 Optimally inefficient allocation

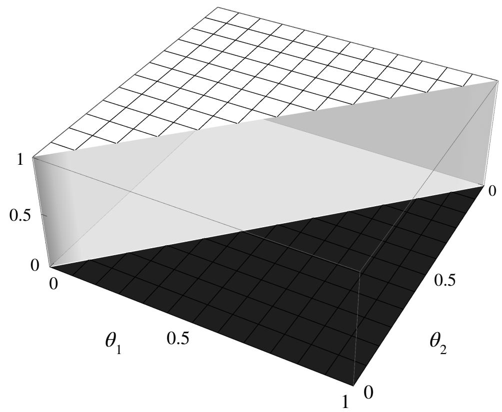
Since allocating efficiently leads to price wars, a natural question arises: When price wars are taken into account, is it actually optimal to allocate efficiently? The first main result of this paper, Theorem 1, shows that the answer is "no": The cartel should give up some allocative efficiency in order to reduce the cost of price wars.

Theorem 1. *Under Assumption 1, a two-firm cartel optimally allocates inefficiently.*

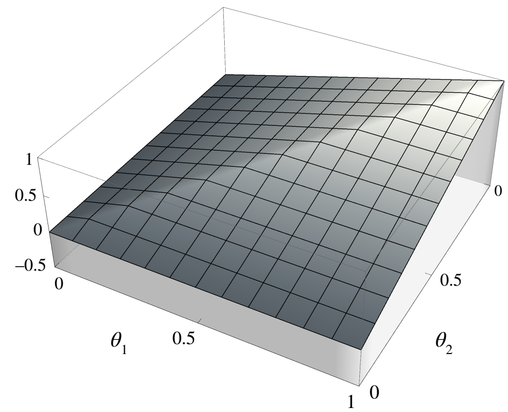
Starting from an efficient allocation rule, the proof constructs a modified allocation rule $\hat{x}$ that allocates efficiently except when θ falls within a small rectangular region $E \subset \Theta$ close to the origin. On E the modified rule always assigns market share to the same firm. This rule still gives each firm a market share that is monotonic in its own signal, so the modified rule is still implementable under EPIC. When $\theta \in E$ the severity of the price wars can be reduced, since no incentive payments are needed. For $\theta \notin E$, price wars are no worse than those that would occur under an efficient mechanism. If E is sufficiently small, the gain from reducing the severity of price wars is guaranteed to be greater than the loss in allocative efficiency. The formal proof is in Appendix D; its intuition is illustrated in the following example.

Example 2 (Private values, inefficient allocation)**.** Suppose that $N = 2$, $w_i(\theta) = \theta_i$ for all i, and θ is distributed according to the uniform distribution on $[0,1]^2$, as in Example 1.

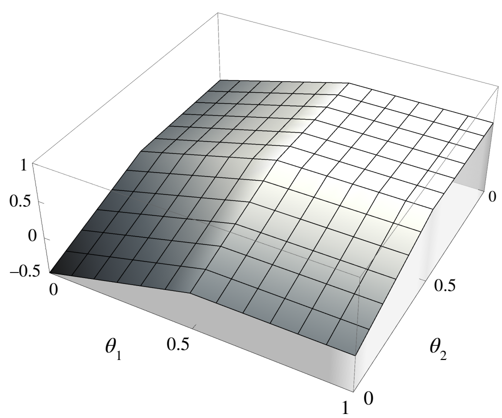
Consider the following slight modification to the mechanism from Example 1, for small $\varepsilon > 0$:



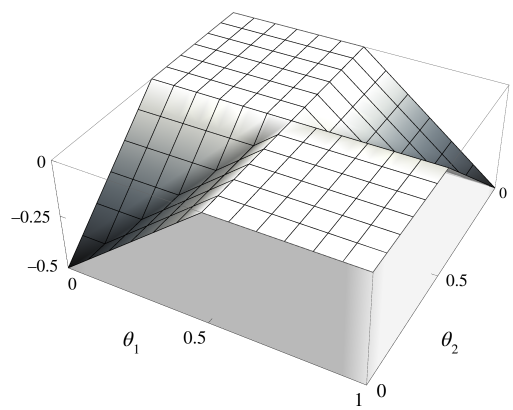
(a) Allocation rule $x_1(\theta)$



(b) Revenues $R(\theta; x)$



(c) Rebates $\sum_i h_i(\theta_{-i})$



(d) Aggregate transfers $\sum_i y_i(\theta)$

Figure 1. Efficient mechanism for Example 1 In this example, each firm's value is $w_i(\theta) = \theta_i$, and each firm's signal θ_i is distributed uniformly on the interval $[0, 1]$. Panel (a) displays the efficient allocation rule, expressed as firm 1's market share $x_1^*(\theta)$; firm 2's market share is $1 - x_1^*(\theta)$. Panel (b) displays the revenues that arise from a second-price knockout auction, $R(\theta; x) = \min_i \theta_i$. Panel (c) displays the rebates that are optimal conditional on the allocation rule, where each firm receives a rebate of $h_i(\theta_{-i}) = \min\{\theta_{-i}, \frac{1}{2}\} - \frac{1}{4}$. Panel (d) displays the aggregate transfers $\sum_i y_i(\theta) = -R(\theta; x) + \sum_i h_i(\theta_{-i})$. Money is burned, via price wars, whenever $\sum_i y_i(\theta) < 0$.

- Allocate to firm 1 if $\max_i \theta_i < \varepsilon$, and otherwise allocate efficiently;
- The firms receive the following transfers:

$$y_1(\theta) = -x_1(\theta)\mathbb{I}(\theta_2 \geq \varepsilon)\theta_2 + \mathbb{I}(\theta_2 \geq \varepsilon)\min\{\theta_2, \tfrac{1}{2}\} - \tfrac{1}{4}, \tag{5}$$

$$y_2(\theta) = -x_2(\theta)\max\{\theta_1, \varepsilon\} + \min\{\max\{\theta_1, \varepsilon\}, \tfrac{1}{2}\} - \tfrac{1}{4}. \tag{6}$$

This mechanism is illustrated in Figure 2. For each firm, the first term in its transfer provides EPIC incentives, while the remaining terms do not depend on its own signal. Observe that whenever $\max_i \theta_i > \varepsilon$, this mechanism yields price wars identical to those in Example 1. The advantage of this mechanism is that, conditional on $\max_i \theta_i < \varepsilon$, the allocation does not depend on the firms' values. Within this region, no monetary incentives are needed, so the price wars can be less severe. Of course, the disadvantage is that sometimes the allocation is inefficient. Compared to the mechanism in Example 1, this mechanism loses $\frac{1}{6}\varepsilon^3$ to by allocating inefficiently, but gains $\frac{1}{2}\varepsilon^3$ by reducing the severity of price wars, and hence represents an improvement for the cartel.

Theorem 1 implies that in constructing an optimal EPPPE the allocation rule and the transfer rule must be optimized simultaneously. This contrasts with PPE, for which, under quite broad conditions, it suffices to first select an efficient allocation rule and only then design transfers to implement it. Accordingly, computing an optimal EPPPE is more difficult than computing an optimal PPE. This is exemplified by the fragility of the following, seemingly simple, example.

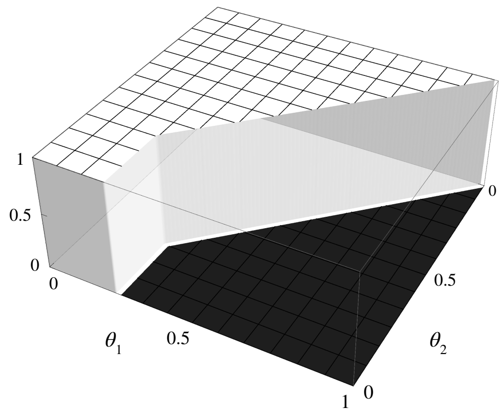
Example 3 (Private values optimal mechanism)**.** Suppose that $N = 2$, $w_i(\theta) = \theta_i$ for all i, and θ is distributed according to the uniform distribution on $[0, 1]^2$, as in Examples 1–2.

An optimal mechanism, illustrated in Figure 3, is as follows (where $\mathbb{I}$ is the indicator function that takes the value 1 if its argument is true and takes the value 0 otherwise):

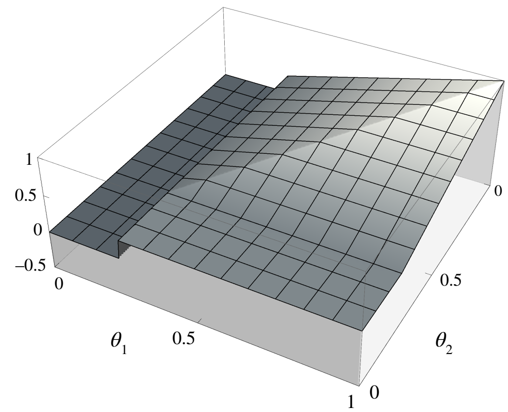
- Allocate efficiently if $\max_i \theta_i \geq \frac{1}{2} > \min_i \theta_i$, and otherwise split the market equally;
- Each firm receives a transfer $y_i(\theta) = -\frac{1}{4}\mathbb{I}(\theta_i \geq \frac{1}{2}) + \frac{1}{4}\mathbb{I}(\theta_{-i} \geq \frac{1}{2})$.

This mechanism takes the intuition of Example 2 to the extreme: the allocation is fixed whenever $\max_i \theta_i < \frac{1}{2}$ or $\min_i \theta_i > \frac{1}{2}$, so no monetary incentives are needed over large swaths of the signal space. Indeed, this mechanism allocates so inefficiently that price wars are eliminated entirely. It can be implemented by allocating "property rights" over the market to each firm with equal probability, and allowing them to trade at a posted price of $\frac{1}{2}$. The value to the cartel is $\frac{5}{8}$, better than the $\frac{7}{12}$ attained in Example 1.

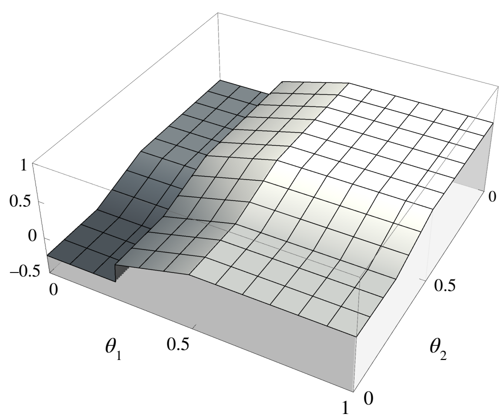
Shao and Zhou (2008) prove that the mechanism described in Example 3 is optimal. However, their proof relies on the specific properties of the uniform distribution. The following example shows that price wars can arise even under private values.



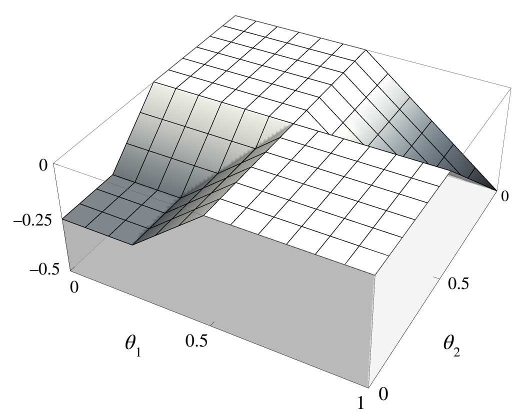
(a) Allocation rule $x_1(\theta)$



(b) Revenues $R(\theta; x)$

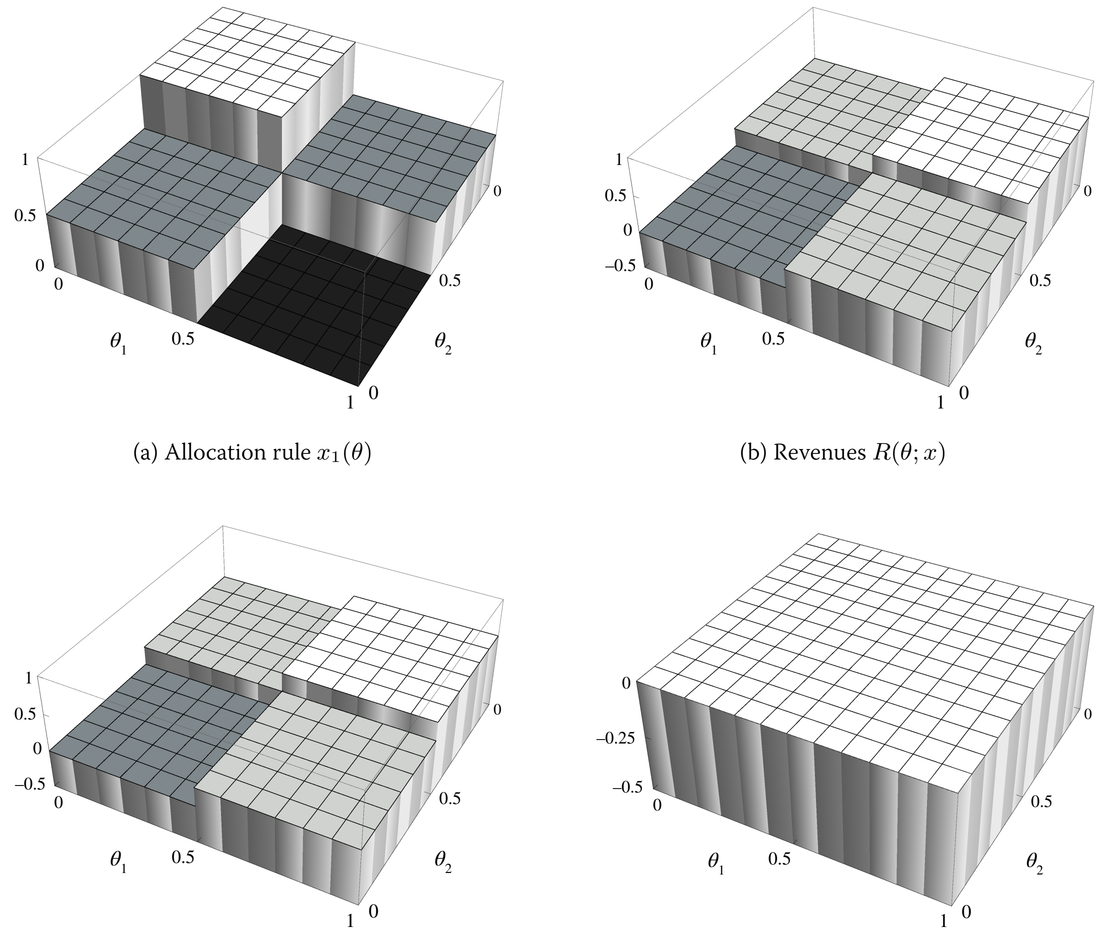


(c) Rebates $\sum_i h_i(\theta_{-i})$



(d) Aggregate transfers $\sum_i y_i(\theta)$

Figure 2. Inefficient mechanism for Example 2 In this example, each firm's value is $w_i(\theta) = \theta_i$, and each firm's signal θ_i is distributed uniformly on the interval $[0, 1]$. Panel (a) displays the inefficient allocation rule for $\varepsilon = \frac{1}{4}$. The allocation is expressed as firm 1's market share $x_1^*(\theta)$; firm 2's market share is $1 - x_1^*(\theta)$. Panel (b) displays the revenues that arise from an inefficient auction with this allocation rule, $R(\theta; x) = \min_i \theta_i$. Panel (c) displays the rebates that are optimal conditional on the allocation rule, where each firm receives a rebate of $h_i(\theta_{-i}) = \min\{\theta_{-i}, \frac{1}{2}\} - \frac{1}{4}$. Panel (d) displays the aggregate transfers $\sum_i y_i(\theta) = -R(\theta; x) + \sum_i h_i(\theta_{-i})$. Price wars are less severe than in Example 1.



(a) Allocation rule $x_1(\theta)$

(b) Revenues $R(\theta; x)$

(c) Rebates $\sum_i h_i(\theta_{-i})$

(d) Aggregate transfers $\sum_i y_i(\theta)$

FIGURE 3. OPTIMAL MECHANISM FOR EXAMPLE 3 Suppose that $N = 2$, each firm's value is $w_i(\theta) = \theta_i$, and each firm's signal θ_i is distributed uniformly on the interval $[0, 1]$. Panel (a) displays the optimal allocation rule, expressed as firm 1's market share $x_1(\theta)$; firm 2's market share is $1 - x_1(\theta)$. Panel (b) displays the revenues that arise from a posted price mechanism with random property rights, $R(\theta; x) = \sum_i \frac{1}{4}\mathbb{I}(\theta_i \geq \frac{1}{2})$. Panel (c) displays the rebates that are optimal conditional on the allocation rule, where each firm receives a rebate of $h_i(\theta_{-i}) = \frac{1}{4}\mathbb{I}(\theta_{-i} \geq \frac{1}{2})$. Panel (d) displays the aggregate transfers $\sum_i y_i(\theta) = -R(\theta; x) + \sum_i h_i(\theta_{-i})$. Since the transfers always sum to zero, there are no price wars.

Example 4 (Private values with price wars)**.** Suppose that $N = 2$, $w_i(\theta) = \theta_i$ for all i, and each θ_i is independently distributed on $[0, 1]$ according to a step distribution with a density of $\frac{2}{5}$ for $\theta_i < \frac{1}{2}$, and a density of $\frac{8}{5}$ otherwise. An approximate optimal mechanism is illustrated in Figure 4. This mechanism was computed numerically, following the linear programming approach outlined in Lemma 4. Since the transfers do not always sum to zero, price wars arise for many realizations of θ.

Price wars in this example arise because the probability distribution places relatively high probability on regions of Θ where the allocation would be inefficient under the mechanism in Example 3. The allocation rule illustrated in Figure 4 allocates more efficiently, particularly in regions that arise with high probability. At the same time, the transfers illustrated in Figure 4 imply price wars mainly in regions that arise with low probability. Taking expectations, this mechanism with price wars is superior to any mechanism that would eliminate price wars.

2.2.2 Optimal price wars

Example 3 shows that price wars may not arise in robust equilibria under private values, although inefficient allocation is pervasive. However, the following theorem shows that it is optimal for a two-firm cartel to engage in price wars if the firms' costs are interdependent. Interdependent costs can be interpreted as representing an industry with underlying common shocks as well as idiosyncratic shocks, in which the shocks are observed imperfectly and privately, and such that the uncertainty is not resolved until after the communication stage. Since the cartel in this model produces an undifferentiated product, underlying common shocks can arise from shocks to the markets for inputs. If each firm has relationships with different suppliers, the underlying common shock could easily be confounded with their suppliers' idiosyncratic shocks.

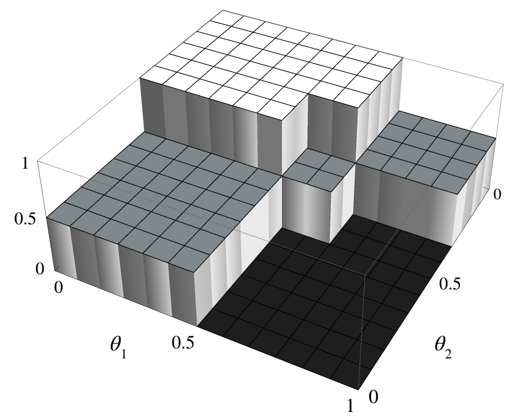
When the firms' valuations are interdependent, the transfers they must make to implement all but the most degenerate allocation rules depend on the fine details of the interdependence. It is typically impossible to rebate such payments back to the firms without disrupting their incentives.

Assumption 2 (Global monotonic interdependence)**.** *There exists $B \in (1, \infty)$ such that $\frac{\partial w_i}{\partial \theta_{-i}} > \frac{1}{B}$ for all θ and all i.*

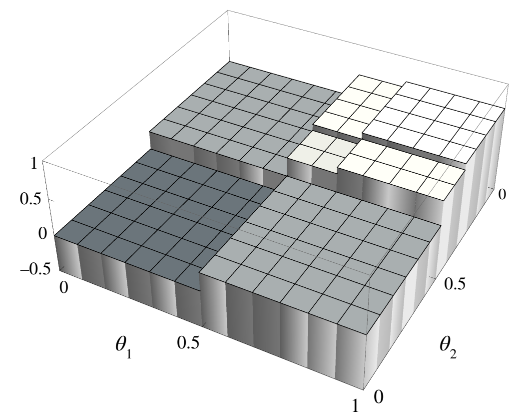
Theorem 2. *Under Assumption 1 and Assumption 2, generically*[11] *an optimal EPPPE involves price wars on the equilibrium path.*

The proof, in Appendix C, begins with Lemma 7, which shows that achieving ex post budget balance under globally interdependent valuations requires the allocation rule to satisfy a partial differential equation. There are two classes of solutions to this PDE: any constant outcome rule, and a particular set of differentiable non-constant outcome rules. However, any solution in either

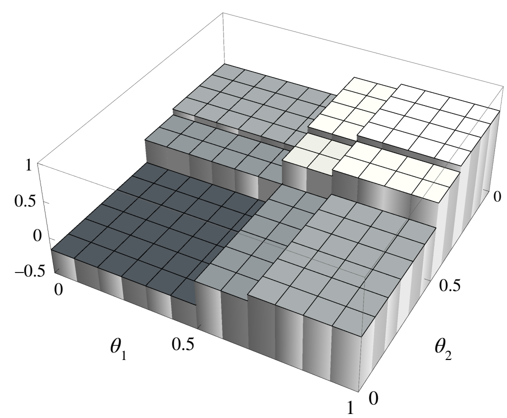
[11] The proof fails if w takes a particular non-generic form along the curve $w_1(\theta) = w_2(\theta)$.



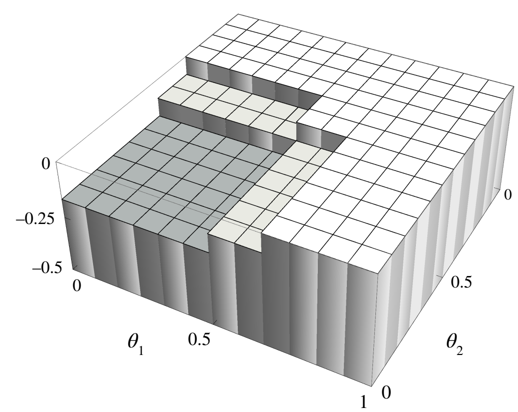
(a) Allocation rule $x_1(\theta)$



(b) Revenues $R(\theta; x)$



(c) Rebates $\sum_i h_i(\theta_{-i})$



(d) Aggregate transfers $\sum_i y_i(\theta)$

Figure 4. Approximate optimal mechanism for Example 4 In this example, each firm's value is $w_i(\theta) = \theta_i$, and each firm's signal θ_i is distributed uniformly on the interval $[0, 1]$. Panel (a) displays the optimal allocation rule, expressed as firm 1's market share $x_1(\theta)$; firm 2's market share is $1 - x_1(\theta)$. Panel (b) displays the revenues that arise from the firms' incentive payments, $R(\theta; x)$. Panel (c) displays the rebates that are optimal conditional on the allocation rule, where each firm receives a rebate of $h_i(\theta_{-i})$. Panel (d) displays the aggregate transfers $\sum_i y_i(\theta) = -R(\theta; x) + \sum_i h_i(\theta_{-i})$. Since the rebates do not match the revenues, price wars arise following some realizations of θ. This mechanism was computed numerically using linear programming, where the signal space is discretized to the grid shown in the graphs.

of these classes can be improved upon by allocating market share to each firm i when θ_i is very high and θ_{-i} is very low, even though doing so requires some price wars. That is, under interdependence, eliminating price wars entirely requires an extremely inefficient allocation rule. The players are better off salvaging some efficiency even at the expense of engaging in price wars. This is illustrated in Example 5.

Example 5. Suppose that $N = 2$, $w_i(\theta) = \frac{3}{4}\theta_i + \frac{1}{4}\theta_{-i}$ for all i, and θ is distributed according to the uniform distribution on $[0, 1]^2$. An approximate optimal mechanism is illustrated in Figure 5. This mechanism was computed numerically, following the linear programming approach outlined in Lemma 4. Since the transfers in this mechanism do not always sum to zero, price wars arise. The allocation rule takes the form of a smoothed-over version of the posted-price mechanism in Example 3, and manages to keep the price wars quite mild (relative to Example 1, for instance).

2.3 Generalizations

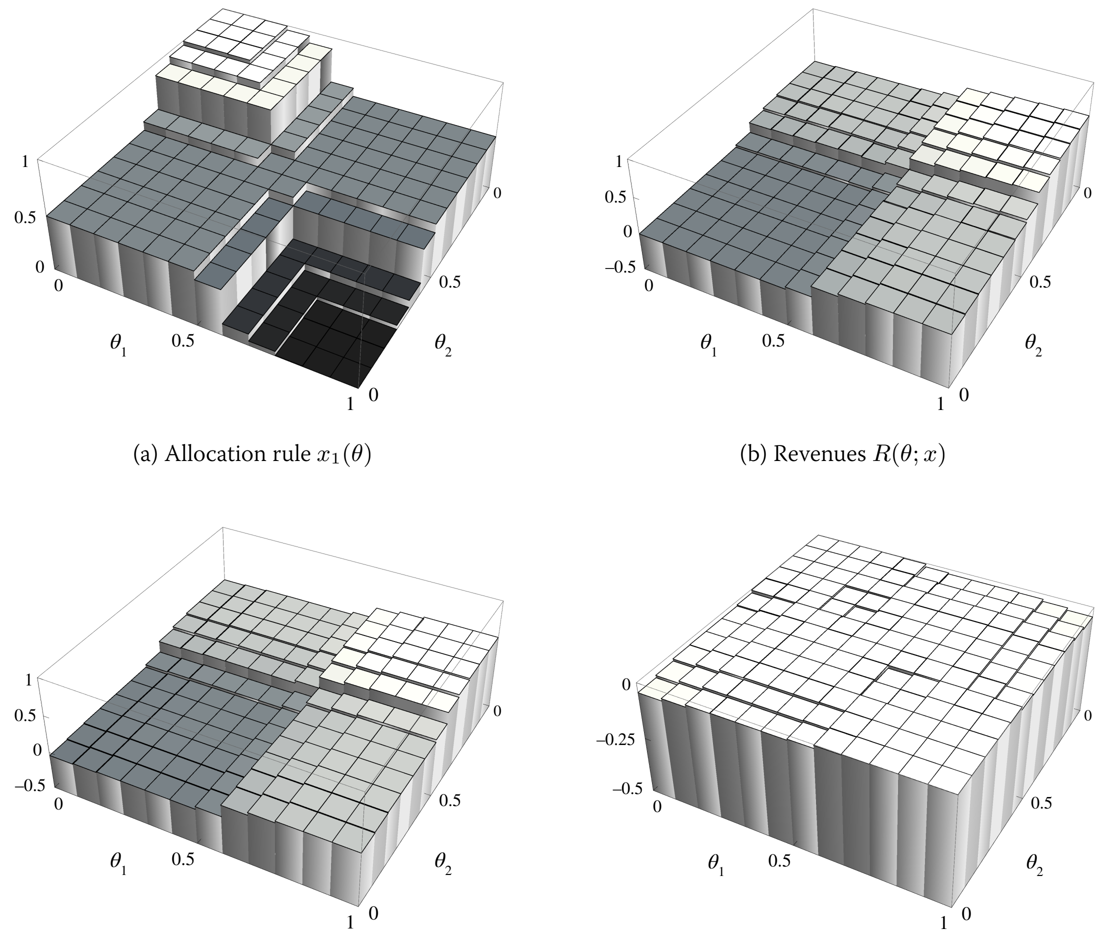
2.3.1 No monetary transfers

The assumption that the firms can transfer money is convenient, but not necessarily realistic in environments with antitrust enforcement. If the firms cannot make transfers, then they must seek other ways to transfer utility. For example, they can play a non-stationary equilibrium in which changes in continuation play substitute for transfers. The cartel can construct such equilibria using the techniques of Fudenberg and Levine (1994). Specifically, for each direction $\xi \in \mathbb{R}^N$, the cartel must identify the mechanism $\langle x^\xi, y^\xi \rangle$ that solves a ξ-weighted version of Eq. 3. Each $\langle x^\xi, y^\xi \rangle$ defines a half-space normal to ξ that attains the ξ-weighted value of $\langle x^\xi, y^\xi \rangle$. The intersection of all such half spaces is the limiting set of utilities attainable under EPPPE as $\delta \to 1$. Naturally, this attainable set is a subset of the particular half-space associated with Eq. 3 (unweighted); whether the boundary of this set attains the value of Eq. 3 may depend on the details of the game. That is, the cartel's best-case scenario is to be able to make monetary transfers, but even if it cannot make monetary transfers it may be able to approach the same value as $\delta \to 1$.

2.3.2 Private strategies

Kandori and Obara (2006) show that equilibria in private strategies can outperform PPE in games with imperfect monitoring. The intuition underlying their results is that by performing a private experiment, a firm can observe a more informative signal of whether its cartel partners are cooperating. More informative signals enable better-targeted punishments, which can be enacted less often while still providing strong incentives.

In collusion with private information, however, the only possible "private experiment" a firm can perform is to misreport its own signal. Doing so does not affect its conditional beliefs about



(a) Allocation rule $x_1(\theta)$

(b) Revenues $R(\theta; x)$

(c) Rebates $\sum_i h_i(\theta_{-i})$

(d) Aggregate transfers $\sum_i y_i(\theta)$

Figure 5. Approximate optimal mechanism for Example 5 In this example, each firm's value is $w_i(\theta) = \theta_i$, and each firm's signal θ_i is distributed uniformly on the interval $[0, 1]$. Panel (a) displays the optimal allocation rule, expressed as firm 1's market share $x_1(\theta)$; firm 2's market share is $1 - x_1(\theta)$. Panel (b) displays the revenues that arise from the firms' incentive payments, $R(\theta; x)$. Panel (c) displays the rebates that are optimal conditional on the allocation rule, where each firm receives a rebate of $h_i(\theta_{-i})$. Panel (d) displays the aggregate transfers $\sum_i y_i(\theta) = -R(\theta; x) + \sum_i h_i(\theta_{-i})$. Although the rebates resemble the revenues, they do not precisely match. Therefore, for many realizations of θ, a non-negligible price war arises. This mechanism was computed numerically using linear programming, where the signal space is discretized to the grid shown in the graphs.

the other firms' signals, and therefore cannot aid in detecting their deviations. So the Kandori and Obara critique of PPE does not apply to games with private information and communication.

Moreover, EPIC requires each firm to be willing to send its equilibrium message regardless of the information on which the other firms are conditioning their messages, including their private histories. Private histories are like payoff-irrelevant information—they generate potentially complex higher order beliefs over the payoff-relevant signals. Hence EPIC rules out any non-trivial use of private strategies.[12]

2.3.3 Imperfect monitoring

This paper focuses on private information about costs, but realistically the problem of monitoring the firms' pricing behavior may also be important. In general, the approach of the literature has been to model these issues separately, with the implicit conjecture that there should be no worrisome synergistic interactions when the two problems are combined in the same model. Intuitively, the stage game with both private cost information and imperfect public monitoring of prices can be split into a first phase, in which private signals are realized and then the firms communicate with each other, and a second phase, in which actions are taken privately and then a public monitoring signal is realized. The mechanism design approach outlined here applies to the first phase, and the approach of Fudenberg, Levine, and Maskin (1994) applies to the second phase.

Specifically, each firm's EPIC constraint in the first phase applies with expectations taken over not only future periods but also over the realization of the monitoring signal. If the firms have revealed their cost signals truthfully in the first communication phase, then, following the realization of the monitoring signal, additional transfers can be specified to enforce the desired pricing actions. Under the Fudenberg, Levine, and Maskin (1994) identifiability conditions, such transfers can be budget balanced.

However, there remains the possibility that a firm could benefit from lying about its own cost signal in the first phase in order to trick the other firms into having incorrect beliefs about the distribution of the monitoring signal, and then selecting a deviant price in the second phase. Whether this additional incentive constraint is satisfied depends on the details of the monitoring distribution. Without constraining these details, the most that can be inferred is that the value of an EPPPE with perfect monitoring is an upper bound on the value of an EPPPE in an otherwise equivalent game with imperfect monitoring.[13]

[12]For expositional and notational convenience, this paper defines an EPPPE as a PPE constrained to satisfy EPIC. Pursuant to this discussion of private strategies, however, an EPPPE could be defined as a perfect extended-Bayesian equilibrium (Battigalli 1996, Fudenberg and Tirole 1991) that satisfies a more general notion of EPIC. (Note that "sequential equilibrium" is not well defined for games with continuum strategy spaces.)

[13]Because the payoff-relevant private cost information persists from the first phase to the second phase, this is a dynamic game. The closely related literature on dynamic mechanism design faces the same kind of difficulty, which is why Pavan, Segal, and Toikka (2009), for example, study Bayesian Nash implementation rather than perfect Bayesian implementation.

3 General results

The fact that robust collusive equilibria cannot attain efficiency does not arise from any special properties of the collusion games studied in Section 2. Instead, inefficiency arises generically in a broader class of repeated games. Theorem 3, below, shows that an optimal EPPPE is a solution to a mechanism design problem with a no-subsidy constraint. Theorem 4 provides necessary and sufficient conditions in the separable payoff environment (Chung and Ely 2006) for an EPPPE to attain efficiency, and Theorem 5 then shows that generically every EPPPE is inefficient.

Consider the following general environment:

- $\mathcal{N}$ is a finite set of players, $i = 1, \ldots, N$;
- $\delta \in (0, 1)$ is a common discount factor;
- Θ_i is the set of private signals for player i in the stage game, with $\Theta \equiv \Theta_1 \times \cdots \times \Theta_N$;
- ϕ is the common prior probability measure on Θ;
- $\mathcal{M}_i$ is the set of public messages that player i can send, with $\Theta_i \subset \mathcal{M}_i$ and $\mathcal{M} \equiv \mathcal{M}_1 \times \cdots \times \mathcal{M}_N$;
- $\mathcal{X}_i$ is the set of public actions for player i in the stage game, with $\mathcal{X} \equiv \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$;
- $\pi_i : \Theta \times \mathcal{X} \to \mathbb{R}_+$ is player i's uniformly bounded payoff function, with $\pi \equiv (\pi_1, \ldots, \pi_N)$;
- $\mathbb{T} \equiv \left\{t \in \mathbb{R}^N : \sum_i t_i \leq 0\right\}$ is the space of net monetary payments, where $t_i > 0$ indicates that player i receives a positive quantity of money.

Signals are drawn independently over time and from the same distribution ϕ, although there may be interdependence among the signals observed by different players within any particular stage. The timing of the stage game is as follows: First, $\theta \in \Theta$ is realized according to ϕ, and each player i privately observes θ_i. Next each player i sends a public announcement $m_i \in \mathcal{M}_i$. Then, each player i chooses a public action $\chi_i \in \mathcal{X}_i$.[14] Finally, the players receive a vector of monetary payments $t \in \mathbb{T}$.[15] A player's *utility* in the stage game is her payoff plus the payment she receives: $\pi_i(\theta, \chi) + t_i$.[16] In the repeated game, each player seeks to maximize her discounted sum of stage

[14] I restrict attention to pure strategy equilibria, but $\mathcal{X}$ itself may be viewed as a space of probability measures. For example, the space of correlated action profiles would be $\chi \in \mathcal{X} \equiv \Delta(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)$, where each player non-cooperatively selects an action $\chi_i \in \mathcal{X}_i$ after observing an arbitrary public randomization device.

[15] More operationally, each player can pay any non-negative amount to any other player. These bilateral payments, combined with money burning, must sum to a vector in $\mathbb{T}$.

[16] For simplicity, I assume that players cannot use their ex-post realized payoffs to make useful inferences about their opponents' true signals. One way to motivate this assumption is to assume that the discount factor reflects the probability that the game will terminate at the end of each period, and that players obtain their true payoffs only when the game terminates. In case the game cannot terminate, Mezzetti (2004) shows how to construct efficient EPIC

game utilities. In addition to collusion with private costs, several other important applications fall into this class, such as public goods, risk sharing, team production, externalities, and repeated trade.

A *perfect public equilibrium* (PPE) is a strategy profile in which, in each stage, each player conditions his action on only the public history and his current private signal, and the strategies form a Bayesian Nash equilibrium in the continuation game following any history. An *ex post perfect public equilibrium* (EPPPE) is a PPE that satisfies *ex post incentive compatibility* (EPIC, below) following every history, taking expectations over the future path of play. Appendix A defines both PPE and EPPPE formally. By the revelation principle, it is without loss of generality to restrict attention to equilibria in which players report their signals truthfully, so EPIC can be defined as follows:

Definition 1. A mechanism $\langle x, y\rangle : \Theta \to \mathcal{X} \times \mathbb{T}$ is *ex post incentive compatible (EPIC)* if

$$\theta_i \in \operatorname{argmax}_{\hat{\theta}_i \in \Theta_i} \big(\pi_i(\theta_i, \vartheta_{-i}, x(\hat{\theta}_i, \vartheta_{-i})) + y_i(\hat{\theta}_i, \vartheta_{-i})\big) \tag{7}$$

for all θ and for all i. An outcome rule $x : \Theta \to \mathcal{X}$ is *ex post implementable* if there exists an EPIC mechanism $\langle x, y\rangle$.

3.1 Bounding equilibrium utility

Theorem 3, below, shows that the players' aggregate utility in any EPPPE is bounded, regardless of how patient they are, by the value of the simple static mechanism design problem of maximizing aggregate utility subject to EPIC and a "no-subsidy" constraint:

$$V^* \equiv \sup_{\langle x,y\rangle:\Theta\to\mathcal{X}\times\mathbb{R}^N} \mathbb{E}\Big[\sum_i \big(\pi_i(\theta, x(\theta)) + y_i(\theta)\big)\Big] \text{ s.t. EPIC and } \sum_i y_i(\theta) \le 0 \text{ for all } \theta. \tag{8}$$

Under two further assumptions, this bound can actually be attained by sufficiently patient players. The first assumption is purely technical. It casts the action space $\mathcal{X}$ as the space of lotteries on a compact space of pure action profiles $\mathcal{A}$, such as would be implied by the existence of an arbitrary public randomization device.[17]

Assumption 3. *$\mathcal{X}$ is the space of probability measures on a compact space $\mathcal{A}$, and $\pi_i(\theta, \chi) = \int_{\mathcal{A}} \pi_i^a(\theta, a)\, d\chi(a)$, where $\pi_i^a(\theta, \cdot)$ is bounded and continuous on $\mathcal{A}$ for all θ and all i.*

mechanisms in a static setting with interdependent payoffs by adding a second communication phase in which the players announce their realized payoffs. However, such mechanisms cannot simultaneously satisfy EPIC and ex post budget balance, and so they could not be used to construct efficient EPPPEs even if there were a second communication phase at the end of each stage.

[17]Similar results could be proven for non-randomized mechanisms by imposing convexity on the space of outcomes and linearity on the $\pi_i(\theta, \cdot)$ functions.

The second assumption is more substantive: there must exist an ex post equilibrium in the stage game. By backward induction there can be no transfers in such an equilibrium. In some economic settings there exists an autarkic outcome that can serve this purpose, such as in a one-shot trade setting where the seller never gives up the object because the buyer cannot commit to pay for it. In the collusion games studied in Section 2, although no firm has intrinsic property rights over the market, the cartel can still construct a mechanism in which a selected firm serves the entire market, and sets a low price instead of paying a transfer to the other firms.

Assumption 4. *There exists an action rule $x : \Theta \to \mathcal{X}$ such that*

$$\pi_i(\theta, x(\theta)) \geq \pi_i\big(\theta, (\hat{x}_i(\hat{\theta}_i, \theta_{-i}), x_{-i}(\hat{\theta}_i, \theta_{-i})\big) \tag{9}$$

for all $\hat{x}_i(\hat{\theta}_i, \theta_{-i}) \in \mathcal{X}_i$, for all $\hat{\theta}_i \in \Theta_i$, for all θ, and for all i.

Intuitively, if a mechanism $\langle x, y \rangle$ solves Eq. 8, we can use it to construct a stationary, pure strategy EPPPE as follows. Along the equilibrium path, the players reveal their signals truthfully to each other, and implement the recommendations of the mechanism in every period. In particular, to implement y they simply make transfers and burn money as called for. Since $\langle x, y \rangle$ is an EPIC mechanism, and conforming to the mechanism yields the same expected utility in every future period, EPIC is satisfied in every period along the equilibrium path. Upon observing any deviation, such as a deviant action or deviant transfer, they trigger a permanent punishment. Assumption 4 guarantees the existence of a trigger punishment that satisfies EPIC in every period. Since π is uniformly bounded, the punishment then deters any observable deviations if the players are patient enough.

Compared to an equilibrium constructed in this manner, nothing is to be gained by varying continuation utilities rather than transferring or burning money. Thus Eq. 8 bounds the utility that can be attained in any EPPPE, regardless of how patient the players are. This is expressed in the following theorem, which follows from the dynamic programming logic of Abreu, Pearce, and Stacchetti (1990) and Fudenberg and Levine (1994). The proof is in Appendix B.

Theorem 3. *Aggregate utility (in average terms) in any EPPPE is bounded above by V^*, uniformly for all $\delta < 1$. Under Assumptions 3–4, for $\delta < 1$ sufficiently high there exists a stationary, pure strategy EPPPE that attains this bound.*

3.2 EPIC-rich games

I term a game *EPIC-simple* if the no-subsidy constraint does not bind; e.g., the efficient allocation rule can be implemented without any money burning. If on the other hand the no-subsidy constraint binds, I call the game *EPIC-rich*. In an EPIC-rich game, either money must be burned or the action rule must be distorted in order to satisfy the no-subsidy constraint.

Definition 2. A game is *EPIC-simple* if

$$V^* = \sup_{\langle x,y\rangle:\Theta\to\mathcal{X}\times\mathbb{R}^N} \mathbb{E}\left[\sum_i \pi_i(\theta, x(\theta))\right] \quad \text{s.t. EPIC;} \tag{10}$$

otherwise it is *EPIC-rich*.

Intuitively, the properties that make a game EPIC-simple are properties that trivialize either the problem of balancing the budget or the problem of providing incentives. For an example, if one player does not have private information, then she can insure the other players against budget imbalances. This is the case in auction models when the auctioneer has a fixed valuation of zero. A second example is when the efficient action does not vary with θ; then no EPIC payments need be made. This is the case in pure common value auctions, in which it does not matter which player wins the object because all players value it equally. This is also the case in games that satisfy the genericity conditions of Jehiel, Meyer-ter-Vehn, Moldovanu, and Zame (2006) (multidimensional types, interdependent preferences, and consumption externalities; see also Bikhchandani 2005), in which only trivial action rules are implementable under EPIC. Trivial action rules can be implemented without any transfers. Third, if payoffs are identical across players—that is, what is best for any one is best for all—then the game is isomorphic to a one-player decision problem, and truthful revelation may be taken for granted.

For some types of games there are knife-edge cases that are EPIC-simple. Chung and Ely (2006) give an example of this sort: a simple trading game between two players in which $x_i(\theta)$ is the probability that player i gets the object, and $\pi_i(\theta, x(\theta)) = (\theta_i + a\theta_{-i})\, x_i(\theta)$. The game is EPIC-simple if and only if $a = -1$. Liu and Tian (1999) show that a public good game with private values and three or more players is EPIC-simple only if the utility functions take a particular non-generic functional form. For $N = 3$, this is the quadratic form studied by Laffont and Maskin (1980). In general, the functional forms they find are solutions to $(N-2)$-degree polynomial differential equations.

More generally, the literature has found efficient allocation to be generally incompatible with robust implementation. Hurwicz and Walker (1990) show for exchange economies that dominant strategy mechanisms generically cannot attain efficiency with budget balance.[18] Gärtner and Schmutzler (2009) derive broad conditions under which EPIC merger mechanisms, in which firms' owners can exchange both money and shares in the merged entity, cannot implement efficient merger rules under budget balance. Optimal (but inefficient) dominant strategy mechanisms under budget balance have been studied by Hagerty and Rogerson (1987) and Barberà and Jackson (1995), among others.

To illuminate which games are EPIC-rich, I consider a special class of games that contains the

[18] Their environment has private values, so dominant strategy implementation is equivalent to EPIC implementation.

economic models most commonly used to study settings with private information. This class is called the *separable payoff environment* (Chung and Ely 2006), and it is characterized by one-dimensional signals and multiplicatively separable payoffs.

Definition 3. A game is in the *separable payoff environment* if, for each $i \in N$:

(i) $\Theta_i = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}$;

(ii) $\pi_i(\theta, \chi) = z_i(\chi)\, w_i(\theta)$;

(iii) $z_i : \mathcal{X} \to \mathbb{R}_+$ is uniformly bounded;

(iv) $w_i : \Theta \to \mathbb{R}$ is continuous, uniformly bounded, differentiable, and strictly increasing in θ_i for all θ_{-i};

(v) If x is efficient, then $z_i(x(\theta))$ is non-decreasing in θ_i for all θ_{-i}.

Conditions (i–iv) place the game in the separable payoff environment as defined by Chung and Ely (2006). Condition (v) is added for convenience, to ensure that efficiency is ex post implementable (by Chung and Ely 2006, Theorem 2). Lemma 4, in Appendix C shows how to use linear programming to compute an optimal EPPPE in the separable payoff environment.

Given an ex post implementable action rule x, in an EPIC mechanism each player's transfer function must have a particular shape. Specifically, Chung and Ely (2006) Proposition 5 implies that the envelope theorem (e.g., Milgrom and Segal 2002 Corollary 1) fully determines each player's equilibrium utility for each θ_{-i}, up to an additive constant $y_i(\underline{\theta}, \theta_{-i})$: for all θ,

$$\begin{aligned}\pi_i(\theta, x(\theta)) + y_i(\theta) = {} & \pi_i\big((\underline{\theta}, \theta_{-i}), x(\underline{\theta}, \theta_{-i})\big) + y_i(\underline{\theta}, \theta_{-i}) \\ & + \int_{\underline{\theta}}^{\theta_i} \frac{\partial w_i(s_i, \theta_{-i})}{\partial s_i} z_i\big(x(s_i, \theta_{-i})\big)\, ds_i. \end{aligned} \tag{11}$$

Each player i's transfer $y_i(\theta)$ can be decomposed as $y_i(\theta) = h_i(\theta_{-i}) - r_i(\theta; x)$, where $h_i(\theta_{-i}) \equiv y_i(\underline{\theta}, \theta_{-i})$ is her *fixed transfer*, the portion of her transfer that does not vary with θ_i; and $r_i(\theta; x)$ is her *EPIC payment*, the portion of her transfer that is pinned down by EPIC and the choice of action rule x. Then rearranging Eq. 11 gives

$$r_i(\theta; x) = \pi_i(\theta, x(\theta)) - \pi_i\big((\underline{\theta}, \theta_{-i}), x(\underline{\theta}, \theta_{-i})\big) - \int_{\underline{\theta}}^{\theta_i} \frac{\partial w_i(s_i, \theta_{-i})}{\partial s_i} z_i\big(x(s_i, \theta_{-i})\big)\, ds_i. \tag{12}$$

Let $R(\cdot; x) \equiv \sum_i r_i(\cdot; x)$ be the *aggregate EPIC payment function*.

Here I extend the seminal result of Hurwicz and Walker (1990) to the present context, where valuations may be interdependent. I identify a simple necessary and sufficient condition for EPIC-richness in the separable payoff environment.

Definition 4. A function $a : \Theta \to \mathbb{R}$ is $(N-1)$–*additively separable* if there exist functions $\alpha_i : \Theta_{-i} \to \mathbb{R}$, $i = 1, \ldots, N$, such that $a(\theta) = \sum_i \alpha_i(\theta_{-i})$ for all θ.

Theorem 4. *In the separable payoff environment, a game is EPIC-simple if and only if, for any action rule $\hat{x}$ that solves the right hand side of Eq. 10, the aggregate EPIC payment function $R(\cdot;\hat{x})$ is $(N-1)$–additively separable.*

The proof of sufficiency, in Appendix C, is straightforward. If there is an efficient action rule with an $(N-1)$-additively separable EPIC payment function, then by definition there exist functions $\alpha_i : \Theta_{-i} \to \mathbb{R}$, $i = 1, \ldots, N$, such that $R(\theta;\hat{x}) = \sum_i \alpha_i(\theta_{-i})$ for all θ. Then each player i can be given a rebate equal to $\alpha_i(\theta_{-i})$ without disrupting incentives, and the budget will balance ex post. The argument for necessity is more subtle, and relies on the fact that an optimal mechanism always exists in the separable payoff environment, as follows from Lemma 3.

$(N-1)$–additive separability relates closely to the "decomposability" property that Hurwicz and Walker (1990) identify. However, because valuations may be interdependent, $(N-1)$–additive separability does not translate into a criterion that can be expressed directly in terms of the primitives. Instead, as the interdependent trading game example of Chung and Ely (2006) illustrates, interdependence of just the right form can generate $(N-1)$–additive separability. So in principle, to know whether a game is EPIC-rich or EPIC-simple, one must actually examine the aggregate EPIC payment functions for the efficient action rules and identify whether any is $(N-1)$-additively separable. This is illustrated in the following simple example regarding public goods.

Example 6. Consider an indivisible public good, which can be provided with any probability $\chi \in [0,1]$. Agent i's valuation is $w_i(\theta) = \theta_i \in [-1,1]$ and her evaluation is $\pi_i(\chi) = \chi$. An efficient provision rule is $\hat{x}(\theta) = \mathbb{I}\big(\sum_i \theta_i > 0\big)$. To implement this rule under EPIC requires a Groves (1973) mechanism, in which $r_i(\theta;\hat{x}) = \big(\hat{x}(\theta) - \hat{x}(-1,\theta_{-i})\big)\sum_{j\neq i}\theta_j$. But $\hat{x}$ is not $(N-1)$-additively separable, and $R(\cdot;\hat{x})$ is a non-degenerate linear transformation of $\hat{x}$, so the game is not $(N-1)$-additively separable. Therefore the game is EPIC-rich.

Though $(N-1)$-additive separability of $R(\cdot;\hat{x})$ is a necessary and sufficient condition for EPIC-efficiency, it is not expressed directly in terms of the primitives of the game. Though Liu and Tian (1999) find primitive conditions for a narrow subset of the separable payoff environment, necessary and sufficient conditions in terms of primitives across the entire separable payoff environment remain elusive. Instead, I prove a genericity result: Games in the separable payoff environment are *generically* EPIC-rich.

Theorem 5. *Suppose that $\mathcal{X}$ is the space of probability measures on a finite set $\mathcal{A}$. In the separable payoff environment, fix Θ, and endow the space of possible w and z functions with the topology of pointwise convergence. Then every non-degenerate game in the separable payoff environment for $\mathcal{X}$ and Θ is contained in the closure of an open set of EPIC-rich games.*

The proof, in Appendix C, exploits the fact that games in the separable payoff environment generically have interdependent valuations. Starting from a game with private values that is EPIC-simple, the proof constructs a perturbation that introduces some interdependence over a small region, disrupting $(N-1)$-additive separability. More intuitively, since the EPIC payments called for in a generalized Groves mechanism are non-linear transformations of the individual valuation functions w_i, if each individual's valuation function w_i is itself not $(N-1)$-additively separable, then generically the aggregate EPIC payment function $R(\cdot;\hat{x})$—a nonlinear combination of the individual valuation functions—is also not $(N-1)$-additively separable.[19]

4 Conclusion

This paper provides a new explanation for why even a patient cartel may engage in price wars. If the information environment is complex, the cartel may seek an arrangement that is robust to payoff-irrelevant signals that disrupt common knowledge. A robust equilibrium—termed an ex post perfect public equilibrium (EPPPE)—cannot attain efficiency, but eliminates the unmodeled burden of maintaining rigid protocols in the face of complexity. In an optimal EPPPE, the firms do not allocate market shares efficiently. Depending on the environment, their optimal EPPPE may or may not involve price wars. A sufficient condition for price wars to be optimal is that the firms' costs are interdependent, e.g., driven by both idiosyncratic and common underlying shocks. In the literature, this is the first explanation for why a patient, optimizing cartel should engage in price wars if it is able to communicate and play an asymmetric equilibrium.

The cartel's inability to collude efficiently does not arise from special features of the particular collusion game under study. Rather, this paper shows that inefficiency arises generically among games in the separable payoff environment, which contains a great array of economically relevant models, including public goods and trade. For an EPPPE to attain efficiency in the separable payoff environment, it is necessary and sufficient for the aggregate EPIC payment function, which can be computed from the primitives of the stage game, to be $(N-1)$-additively separable. However, this condition is generically violated, so efficiency is generically unattainable.

Though the mechanism design literature has long recognized that efficiency is typically unattainable in economic models under EPIC and ex post budget balance, the problem of optimal ex post implementation under a no-subsidy condition has largely been ignored. (To my knowledge, Shao and Zhou 2008—developed concurrently with the present paper—is the only exception to this

[19] Indeed, this same basic intuition underlies Theorem 2.

generalization.) Perhaps this is because it strains the imagination that a mechanism designer who can commit to enforcing transfers (a necessary assumption in static contexts) should object to a budget that is balanced ex ante but unbalanced ex post. Only when the mechanism encompasses an entire society should such a designer object to an ex post budget imbalance.[20] If instead a static mechanism is interpreted as a contract among the parties, then in addition to EPIC and the no-subsidy condition, ex post participation constraints ought to be imposed. For instance, participation constraints are the source of inefficiency in the seminal Myerson and Satterthwaite (1983) theorem.

In the context of collusion, however, the no-subsidy condition follows naturally from the restriction that firms should not bring in money from outside the game. Their commitment to the mechanism arises from the threat of endogenous punishments, relaxing participation constraints. Since there is no exogenous authority, there is no presumption that the firms ought to be able to obtain insurance against budget imbalances. If they could indeed accept budget imbalances, such abilities should be modeled within the game; this is the approach taken by Athey and Miller 2007 for a bilateral trading relationship.[21]

[20]Indeed, debate over the budget imbalance properties of dominant strategy implementation arose in the 1970s regarding society-wide mechanisms; see Greenberg, Mackay, and Tideman (1977) and Groves and Ledyard (1977a,b).

[21]In addition, a new literature on dynamic mechanism design has arisen contemporaneously with this paper, studying the related problem faced by a mechanism designer with commitment power who wants to implement a particular outcome in a dynamic environment (Athey and Segal 2007, Bergemann and Välimäki 2010, Parkes and Singh 2003, Pavan et al. 2009). This literature primarily addresses efficient and revenue-maximizing mechanisms, rather than constrained optimal mechanisms.

Appendix A Equilibrium

This section provides the notation necessary to describe strategies and equilibria, given any message space $\mathcal{M} = \mathcal{M}_1 \times \cdots \times \mathcal{M}_N$ satisfying $\mathcal{M}_i \supset \Theta_i$ for all i. A behavioral stage strategy for player i is a triplet $s_i = \langle \hat{m}_i, \hat{x}_i, \hat{t}_i \rangle$ that contains a reporting rule $\hat{m}_i : \Theta_i \to \Delta\mathcal{M}_i$, an action rule $\hat{x}_i : \Theta_i \times \mathcal{M} \to \Delta\mathcal{X}_i$, and a payment rule $\hat{t}_i : \Theta_i \times \mathcal{M} \times \mathcal{X} \to \Delta\mathbb{R}^N_+$, where the realization of $\hat{t}_{i,j}(\theta_i, \mu, \chi) \geq 0$ is the amount that player i pays player j, and $\hat{t}_{i,i}(\theta_i, \mu, \chi) \geq 0$ is the amount that player i burns. Abusing notation, I write pure stage strategies, behavioral stage strategies, and the realizations of behavioral stage strategies all in the same way, relying on the context to distinguish them.

A behavioral stage strategy profile is a vector $s \equiv (s_1, \ldots, s_N)$, or, equivalently, $s \equiv \langle \hat{m}, \hat{x}, \hat{t} \rangle$. The public history at the end of period τ is $H^\tau \equiv (h^1, \ldots, h^\tau)$, and the private history for player i is $H_i^\tau \equiv (h_i^1, \ldots, h_i^\tau)$; H^0 and H_i^0 are null histories. A behavioral strategy for player i is a function σ_i that maps player i's private history (of any length) to a probability distribution over behavioral stage strategies. Players can choose their behavioral stage strategies using an arbitrary public randomization device, so that (indulging in more abuse of notation) $\sigma(\{H_i^{\tau-1}\}_{i=1}^N)$ need not be statistically independent. Given a behavioral strategy profile σ and a set of private histories $\{H_i^{\tau-1}\}_{i=1}^N$, the ex post stage game payoff for player i in period τ is

$$\begin{aligned} \hat{\pi}_i\big(\theta; \sigma(\{H_i^{\tau-1}\}_{i=1}^N)\big) = \pi\big(\theta, \hat{x}(\theta, \hat{m}(\theta))\big) &+ \sum_{j \neq i} \hat{t}_{j,i}\big(\theta_j, \hat{m}(\theta), \hat{x}(\theta, \hat{m}(\theta))\big) \\ &- \sum_j \hat{t}_{i,j}\big(\theta_j, \hat{m}(\theta), \hat{x}(\theta, \hat{m}(\theta))\big), \end{aligned} \tag{13}$$

where the understanding is that $\langle \hat{m}, \hat{x}, \hat{t} \rangle$ is the outcome of the randomization specified by the stage strategy $\sigma(\{H_i^{\tau-1}\}_{i=1}^N)$.

A public strategy for player i is a behavioral strategy such that $\sigma_i(H_i^{\tau-1}) = \sigma_i(\tilde{H}_i^{\tau-1})$ whenever $H^{\tau-1} = \tilde{H}^{\tau-1}$; i.e., a behavioral strategy in which player i ignores her private history. When σ_i is a public strategy, I write $\sigma_i(H^{\tau-1})$ for simplicity. Given a profile of public strategies, observe that the payoff player i earns by deviating to any alternative strategy can be attained by deviating to a public strategy, since her private history does not covary with the public strategies of the other players. Given a profile of public strategies σ, the value to player i of the public history H^τ is

$$\hat{v}_i(H^\tau; \sigma) = (1-\delta)\,\mathbb{E}\left[\sum_{\tilde{\tau}=\tau+1}^{\infty} \delta^{\tilde{\tau}-1} \hat{\pi}_i(\vartheta^{\tilde{\tau}}; \sigma(H^{\tilde{\tau}-1}\}) \,\middle|\, H^\tau, \sigma\right], \tag{14}$$

where $\vartheta^{\tilde{\tau}}$ is the random signal vector in period $\tilde{\tau}$.

Definition 5. A *perfect public equilibrium*, or PPE, is a public strategy profile σ such that, for all public histories H, the behavioral stage strategy $\sigma(H) = \langle \hat{m}, \hat{x}, \hat{t} \rangle$ satisfies

(i) for all signals $\theta_i \in \Theta_i$ and for all i,

$$\theta_i \in \arg\max_{\mu_i \in \Theta_i} \mathbb{E}\left[\begin{array}{l} \pi_i\big(\theta_i, \vartheta_{-i}; \chi\big) + \sum_{j \neq i} \hat{t}_{j,i}\big(\vartheta_j, \mu_i, \hat{m}_{-i}(\vartheta_{-i}), \chi\big) \\ -\sum_j \hat{t}_{i,j}\big(\theta_i, \mu_i, \hat{m}_{-i}(\vartheta_{-i}), \chi\big) + \dfrac{\delta}{1-\delta}\hat{v}_i(H'; \sigma) \end{array}\middle| H, \theta_i\right], \tag{15}$$

where

$$\chi = \big(\hat{x}_i(\theta_i, \mu_i, \hat{m}_{-i}(\vartheta_{-i})), \hat{x}_{-i}(\vartheta_{-i}, \mu_i, \hat{m}_{-i}(\vartheta_{-i}))\big), \tag{16}$$

$$H' = \big(H, (\mu_i, \hat{m}_{-i}(\vartheta_{-i}), \chi, \hat{t}(\mu_i, \hat{m}_{-i}(\vartheta_{-i}), \chi)\big); \tag{17}$$

(ii) for all signals $\theta_i \in \Theta_i$, for all message profiles $\mu \in \Theta$, and for all i,

$$\hat{x}_i(\mu) \in \arg\max_{\chi_i \in \mathcal{X}_i} \mathbb{E}\left[\begin{array}{l} \pi_i\big(\theta_i, \vartheta_{-i}; \chi\big) + \sum_{j \neq i} \hat{t}_{j,i}\big(\vartheta_j, \mu, \chi\big) - \sum_j \hat{t}_{i,j}\big(\theta_j, \mu, \chi\big) \\ + \dfrac{\delta}{1-\delta}\hat{v}_i(H'; \sigma) \end{array}\middle| H, \theta_i\right], \tag{18}$$

where $\chi = \big(\chi_i, \hat{x}_{-i}(\vartheta_{-i}, \mu)\big)$ and $H' = \big(H, (\mu, \chi, \hat{t}(\mu, \chi)\big)$;

(iii) for all signals $\theta_i \in \Theta_i$, for all message profiles $\mu \in \Theta$, for all action profiles $\chi \in \mathcal{X}$, and for all i,

$$\hat{t}_i(\mu, \chi) \in \arg\max_{t_i \in \mathbb{R}^N} \mathbb{E}\Big[\sum_{j \neq i} \hat{t}_{j,i}\big(\vartheta_j, \mu, \chi\big) - \sum_j t_{i,j}\big(\theta_j, \mu, \chi\big) + \frac{\delta}{1-\delta}\hat{v}_i(H'; \sigma)\Big| H, \theta_i\Big], \tag{19}$$

where $H' = \big(H, (\mu, \chi, t_i, \hat{t}_{-i}(\mu, \chi)\big)$.

Definition 6. An *ex post perfect public equilibrium*, or EPPPE, is a PPE σ such that for all signal vectors θ, for all public histories H, and for all i,

$$\theta_i \in \arg\max_{\mu_i \in \Theta_i} \left(\begin{array}{l} \pi_i\big(\theta; \chi\big) + \sum_{j \neq i} \hat{t}_{j,i}\big(\theta_j, \mu_i, \hat{m}_{-i}(\theta_{-i}), \chi\big) \\ -\sum_j \hat{t}_{i,j}\big(\theta_j, \mu_i, \hat{m}_{-i}(\theta_{-i}), \chi\big) + \dfrac{\delta}{1-\delta}\hat{v}_i(H'; \sigma) \end{array}\right), \tag{20}$$

where

$$\chi = \big(\hat{x}_i(\theta_i, \mu_i, \hat{m}_{-i}(\theta_{-i})), \hat{x}_{-i}(\theta_{-i}, \mu_i, \hat{m}_{-i}(\theta_{-i}))\big), \tag{21}$$

$$H' = \big(H, (\mu_i, \hat{m}_{-i}(\theta_{-i}), \chi, \hat{t}(\mu_i, \hat{m}_{-i}(\theta_{-i}), \chi)\big). \tag{22}$$

Appendix B Proof of Theorem 3

Theorem 3 (page 24) is an immediate consequence of the following three lemmas.

Lemma 1. *Given a discount factor $\delta < 1$, suppose that σ is a PPE and H is a public history. Then*

$$\sum_i \hat{v}_i(H;\sigma) \leq \sup_{\langle x,y\rangle:\Theta\to\mathcal{X}\times\mathbb{R}^N} \mathbb{E}\Big[\sum_i \big(\pi_i(\theta, x(\theta)) + y_i(\theta)\big)\Big] \text{ s.t. IIC and } \sum_i y_i(\theta) \leq 0 \text{ for all } \theta. \tag{23}$$

If in addition σ is an EPPPE, then $\sum_i \hat{v}_i(H;\sigma) \leq V^*$.

Proof. Taking parts (i)–(iii) of Definition 5 as given, substitute

$$\begin{aligned} y_i(\theta) = &\sum_{j\neq i} \hat{t}_{j,i}\big(\theta_j, \theta, x(\theta,)\big) - \sum_{j\neq i} \hat{t}_{i,j}\big(\theta_i, \theta, x(\theta)\big) \\ &+ \frac{\delta}{1-\delta}\mathbb{E}\big[\hat{v}_i(H, \theta, x(\theta), \hat{t}(\theta, \theta, x(\theta);\sigma)\big|H,\sigma\big] \end{aligned} \tag{24}$$

into Eq. 15 to obtain the *interim incentive compatibility* (IIC) constraint for a mechanism $\langle x,y\rangle : \Theta \to \mathcal{X}\times\mathbb{T}$:

$$\theta_i \in \arg\max_{\mu_i\in\Theta_i} \mathbb{E}\left[\pi_i\big(\theta_i, \vartheta_{-i}; x(\mu_i, \vartheta_{-i})\big) + y_i(\mu_i, \vartheta_{-i})\big|\theta_i\right] \text{ for all } \theta_i \text{ and all } i. \tag{25}$$

That is, a PPE must satisfy IIC after every history, taking expectations over the future path of play. Let $\tilde{V} \equiv \sup_{H'} \sum_i \hat{v}_i(H';\sigma)$; then $\sum_i \hat{v}_i(H;\sigma) \leq \tilde{V}$, so $\sum_i \hat{y}_i(\theta) \leq \frac{\delta}{1-\delta}\tilde{V}$. It follows that

$$\begin{aligned} \frac{1}{1-\delta}\sum_i \hat{v}_i(H;\sigma) \leq &\sup_{\langle x,y\rangle:\Theta\to\mathcal{X}\times\mathbb{R}^N} \mathbb{E}\Big[\sum_i \big(\pi_i(\theta, x(\theta)) + y_i(\theta)\big)\Big] \\ &\text{s.t. IIC and } \sum_i y_i(\theta) \leq \frac{\delta}{1-\delta}\tilde{V} \text{ for all } \theta; \end{aligned} \tag{26}$$

rearranging yields Eq. 23. If in addition σ is an EPPPE it must also satisfy Eq. 20, so $\sum_i \hat{v}_i(H;\sigma) \leq V^*$. □

An EPPPE is *stationary* if $\sigma(H) = \sigma(H')$ for all equilibrium path public histories H and H'.

Lemma 2. *Under Assumption 4, if there exists a mechanism $\langle x, y\rangle$ that solves Eq. 8, then for $\delta < 1$ sufficiently high there exists a stationary pure strategy EPPPE that attains $\sum_i \hat{v}_i(H^0;\sigma) = V^*$.*

Proof. Suppose that $\langle x, y\rangle$ solves Eq. 8. Construct a public strategy profile σ as follows. First, consider any public history H that contains an observable deviation. By Assumption 4 there exists an ex post equilibrium in the stage game. Let $\sigma(\cdot;H)$ be this equilibrium, and let p be stage game payoff vector associated with $\sigma(\cdot;H)$. If $\sum_i p_i = V^*$, then also play this stage game equilibrium following every history to complete the proof.

Next, suppose that $\sum_i p_i < V^*$. Consider any public history H that contains no observable deviations. For each i, let $\sigma(H) = \langle \hat{m}, \hat{x}, \hat{t}\rangle$ be a pure public strategy profile such that, for all $i \in \mathcal{N}$,

$$\hat{m}_i(\theta_i) = \theta_i, \tag{27}$$

$$\hat{x}_i(\theta_i, \theta) = x_i(\theta), \tag{28}$$

$$\sum_{j\neq i} \hat{t}_{j,i}\big(\theta_j, \hat{m}(\theta), \hat{x}(\theta, \hat{m}(\theta))\big) - \sum_{j\neq i} \hat{t}_{i,j}\big(\theta_j, \hat{m}(\theta), \hat{x}(\theta, \hat{m}(\theta))\big) = y_i(\theta) + f_i, \tag{29}$$

where f is chosen to set $\sum_i f_i = 0$ and $\mathbb{E}\big[\pi_i(\theta, x(\theta) + y_i(\theta)\big] + f_i > p_i$. A profile of such transfers $\hat{t}$ exists because $\sum_i y_i(\theta) \leq 0$ and $\sum_i p_i < V^*$. In addition, following the realization of any message profile $\mu \notin \Theta$, choose $\hat{x}(\theta, \mu)$ and $\hat{t}(\theta, \mu, \hat{x}(\theta, \mu)) = (0, \ldots, 0)$ so as to constitute an ex post equilibrium in the remainder of the stage game (which exists by Assumption 4). Similarly, following the realization of any $\chi \neq \hat{x}(\theta, \mu)$, choose $\hat{t}(\theta, \mu, \chi) = (0, \ldots, 0)$ so as to constitute an ex post equilibrium in the remainder of the stage game.

By construction, this stationary pure strategy profile satisfies EPIC and yields average utility of V^*. After every observable deviation, the players make zero transfers and then play a stage game ex post equilibrium forever. If they are sufficiently patient, the threat of the trigger punishment suffices to discourage any observable deviations. This completes the proof. □

Lemma 3. *Under Assumptions 3–4, there exists a mechanism that solves Eq. 8.*

Proof. The proof follows (Balder 1996, Example 3.3). Define the topology $\mathscr{T} \equiv \{\emptyset, \mathcal{N}\} \times \mathscr{F}$ on $\mathcal{N} \times \Theta$, where $\mathscr{F}$ is the topology associated with ϕ. Let $\mathcal{K} \equiv \mathcal{X} \times [-B, B]^N$, where $B \in \mathbb{R}_+$ is large relative to the uniform bound on π. Let $u(i, \theta, (\chi, t)) \equiv \pi_i(\theta, \chi) + t_i$, $U(i, \theta, (\chi, \psi)) \equiv \sum_j \hat{u}(j, \theta, (\chi, t))$, and $W(i, \theta) \equiv \big\{(\chi, t) \in \mathcal{K} : \sum_j t_j \leq 0\big\}$.

The main task is to establish that Balder's Assumptions 2.1–2.7 are satisfied. Assumption 3 implies that $\mathcal{X}$ is a compact, convex metric space (by Aliprantis and Border 1999, Theorem 14.11), so $\mathcal{K}$ is convex, metrizable, and compact, satisfying Balder's Assumption 2.1. Assumption 3 implies that $\pi_i(\theta, \cdot)$ is a linear function on $\mathcal{X}$ for all θ and all i, so that Balder's Assumption 2.3 is satisfied; it also implies that $W(i, \theta)$ is a closed and convex-valued correspondence, satisfying Balder's Assumption 2.2. Although $u(\cdot)$ is not $\mathscr{T}$-measurable, $u(i, \cdot)$ clearly is, and so $U(\cdot, \cdot, (\chi, t))$ is $\mathscr{T}$-measurable—satisfying Balder's Assumption 2.4. Since π is uniformly bounded, U is uniformly bounded as well, and so Balder's Assumption 2.5 is satisfied. The linearity of $\pi_i(\theta, \cdot)$ also implies that U is concave and continuous, satisfying Balder's Assumption 2.6. Finally, Assumption 4 guarantees that the set of EPIC mechanisms is non-empty, implying Balder's Assumption 2.7. Now Balder's Theorem 2.1 implies that there exists a solution to Eq. 8. □

Appendix C Proofs for Section 3

Let $Y(\theta) \equiv \sum_i h_i(\theta_{-i}) - R(\theta; x)$, $r(\theta; x) \equiv \big(r_i(\theta_{-i}); x\big)_{i=1}^N$, and $h(\theta) \equiv \big(h_i(\theta_{-i})\big)_{i=1}^N$.

Lemma 4. *Let* $\Gamma \equiv \big\{\pi^x(\chi) : \chi \in \mathcal{X}\big\} \subset \mathbb{R}^N$ *and, given* $\gamma : \Theta \to \Gamma$, *let*

$$\tilde{R}(\theta; \omega) \equiv \sum_i \left(\pi_i^\theta(\theta)\gamma(\theta) - \pi_i^\theta(\underline{\theta}, \theta_{-i})\gamma(\underline{\theta}, \theta_{-i}) - \int_{\underline{\theta}}^{\theta_i} \frac{\partial \pi_i^\theta(s_i, \theta_{-i})}{\partial s_i} \gamma(s_i, \theta_{-i})\, ds_i \right). \tag{30}$$

Then

$$\begin{aligned} & max_{\gamma:\Theta\to\mathcal{C},\ (h_i:\Theta_{-i}\to\mathbb{R})_{i=1}^N} \mathbb{E}\left[\sum_i \gamma(\vartheta) w_i(\vartheta) + \sum_i h_i(\vartheta_{-i}) - \tilde{R}(\vartheta; \gamma)\right] \\ & \text{s.t.} \quad \sum_i h_i(\theta_{-i}) \leq \tilde{R}(\theta; \gamma) \text{ for all } \theta; \\ & \qquad\quad \gamma(\theta) \geq \gamma(\theta_i', \theta_{-i}) \text{ for all } \theta_i' \leq \theta_i, \text{ all } \theta, \text{ and all } i; \end{aligned} \tag{31}$$

is a linear program. Furthermore, if $\langle \gamma^*, h^* \rangle$ solves Eq. 31, x^* solves $\pi^x(x^*(\theta)) = \gamma^*(\theta)$ for all θ, and $y_i^*(\theta) \equiv h_i^*(\theta_{-i}) - r_i(\theta; x^*)$ for all θ and all i, then $\langle x^*, y^* \rangle$ is an optimal mechanism.

Proof. By Assumption 3, $\mathcal{C}$ is compact and convex. Hence Eq. 31 is evidently a linear program. EPIC is redundant with the functional form of the objective and the monotonicity requirement on σ (Chung and Ely 2006, Theorem 2), so it can be imposed without altering the program. Then substituting x and y into the program to eliminate σ and h, and deleting the redundant monotonicity constraint, yields Eq. 8. $\square$

Proof of Theorem 4 (page 27). "If": Suppose there exists x^* efficient such that $R(\cdot; x^*)$ is $(N-1)$–additively separable; then by definition there exist functions $h_i : \Theta_{-i} \to \mathbb{R}$, $i = 1, \ldots, N$, such that $\sum_i h_i(\theta_{-i}) = R(\theta; x^*)$ for all θ. Let $y_i(\theta) = h_i(\theta_{-i}) - r_i(\theta; x^*)$; then $\langle x^*, y \rangle$ is EPIC by construction. As claimed, the game is EPIC-simple because x^* is efficient and $Y(\theta) = \sum_i \big(h_i(\theta_{-i}) - r_i(\theta; x^*)\big) = 0$ for all θ.

"Only if": Suppose the game is EPIC-simple; i.e., $V^* = \sup_{\langle x,y \rangle : \Theta \to \mathcal{X} \times \mathbb{R}^N} \mathbb{E}\big[\sum_i \pi_i(\theta, x(\theta))\big]$ subject to EPIC. Then there exists some sequence of (possibly randomized) EPIC mechanisms $\{\langle x_k, y_k \rangle\}_{k=1}^{\infty}$ such that, as $k \to \infty$, (i) $\mathbb{E}\big[\sum_i \pi_i(\theta, x_k(\theta))\big]$ approaches V^* and (ii) $Y_k(\theta) = 0$ has for all θ and for all k. This implies that an optimal mechanism $\langle x^*, y^* \rangle$ must achieve both EPIC efficiency and $Y^*(\theta) = 0$ for all θ, since any other mechanism can be outperformed by some member of the sequence. Since by Lemma 3 an optimal mechanism must exist, y^* can be decomposed into $Y^*(\theta) = \sum_i h_i^*(\theta_{-i}) - R(\theta; x^*)$, which, since $Y^*(\theta) = 0$ for all θ, implies that $R(\theta; x^*)$ is additively separable, as claimed. $\square$

Proof of Theorem 5 (page 28). Fixing $\mathcal{X}$ (the space of probability measures on a finite set $\mathcal{A}$) and $\Theta = [\underline{\theta}, \overline{\theta}]$, any game in the separable utility environment is defined by (w, z), where z is linear and w is continuous, uniformly bounded, and differentiable. Let $\mathcal{G}$ be the vector space containing all games satisfying these conditions, endowed with the topology of pointwise convergence. The separable payoff environment, $\mathcal{G}^*$, is a subset of this space, requiring also that $w_i(\theta)$ and $z_i(x(\theta))$ be strictly increasing in θ_i for all i, where x is an efficient allocation rule. Say that a game in the separable payoff environment is *degenerate* if every EPIC-efficient action rule is constant on Θ. Since a non-degenerate allocation rule x is EPIC-implementable if $z_i(x(\theta))$ is strictly increasing in θ_i for all i, evidently the set of non-degenerate games in the separable utility environment, $\mathcal{G}^* \in \mathcal{G}$, has non-empty interior.

Say that a game is *strict* if there exists a sublattice $\mathrm{Z} \subset \Theta$ and an efficient action rule x such that (i) $|\mathrm{Z}| = 2^N$; (ii) there exist $\theta, \theta' \in \mathrm{Z}$ with $\theta_i \neq \theta_i'$ for all i; (iii) for each $\theta \in \mathrm{Z}$, $\arg\max_{\chi \in \mathcal{X}} \sum_i \pi_i(\theta, \chi)$ is a singleton; and (iv) there exist $\zeta, \zeta' \in \mathrm{Z}$ such that $\arg\max_{\chi \in \mathcal{X}} \sum_i \pi_i(\zeta, \chi) \neq \arg\max_{\chi \in \mathcal{X}} \sum_i \pi_i(\zeta', \chi)$. Since $\mathcal{G}^*$ contains only non-degenerate games and z is linear, the set of strict games is open and dense in $\mathcal{G}^*$. Moreover, since $\mathcal{A}$ is finite, by continuity (Definition 3(iii–iv)) for any strict game g there exists a lattice $\mathrm{Z} \subset \Theta$, satisfying conditions (i–iv) above, such that g is contained in an open set of strict games whose efficient action rules, restricted to Z, are identical.

Consider a strict game g that is EPIC-simple. Select any lattice $\mathrm{Z} \subset \Theta$ and efficient action rule x satisfying conditions (i–iv) above. It is without loss of generality to suppose that ζ and ζ' differ on only the Θ_i dimension for some i, by transitivity, and that $\zeta_i' > \zeta_i$. Let $E(\zeta, \zeta')$ be the "edge" of Z between ζ and ζ'; i.e., $E(\zeta, \zeta') \equiv \{\beta\zeta + (1-\beta)\zeta' : \beta \in (0,1)\}$. Consider a set $\Upsilon \subset \Theta$ that contains $E(\zeta, \zeta')$ but does not intersect Z or any other edge of Z. By Eq. 12, perturbations of w_i that are restricted to Υ alter $r_i(\zeta'; x)$ but do not affect the EPIC payments at any other point in Z. Inspection of Eq. 12 makes clear that an open set of such perturbations, whose closure contains the null perturbation, strictly changes $r_i(\zeta'; x)$ while leaving the efficient action rule unchanged on Z. (If game g had private values, such a perturbation necessarily introduces interdependence, in order for the other edges of Z to remain unperturbed.) Restrict attention to such perturbations.

By Theorem 4, for game g there exist functions $\alpha_i : \Theta_{-i} \to \mathbb{R}$, $i = 1, \ldots, N$, such that $\sum_i \alpha_i(\theta_{-i}) = \sum_i r_i(\theta; x)$ for all $\theta \in Z$. But then the perturbed game is EPIC-rich, since there do not exist functions $\alpha'_i : \Theta_{-i} \to \mathbb{R}$, $i = 1, \ldots, N$, such that $\sum_i \alpha'_i(\theta_{-i}) = \mathbb{I}(\theta = \zeta')$ for all $\theta \in Z$.

Finally, although the perturbations considered thus far were restricted to a subspace of $\mathcal{G}$, since $R(\cdot; x)$ restricted to Z is continuous on $\mathcal{G}^*$, every such perturbed game itself is contained in an open set of EPIC-rich games. □

Appendix D Proofs for Section 2

Definition 7. Fixed transfer function h^* is *conditionally optimal given* x if it solves

$$\max_{(h_i:\Theta_{-i}\to\mathbb{R})_{i=1}^N} \mathbb{E}\Big[\sum_i h_i(\vartheta_{-i}) - R(\vartheta; x)\Big] \quad \text{s.t.} \sum_i h_i(\theta_{-i}) - R(\theta; x) \leq 0 \;\forall\, \theta. \tag{32}$$

Let $h^*(\cdot; x)$ be conditionally optimal given x. The next two lemmas are used in the proof of Theorem 1.

Lemma 5. *Under Assumption 1, given an efficient allocation rule $\hat{x}$, there exists a conditionally optimal fixed transfer function $h^*(\cdot; \hat{x})$ and a point b on the interior of Θ such that*

(i) $w_1(b) = w_2(b)$;

(ii) $h_i^*(\theta_{-i}; \hat{x}) = \min\{w_i(\lambda_i^*(\theta_{-i}), \theta_{-i}), w_i(b)\} - \frac{1}{2} w_i(b)$;

where $\lambda_i^*(\theta_{-i})$ solves $w_i(\lambda_i, \theta_{-i}) = w_{-i}(\lambda_i, \theta_{-i})$.[22] *Furthermore, $H^*(\theta; \hat{x}) - R(\theta; \hat{x}) = 0$ for all $\theta \in [b_1, 1] \times [0, b_2] \cup [0, b_1] \times [b_2, 1]$.*

Proof. An allocation rule $\hat{x}$ is efficient if and only if $\hat{x}$ allocates the object to a player with a highest valuation; i.e., if $\hat{x}_i(\theta) = 0$ for all $i \notin \arg\max_j [w_j(\theta)]$, for all θ. The EPIC payments for an efficient allocation rule take the form $R(\theta; \hat{x}) \equiv \sum_i r_i(\theta; \hat{x}) = \sum_i \mathbb{I}_{\{\hat{x}_i(\theta)=1\}} w_i(\lambda_i^*(\theta_{-i}), \theta_{-i})$.

Fix h_{-i}, assuming nothing about its form, and let $k_{-i} = \sup_{\vartheta_i} [h_{-i}(\vartheta_i)]$. The solution to Eq. 32, given $\hat{x}$ and h_{-i} fixed, is to maximize h_i pointwise subject to $h_i(\theta_{-i}) \leq R(\theta; \hat{x}) - h_{-i}(\theta_i)$ for all θ. For any particular θ_i, the constraint $h_i(\theta_{-i}) \leq R(\theta; \hat{x}) - h_{-i}(\theta_i)$ requires

$$h_i(\theta_{-i}) \leq \begin{cases} w_i\big(\lambda_i^*(\theta_{-i}), \theta_{-i}\big) - h_{-i}(\theta_i) & \text{if } \theta_{-i} \in [0, \lambda_{-i}^*(\theta_i)], \\ w_{-i}(\lambda_{-i}^*(\theta_i), \theta_i) - h_{-i}(\theta_i) & \text{otherwise.} \end{cases} \tag{33}$$

The "otherwise" part of the constraint is flat, while the part of the constraint for $\theta_{-i} \in [0, \lambda_{-i}^*(\theta_i)]$ strictly increases in θ_{-i} under the regularity conditions. Since $w_i\big(\theta_{-i}, \lambda_i^*(\theta_{-i})\big) = w_{-i}\big(\theta_{-i}, \lambda_i^*(\theta_{-i})\big)$, the two parts of the constraint coincide at $\theta_{-i} = \lambda_{-i}^*(\theta_i)$ and at no other point. Thus the constraints for all $\theta_i \in \Theta_i$ can be simplified as follows:

$$h_i(\theta_{-i}) \leq \min\Big\{w_i\big(\lambda_i^*(\theta_{-i}), \theta_{-i}\big) - k_{-i}, \inf_{\vartheta_i}\big[w_{-i}\big(\lambda_{-i}^*(\vartheta_i), \vartheta_i\big) - h_{-i}(\vartheta_i)\big]\Big\}. \tag{34}$$

[22] That is, given θ_{-i}, $\lambda_i^*(\theta_{-i})$ is the signal for player i at which she is pivotal.

Let h_i^* equal this upper bound pointwise. If b_{-i} satisfies

$$w_i\big(\lambda_i(b_{-i}), b_{-i}\big) - k_{-i} = \inf_{\vartheta_i}\big[w_{-i}(\lambda_{-i}^*(\vartheta_i), \vartheta_i) - h_{-i}(\vartheta_i)\big], \tag{35}$$

then

$$h_i^*(\theta_{-i}) = \min\big\{w_i\big(\lambda_i^*(\theta_{-i}), \theta_{-i}\big), w_i\big(\lambda_i^*(b_{-i}), b_{-i}\big)\big\} - k_{-i}. \tag{36}$$

Applying these results to both h_i and h_{-i} implies that $w_i\big(\lambda_i^*(b_{-i}), b_{-i}\big) = w_{-i}\big(\lambda_{-i}^*(b_i), b_i\big)$ and $\sup_{\vartheta_{-i}}\big[h_i(\vartheta_{-i})\big] = w_i(b) - k_{-i} = k_i$. Since lump sum adjustments do not affect incentives, set $k_i = k_{-i} = \frac{1}{2}w_i(b)$ to yield (i) and (ii) as claimed.

Finally, on the region $[b_1, 1] \times [0, b_2]$,

$$\sum_i h_i^*(\theta_{-i}; \hat{x}) - R(\theta; \hat{x}) = w_1(\lambda_1^*(\theta_2), \theta_2) + w_2(b) - w_2(b) - w_1(\lambda_1^*(\theta_2), \theta_2) = 0, \tag{37}$$

and similarly for the region $[0, b_1] \times [b_2, 1]$. □

Lemma 6. *In the context of Lemma 5, suppose that an ex post implementable allocation rule x is identical to x^* except on a rectangle $[0, \varepsilon_1] \times [0, \varepsilon_2]$ with ε_1 and ε_2 sufficiently small. Then $\sum_i h_i^*(\theta_{-i}; x) - R(\theta; x) = Y^*(\theta; x^*)$ for all $\theta \notin [0, \varepsilon_1] \times [0, \varepsilon_2]$.*

Proof sketch. Use the proof of Lemma 5, but replace $w_i(\lambda_i^*(\theta_{-i}), \theta_{-i})$ with $r_i((1, \theta_{-i}); x)$. Since $r_i((1, \theta_{-i}); x) = r_i((1, \theta_{-i}); x^*)$, for $\theta_{-i} \in [\varepsilon_{-i}, 1]$, this will imply that $h_i^*(\theta_{-i}; x^*) = h_i^*(\theta_{-i}; x)$. For $\theta_{-i} \in [0, \varepsilon_{-i}]$, it will imply that $r_i((1, \theta_{-i}); x) - r_i((1, \theta_{-i}); x^*) = h_i^*(\theta_{-i}; x^*) - h_i^*(\theta_{-i}; x)$, with $y_i^*(\theta; x^*) = y_i^*(\theta; x)$ for all $\theta \notin [0, \varepsilon_1] \times [0, \varepsilon_2]$. □

Proof of Theorem 1 (page 12). On a sufficiently small rectangle that borders the origin, first order approximations are valid under Assumption 1. So for $\eta > 0$ small, for all $\theta \in E^\eta \equiv [0, \eta] \times [0, \beta\eta]$ assume that $w_1(\theta) = c_{10} + c_{11}\theta_1 + c_{12}\theta_2$, $w_2(\theta) = c_{20} + c_{21}\theta_1 + c_{22}\theta_2$, and $\phi(\theta) = c_{30} + c_{31}\theta_1 + c_{32}\theta_2$, and let $\beta \equiv \big(c_{11} - c_{21}\big)/\big(c_{22} - c_{12}\big)$. By Assumption 1, $\min\{c_{30}, c_{11}, c_{22}\} > 0$ and $0 < \beta < \infty$.

Let $x_1^\eta(\theta) = \chi_1 = 1 - x_2^\eta(\theta) \in [0, 1]$ if $\theta \in E^\eta$, and $x^\eta(\theta) = x^*(\theta)$ otherwise; note that any $\chi_1 \in [0, 1]$ will preserve monotonicity. By Lemma 6, $Y^*(\theta; x^*) = Y^*(\theta; x^\eta)$ on all of $\Theta \backslash E^\eta$. Hence to demonstrate an improvement we need only to show that $\Delta V > 0$, where

$$\Delta V \equiv \int_E^\eta \Big(\big(w_1(\vartheta) - w_2(\vartheta)\big)\big(\chi_1 - x_1^*(\vartheta)\big) + Y^*(\vartheta; x^\eta) + Y^*(\vartheta; x^*)\Big)\, d\phi(\vartheta). \tag{38}$$

Define the following, and note their values for $\theta \in E^\eta$:

$$\Delta h_2^*(\theta_1) \equiv h_2^*(\theta_1; x^\eta) - h_2^*(\theta_1; x^*) = w_2(\theta_1, \beta\eta)\,\chi_1 - w_2(\theta_1, \beta\theta_1), \tag{39}$$

$$\Delta h_1^*(\theta_2) \equiv h_1^*(\theta_2; x^\eta) - h_1^*(\theta_2; x^*) = w_1(\eta, \theta_2)\,(1 - \chi_1) - w_1(\tfrac{1}{\beta}\theta_2, \theta_2), \tag{40}$$

$$\Delta r_1(\theta) \equiv r_1(\theta; x^\eta) - r_1(\theta; x^*) = -w_1(\tfrac{1}{\beta}\theta_2, \theta_2)\,\mathbb{I}_{\{\theta_1 > \theta_2/\beta\}}, \tag{41}$$

$$\Delta r_2(\theta) \equiv r_2(\theta; x^\eta) - r_2(\theta; x^*) = -w_2(\theta_1, \beta\theta_1)\,\mathbb{I}_{\{\theta_2 > \beta\theta_1\}}. \tag{42}$$

Divide E^η into two parts, $E_1^\eta \equiv \{\theta \in E^\eta : \theta_2 < \beta\theta_1\}$ and $E_2^\eta \equiv \{\theta \in E^\eta : \theta_2 \geq \beta\theta_1\}$; then

$$\Delta V = \int_{E_1^\eta} \Big(\big(w_1(\vartheta) - w_2(\vartheta)\big)\,(\chi_1 - 1) + \Delta h_2^*(\vartheta_1) + \Delta h_1^*(\vartheta_2) - \Delta r_1(\vartheta) \Big)\, d\phi(\vartheta) \\ + \int_{E_2^\eta} \Big(\big(w_1(\vartheta) - w_2(\vartheta)\big)\chi_1 + \Delta h_2^*(\vartheta_1) + \Delta h_1^*(\vartheta_2) - \Delta r_2(\vartheta) \Big)\, d\phi(\vartheta). \tag{43}$$

Note that ΔV is linear in χ_1, so that a maximum is to be found either at $\chi_1 = 0$ or at $\chi_1 = 1$. Thus it suffices to show that $\overline{\Delta V} \equiv \frac{1}{2}\big(\Delta V|_{\chi_1=0} + \Delta V|_{\chi_1=1}\big) > 0$. Solving explicitly for $\overline{\Delta V}$ yields $\overline{\Delta V} = \frac{\beta}{12}(c_{11} + \beta c_{22})c_{30}\eta^3 + A\eta^4$, where A is a term that does not vary with η. Since the first term is an order of η larger than the second term, it suffices to note that the first term is strictly positive under the regularity conditions. □

The next three lemmas are used in the proof of Theorem 2.

Lemma 7. *In a two-firm collusion game, suppose that the firms' valuation functions satisfy globally interdependent valuations. Then generically any optimal EPIC mechanism satisfying ex post budget balance must employ a differentiable allocation rule satisfying*

$$3\big(w_2(\theta) - w_1(\theta)\big)\frac{\partial^2 x_1}{\partial\theta_2\partial\theta_1} = \frac{\partial w_1}{\partial\theta_2}\frac{\partial x_1}{\partial\theta_1} + \frac{\partial w_2}{\partial\theta_1}\frac{\partial x_2}{\partial\theta_2} \tag{44}$$

for all $\theta \in int(\Theta)$.

Proof. Since Eq. 3 is a linear program with continuously differentiable constraints, the solution is piecewise differentiable. That is, there exists a countable partition of Θ, such that each partition element is a Borel set with a piecewise continuously differentiable boundary, and the optimal mechanism is continuously differentiable on the interior of each partition element. Each such partition element is a *regular region*; a *regular boundary* is a shared boundary of adjacent regular regions.

Given $x : \Theta \to \mathcal{X}$, $\theta \in \text{int}(\Theta)$, and $\varepsilon = (\varepsilon_1^-, \varepsilon_1^+, \varepsilon_2^-, \varepsilon_2^+) \in \mathbb{R}_+^4$, let $\theta^- \equiv (\theta_1 - \varepsilon_1^-, \theta_2 - \varepsilon_2^-)$, $\theta^\pm \equiv (\theta_1 + \varepsilon_1^+, \theta_2 - \varepsilon_2^-)$, $\theta^\mp \equiv (\theta_1 - \varepsilon_1^-, \theta_2 + \varepsilon_2^+)$, and $\theta^+ \equiv (\theta_1 + \varepsilon_1^+, \theta_2 + \varepsilon_2^+)$. Assume that ε is small enough that $\{\theta^-, \theta^\pm, \theta^\mp, \theta^+\} \subset \Theta$, and let

$$\Delta^2(\theta, \varepsilon) \equiv \big(R(\theta^-; x) - R(\theta^\pm; x)\big) - \big(R(\theta^\mp; x) - R(\theta^+; x)\big). \tag{45}$$

A mechanism can satisfy ex post budget balance if and only if $\Delta^2(\theta, \varepsilon) = 0$ for every θ and ε. When $\varepsilon \to (0, 0, 0, 0)$ along some path, I write $\varepsilon \to 0$.

By Eq. 12, for any ex post budget balanced EPIC mechanism, $\Delta^2(\theta, \varepsilon)$ can be expressed in terms of x

as follows:

$$\Delta^2(\theta,\varepsilon) = \sum_i \left(\begin{array}{c} x_i(\theta^+)w_i(\theta^+) - x_i(\theta^\mp)w_i(\theta^\mp) - \displaystyle\int_{\theta_i^-}^{\theta_i^+} \frac{\partial w_i(s_i,\theta_{-i}^+)}{\partial s_i} x_i(s_i,\theta_{-i}^+)\,ds_i \\ - x_i(\theta^\pm)w_i(\theta^\pm) + x_i(\theta^-)w_i(\theta^-) + \displaystyle\int_{\theta_i^-}^{\theta_i^+} \frac{\partial w_i(s_i,\theta_{-i}^-)}{\partial s_i} x_i(s_i,\theta_{-i}^-)\,ds_i \end{array} \right). \tag{46}$$

Given a path along which $\varepsilon \to 0$, for $i = 1, 2$ let

$$\eta_i^- \equiv \lim_{\varepsilon\to 0} \big(x_i(\theta_i^+, \theta_{-i}^-) - x_i(\theta^-) \big) > 0, \tag{47}$$

$$\eta_i^+ \equiv \lim_{\varepsilon\to 0} \big(x_i(\theta^+) - x_i(\theta_i^-, \theta_{-i}^+) \big) > 0, \tag{48}$$

and let $\left(\alpha_i^+, \alpha_i^-\right) \in [0,1]^2$ solve

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon_i^- + \varepsilon_i^+} \int_{\theta_i^-}^{\theta_i^+} \frac{\partial w_i(s_i,\theta_{-i}^+)}{\partial s_i} x_i(s_i,\theta_{-i}^+)\,ds_i = \frac{\partial w_i}{\partial \theta_i}\Big|_\theta \big(\alpha_i^+ x_i(\theta_i^-,\theta_{-i}^+) + (1-\alpha_i^+) x_i(\theta^+)\big), \tag{49}$$

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon_i^- + \varepsilon_i^+} \int_{\theta_i^-}^{\theta_i^+} \frac{\partial w_i(s_i,\theta_2^-)}{\partial s_i} x_i(s_i,\theta_2^-)\,ds_i = \frac{\partial w_i}{\partial \theta_i}\Big|_\theta \big(\alpha_i^- x_i(\theta^-) + (1-\alpha_i^-) x_i(\theta_i^+,\theta_{-i}^-)\big). \tag{50}$$

Differentiable regions Suppose that θ is in the interior of a regular region. Then x is differentiable, implying that $\eta_i^+ = \eta_i^- = 0$ for both i. Then, letting $\varepsilon_1^- = \varepsilon_2^+ = \xi$ and $\varepsilon_1^+ = \varepsilon_2^- = \xi^2$, and taking $\xi \to 0$, so that $\alpha_i^+ = \alpha_i^- = \frac{1}{2}$, it can be shown that[23]

$$\lim_{\xi\to 0} \frac{\Delta^2(\theta,\varepsilon)}{\xi} = 3\big(w_1(\theta) - w_2(\theta)\big)\frac{\partial^2 x_1}{\partial\theta_1\partial\theta_2} + \frac{\partial w_1}{\partial\theta_2}\frac{\partial x_1}{\partial\theta_1} + \frac{\partial w_2}{\partial\theta_1}\frac{\partial x_2}{\partial\theta_2}; \tag{51}$$

i.e., Eq. 44 holds. The remainder of the proof shows that there are no regular boundaries in the interior of Θ, and hence Θ itself is a regular region.

Discontinuities First I rule out discontinuous regular boundaries other than vertical, horizontal, or along the curve $w_1(\theta) = w_2(\theta)$. It can be shown that

$$\lim_{\varepsilon\to 0} \Delta^2(\theta,\varepsilon) = (\eta_2^- - \eta_2^+)\big(w_1(\theta) - w_2(\theta)\big), \tag{52}$$

implying that either $\eta_2^- = \eta_2^+$ or $w_1(\theta) = w_2(\theta)$. Suppose there is a discontinuous regular boundary other than horizontal or vertical. Then choose θ on the discontinuous regular boundary, and let $\varepsilon \to 0$ in a direction such that one of η_2^- and η_2^+ is zero and the other is strictly positive. If $w_1(\theta) \neq w_2(\theta)$, this contradicts $\Delta^2(\theta,\varepsilon) = 0$.

[23]Derivations for the limits given in this proof are available on request.

Next I rule out discontinuities along the curve $w_1(\theta) = w_2(\theta)$. Let $\varepsilon_1^+ = k_1^+\xi$, $\varepsilon_1^- = k_1^-\xi$, $\varepsilon_2^+ = k_2^+\xi$, and $\varepsilon_2^- = k_2^-\xi$, where all four coefficients are strictly positive. Then it can be shown that

$$\begin{aligned}
\lim_{\xi\to 0}\frac{\Delta^2(\theta,\varepsilon)}{\xi} &= k_1^+\left(\frac{\partial w_1}{\partial\theta_1}\big(\alpha_1^+(\eta_1^- + \eta_2^- - \eta_2^+) - \alpha_1^-\eta_1^-\big) + \frac{\partial w_2}{\partial\theta_1}\eta_2^+\right)\\
&\quad + k_2^+\left(\frac{\partial w_2}{\partial\theta_2}\big(\alpha_2^+\eta_2^+ - \alpha_2^-\eta_2^-\big) + \frac{\partial w_1}{\partial\theta_2}\big(\eta_1^- + \eta_2^- - \eta_2^+\big)\right)\\
&\quad + k_1^-\left(\frac{\partial w_1}{\partial\theta_1}\big(\eta_2^+ - \eta_2^- - \alpha_1^-\eta_1^- + \alpha_1^+(\eta_1^- + \eta_2^- - \eta_2^+)\big) + \frac{\partial w_2}{\partial\theta_1}\eta_2^-\right)\\
&\quad + k_2^-\left(\frac{\partial w_2}{\partial\theta_2}\big(\eta_2^- - \alpha_2^-\eta_2^- - (1-\alpha_2^-)\eta_2^+\big) + \frac{\partial w_1}{\partial\theta_2}\eta_1^-\right).
\end{aligned} \tag{53}$$

Suppose there is a discontinuity along the curve $w_1(\theta) = w_2(\theta)$. Then choose k_1^+, k_1^-, k_2^+, and k_2^- so that $\eta_2^+ = \eta_2^- = 0$, making the second and fourth lines of Eq. 53 both strictly negative. Of course this implies that the sum of the first and third lines must be strictly positive, but moreover it implies that $(k_1^+, k_1^-, k_2^+, k_2^-)$ must lie in a 2-dimensional plane in $\mathbb{R}^4$. Now perturb these values slightly out of the plane without violating the condition that $\eta_2^+ = \eta_2^- = 0$, and by doing so generate a contradiction to $\Delta^2(\theta,\varepsilon) = 0$.

Finally, I rule out horizontal discontinuities (and, by symmetry, vertical discontinuities). Suppose there is a horizontal discontinuity at θ, so that $\eta_1^- = \eta_1^+ = 0$ and $\eta_2^+ = \eta_2^- > 0$. Then, taking $\varepsilon = (\xi,\xi,\xi,\xi)$ and $\xi \to 0$, it can be shown that

$$\lim_{\xi\to 0}\frac{\Delta^2(\theta,\varepsilon)}{2\xi} = \frac{\partial w_2}{\partial\theta_1}\eta_2^- - \big(w_1(\theta) - w_2(\theta)\big)\lim_{\xi\to 0}\left(\frac{\partial x_1}{\partial\theta_1}\Big|_{\theta\pm} - \frac{\partial x_1}{\partial\theta_1}\Big|_{\theta+}\right). \tag{54}$$

Since the first term on the right hand side is strictly positive, this discontinuity cannot cross the curve $w_1(\theta) = w_2(\theta)$. If such a discontinuity exists where $w_1(\theta) > w_2(\theta)$ (a similar argument holds for $w_1(\theta) < w_2(\theta)$), there exists a "starting point" $\theta^* \equiv \big(\max_{\tilde\theta_1}\{\tilde\theta_1 : \eta_2^-(\tilde\theta_1,\theta_2) = 0\}, \theta_2\big)$ such that $w_1(\theta^*) \geq w_2(\theta^*)$. Then it can be shown that $\lim_{\xi\to 0}\frac{1}{\xi}\Delta^2(\theta^*,\varepsilon) = -\infty$ unless $\eta_2^+(\theta^*) = 0$ as well (regardless of $\eta_1^-(\theta^*)$ and $\eta_1^+(\theta^*)$). Hence η_2^+, as a function of θ, must be continuous in θ_1. It is also a geometric fact that

$$\frac{\partial\eta_2^+}{\partial\theta_1} = \lim_{\xi\to 0}\left(\frac{\partial x_1}{\partial\theta_1}\Big|_{\theta\pm} - \frac{\partial x_1}{\partial\theta_1}\Big|_{\theta+}\right), \tag{55}$$

which, combined with Eq. 54, yields the differential equation

$$\frac{\partial w_2}{\partial\theta_1}\eta_2^+ = \big(w_1(\theta) - w_2(\theta)\big)\frac{\partial\eta_2^+}{\partial\theta_1}. \tag{56}$$

The boundary condition at θ^* is $\eta_2^+(\theta^*) = 0$. If $\pi_1(\theta^*) > \pi_2(\theta^*)$, then $\eta_2^+(\theta_1,\theta_2^*) = 0$ for all θ_1 is the unique solution, and there can be no horizontal discontinuity. If instead $\pi_1(\theta^*) = \pi_2(\theta^*)$, differentiating both sides of Eq. 56 with respect to θ_1 yields

$$\frac{\partial^2 w_2}{\partial\theta_1^2}\eta_2^+ + \frac{\partial w_2}{\partial\theta_1}\frac{\partial\eta_2^+}{\partial\theta_1} = \left(\frac{\partial w_1}{\partial\theta_1} - \frac{\partial w_2}{\partial\theta_1}\right)\frac{\partial\eta_2^+}{\partial\theta_1} + \big(w_1(\theta) - w_2(\theta)\big)\frac{\partial^2\eta_2^+}{\partial\theta_1^2}, \tag{57}$$

and substituting in θ^*, where $w_1(\theta^*) = w_2(\theta^*)$ and $\eta_2^+ = 0$, implies that

$$2\frac{\partial w_2}{\partial \theta_1}\Big|_{\theta^*}\frac{\partial \eta_2^+}{\partial \theta_1}\Big|_{\theta^*} = \frac{\partial w_1}{\partial \theta_1}\Big|_{\theta^*}\frac{\partial \eta_2^+}{\partial \theta_1}\Big|_{\theta^*}. \tag{58}$$

Since generically $2\frac{\partial w_2}{\partial \theta_1}\big|_{\theta^*} \neq \frac{\partial w_1}{\partial \theta_1}\big|_{\theta^*}$ everywhere along the curve $w_1(\theta) = w_2(\theta)$, generically it must be that $\frac{\partial \eta_2^+}{\partial \theta_1}\big|_{\theta^*} = 0$. Continuing to differentiate in this manner, we see inductively that each succeeding higher order derivative of η_2^+ with respect to θ_1 must satisfy an equation like Eq. 58, and therefore must equal zero generically. That is, generically there can be no horizontal discontinuity at all. Swapping the players, the same argument generically rules out vertical discontinuities, and hence generically there can be no discontinuities anywhere.

Non-differentiabilities First, I rule out the possibility that x could be non-differentiable anywhere $w_1(\theta) \neq w_2(\theta)$. Suppose that x is continuous at such a boundary, but not differentiable. Without loss of generality (we could swap these pairs and make the same argument), let $\varepsilon_1^- = \varepsilon_1^+ = \xi$ and $\varepsilon_2^- = \varepsilon_2^+ = k\xi$ and choose $k \in (0,1)$ sufficiently small that θ^- and $\theta^\mp$ are on one side of the boundary, and $\theta^\pm$ and θ^+ are on the other side, as $\xi \to 0$. It can be shown that

$$\lim_{\xi\to 0}\frac{\Delta^2(\theta,\varepsilon)}{2k\xi} = \big(w_2(\theta) - w_1(\theta)\big)\lim_{\xi\to 0}\Big(\frac{\partial x_1}{\partial \theta_2}\Big|_{\theta^-} - \frac{\partial x_1}{\partial \theta_2}\Big|_{\theta^+}\Big). \tag{59}$$

But since $w_1(\theta) \neq w_2(\theta)$, it must be that $\lim_{\xi\to 0}\frac{\partial x_1}{\partial \theta_2}\big|_{\theta^-} = \lim_{\xi\to 0}\frac{\partial x_1}{\partial \theta_2}\big|_{\theta^+}$, and by symmetry the same is true for $\frac{\partial x_2}{\partial \theta_1}$, so in fact x must be differentiable, contrary to the existence of such a boundary.

Finally, I rule out the possibility that x could be non-differentiable along the curve $w_1(\theta) = w_2(\theta)$. Suppose there is a non-differentiable regular boundary along $w_1(\theta) = w_2(\theta)$. Let $\varepsilon_1^- = \varepsilon_2^+ = \xi$ and $\varepsilon_1^+ = \varepsilon_2^- = \xi^2$; then $\theta^\pm$ is on one side of the boundary, while θ^-, $\theta^\mp$, and θ^+ are on the other side, as $\xi \to 0$. It can be shown that

$$\lim_{\xi\to 0}\frac{\Delta^2(\theta,\varepsilon)}{\xi^2} = \lim_{\xi\to 0}\left(\frac{\partial w_1}{\partial \theta_2}\frac{\partial x_1}{\partial \theta_1}\Big|_{\theta^\pm} + \frac{\partial w_2}{\partial \theta_1}\frac{\partial x_2}{\partial \theta_2}\Big|_{\theta^\pm}\right). \tag{60}$$

But since $\frac{\partial w_1}{\partial \theta_2}$ and $\frac{\partial w_2}{\partial \theta_1}$ are both strictly positive, it must be that $\lim_{\xi\to 0}\frac{\partial x_1}{\partial \theta_2}\big|_{\theta^\pm} = 0$ and $\lim_{\xi\to 0}\frac{\partial x_1}{\partial \theta_2}\big|_{\theta^\mp} = 0$, and by symmetry the same is true for $\frac{\partial x_2}{\partial \theta_1}$, so in fact x must be differentiable, contrary to the existence of such a boundary.

In summary, there can be no discontinuities or non-differentiabilities at any regular boundary. Finally, any solution to a differential equation such as Eq. 44 is smooth, so Θ itself is a regular region. This completes the proof. $\square$

Lemma 8. *In a two-firm collusion game, suppose that the firms' valuation functions satisfy globally interdependent valuations. Then an optimal regular mechanism cannot use a constant allocation rule.*

Proof. Without loss of generality, consider a constant action rule with $\overline{x}_1(\theta) = \chi_1 \leq \frac{1}{2}$ for all θ. Since $r_i(\theta;\overline{x}) = 0$ for all θ and all i, let $\overline{y} = \overline{h} - r(\cdot;\overline{x})$, with $\overline{h}_i(\theta_{-i}) = 0$ for all θ and all i as well. Now, for $\varepsilon > 0$ small, consider an alternative action rule x, such that $x_1(\theta) = 1$ for all $\theta \in \mathcal{T} \equiv [1-\varepsilon, 1] \times [0,\varepsilon]$, and $x_1(\theta) = \chi_1$ otherwise. Choose $h_2(\theta_1) = (1-\chi_2)\,\pi_2(\theta_1,\varepsilon)$ for $\theta_1 \geq 1-\varepsilon$, and $h_2(\theta_1) = 0$ otherwise;

choose $h_1(\theta_2) = 0$ for all θ. Since $r_2(\theta) = (1-\chi_1)\,\pi_2(\theta_1, \varepsilon)$ for $\theta \in [1-\varepsilon, 1] \times (\varepsilon, 1]$ and $r_2(\theta) = 0$ elsewhere, while $r_1(\theta) = (1-\chi_1)\,\pi_1(1-\varepsilon, \theta_2)$ for $\theta \in \mathcal{T}$ and $r_1(\theta) = 0$ elsewhere,

$$\mathbb{E}[Y(\vartheta)] = \int\limits_{\mathcal{T}} (1-\chi_1)\big(w_2(\vartheta_1, \varepsilon) - w_1(1-\varepsilon, \vartheta_2)\big)\, d\phi(\vartheta). \tag{61}$$

At the same time, the increase in expected aggregate payoff is

$$\sum_i \mathbb{E}[\pi_i(\vartheta, x(\vartheta))] - \sum_i \mathbb{E}[\pi_i(\vartheta, \overline{x}(\vartheta))] = \int\limits_{\mathcal{T}} (1-\chi_1)\big(w_1(\theta) - w_2(\theta)\big)\, d\phi(\theta). \tag{62}$$

Hence the improvement in the value of the mechanism is

$$\begin{aligned}\sum_i \mathbb{E}[\pi_i(\vartheta, x(\vartheta))] &- \sum_i \mathbb{E}[\pi_i(\vartheta, \overline{x}(\vartheta))] - \max\nolimits_\vartheta[Y(\vartheta)] \\ &= (1-\chi_1)\int\limits_{\mathcal{T}} \Big(w_1(\vartheta) - \pi_1^\theta(1-\varepsilon, \vartheta_2) + w_2(\vartheta_1, \varepsilon) - \pi_2^\theta(\vartheta)\Big)\, d\phi(\vartheta), \end{aligned} \tag{63}$$

which is strictly positive under the regularity conditions. □

Lemma 9. *In a two-firm collusion game, if the firms have globally interdependent valuations, then, generically, for any mechanism $\langle x, \overline{y}\rangle$ satisfying ex post budget balance, if x is not constant then for any sufficiently small open neighborhood $\mathcal{U} \ni (1,0)$ there exists $B \in (1, \infty)$ such that, for all $\theta \in \mathcal{U} \cap int(\Theta)$, $\frac{\partial x_1(\theta)}{\partial \theta_1} > \frac{1}{B} < -\frac{\partial x_1(\theta)}{\partial \theta_2}$.*

Proof. Any such mechanism $\langle x, y\rangle$ must satisfy Eq. 44 by Lemma 7. Consider the region $\mathcal{D} \equiv \big\{\theta \in \Theta : w_1(\theta) \geq w_2(\theta)\big\}$. Eq. 44 implies that $\frac{\partial x_i}{\partial \theta_i} = 0$ for both i wherever $w_1(\theta) = w_2(\theta)$, so x is constant along this curve. Furthermore, for any point $\theta \in \text{int}(\mathcal{D})$,

$$\left.\frac{\partial x_1}{\partial \theta_1}\right|_\theta = \int\limits_{\theta_2}^{1} -\mathbb{I}_{\{(\theta_1, s_2)\in\mathcal{D}\}} \frac{\partial^2 x_1}{\partial\theta_1 \partial s_2}\, ds_2. \tag{64}$$

By Eq. 44, the integrand on the right hand side of Eq. 64 is non-negative for all θ. At $\theta = (1,0)$, if the left hand side of Eq. 64 is zero, then the integrand on the right hand side must be zero pointwise, because otherwise the regularity condition $\frac{\partial x_1}{\partial \theta_1} \geq 0$ would be violated for some $\theta = (1, \theta_2)$. Hence x is constant on $\{1\} \times [0,1] \cap \mathcal{D}$. By monotonicity, x must also be constant on all of $\mathcal{D}$. Since a similar argument could be made for the point $\theta = (1,0)$, and by assumption x is not constant on all of Θ, to avoid any contradiction it must be that $\frac{\partial x_1}{\partial \theta_1}\big|_{(1,0)} > 0$. Since x is also twice differentiable, there must exist $B \in (1,\infty)$ and a neighborhood $\mathcal{U}$ of $(1,0)$ on which $\frac{\partial x_1}{\partial \theta_1} > \frac{1}{B}$. By similar arguments, the same is true of $-\frac{\partial x_1}{\partial \theta_2}$. □

Proof of Theorem 2 (page 17). By Lemma 7, under globally interdependent valuations a mechanism satisfying ex post budget balance must satisfy Eq. 44. Any constant allocation rule satisfies Eq. 44, but by Lemma 8 such a rule cannot be an optimal regular mechanism. So instead suppose there exists a regular mechanism satisfying Eq. 44, for which the allocation rule $\overline{x}$ is not constant.

Now, choose $\varepsilon > 0$ sufficiently small that $\mathcal{T} \equiv (1-\varepsilon, 1] \times [0, \varepsilon) \subset \mathcal{U}$ satisfies the above claim, and consider an alternative action rule x, such that $x_1(\theta) = 1$ for all $\theta \in \mathcal{T}$, and $x_1(\theta) = \overline{x}_1(\theta)$ otherwise. Then

$$r_1(\theta; x) = r_1(\theta; \overline{x}) + \mathbb{I}_{\{\theta \in \mathcal{T}\}} \left(w_1(\theta)\big(1 - \overline{x}_1(\theta)\big) - \int_{1-\varepsilon}^{\theta_1} \frac{\partial w_1(s_1, \theta_2)}{\partial s_1} \big(1 - \overline{x}_1(s_1, \theta_2)\big) ds_1 \right), \tag{65}$$

$$r_2(\theta; x) = \mathbb{I}_{\{\theta \notin \mathcal{T}\}} \left(r_2(\theta; \overline{x}) + \mathbb{I}_{\{\theta_1 > 1-\varepsilon\}} \left(w_2(\theta_1, 0)\overline{x}_2(\theta_1, 0) + \int_0^{\varepsilon} \frac{\partial w_2(\theta_1, s_2)}{\partial s_2} \overline{x}_2(\theta_1, s_2) ds_2 \right) \right). \tag{66}$$

Choose

$$h_2(\theta_1) \equiv h_2^*(\theta_1; \overline{x}) + \mathbb{I}_{\{\theta_1 > 1-\varepsilon\}} \left(w_2(\theta_1, 0)\overline{x}_2(\theta_1, 0) + \int_0^{\varepsilon} \frac{\partial w_2(\theta_1, s_2)}{\partial s_2} \overline{x}_2(\theta_1, s_2) ds_2 \right) \tag{67}$$

and $h_1 = h_1^*(\cdot; \overline{x})$. Then $Y(\theta) = \sum_i h_i^*(\theta_{-i}; \overline{x}) - R(\theta; \overline{x}) + c = c$ for all $\theta \in \Theta \backslash \mathcal{T}$, while, for $\theta \in \mathcal{T}$,

$$\begin{aligned}
Y(\theta) &= h_1^*(\cdot; \overline{x}) + h_2^*(\theta_1; \overline{x}) - r_1(\theta; \overline{x}) + w_2(\theta_1, 0)\overline{x}_2(\theta_1, 0) - w_1(\theta)\big(1 - \overline{x}_1(\theta)\big) \\
&\quad + \int_0^{\varepsilon} \frac{\partial w_2(\theta_1, s_2)}{\partial s_2} \overline{x}_2(\theta_1, s_2)\, ds_2 + \int_{1-\varepsilon}^{\theta_1} \frac{\partial w_1(s_1, \theta_2)}{\partial s_1} \big(1 - \overline{x}_1(s_1, \theta_2)\big)\, ds_1 \\
&= \int_{\theta_2}^{\varepsilon} \frac{\partial w_2(\theta_1, s_2)}{\partial s_2} \overline{x}_2(\theta_1, s_2) ds_2 + \int_{1-\varepsilon}^{\theta_1} \frac{\partial w_1(s_1, \theta_2)}{\partial s_1} \overline{x}_2(s_1, \theta_2) ds_1 - \big(w_1(\theta) - w_2(\theta)\big)\overline{x}_2(\theta)
\end{aligned} \tag{68}$$

where the last equality holds because the mechanism using $\overline{x}$ is ex post budget balanced. By Lemma 9, $\overline{x}_2$ has uniformly bounded first derivatives on $\mathcal{T}$, so the integral terms are on the order of $\varepsilon \overline{x}_2(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2})$, while the non-integral terms are on the order of $\overline{x}_2(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2})$. Hence for ε small the aggregate transfers for $\theta \in \mathcal{T}$ are approximately $Y(\theta) \approx -\big(w_1(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}) - w_2(\frac{\varepsilon}{2}, \frac{\varepsilon}{2})\big)\overline{x}_2(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2})$, which is strictly negative under the regularity conditions. Hence

$$\mathbb{E}[Y(\vartheta)] = \int_T \left[\begin{aligned} &\int_{\vartheta_2}^{\varepsilon} \frac{\partial w_2(\vartheta_1, s_2)}{\partial s_2} \overline{x}_2(\vartheta_1, s_2)\, ds_2 + \int_{1-\varepsilon}^{\vartheta_1} \frac{\partial w_1(s_1, \vartheta_2)}{\partial s_1} \overline{x}_2(s_1, \vartheta_2)\, ds_1 \\ &- \big(w_1(\vartheta) - w_2(\vartheta)\big)\overline{x}_2(\vartheta) \end{aligned} \right] d\phi(\vartheta). \tag{69}$$

At the same time, the increase in aggregate payoff is

$$\sum_i \mathbb{E}[\pi_i(\vartheta, x(\vartheta))] - \sum_i \mathbb{E}[\pi_i(\vartheta, \overline{x}(\vartheta))] = \int_T \big(w_1(\vartheta) - w_2(\vartheta)\big)\overline{x}_2(\vartheta)\, d\phi(\vartheta). \tag{70}$$

Hence the improvement in the value of the mechanism is

$$\sum_i \mathbb{E}[\pi_i(\vartheta, x(\vartheta))] - \sum_i \mathbb{E}[\pi_i(\vartheta, \overline{x}(\vartheta))] + \mathbb{E}[Y(\vartheta)]$$
$$= \int\limits_T \left[\int\limits_{\vartheta_2}^{\varepsilon} \frac{\partial w_2(\vartheta_1, s_2)}{\partial s_2} \overline{x}_2(\vartheta_1, s_2)\, ds_2 + \int\limits_{1-\varepsilon}^{\vartheta_1} \frac{\partial w_1(s_1, \vartheta_2)}{\partial s_1} \overline{x}_2(s_1, \vartheta_2)\, ds_1 \right] d\phi(\vartheta) > 0. \tag{71}$$

Hence $\overline{x}$ cannot be the allocation rule for an optimal regular mechanism. □

References

Dilip Abreu, David Pearce, and Ennio Stacchetti. Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251–69, June 1986.

Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58(5):1041–1063, September 1990.

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin, 2nd edition, 1999.

Masaki Aoyagi. Bid rotation and collusion in repeated auctions. *Journal of Economic Theory*, 112(1):79–105, September 2003.

Masaki Aoyagi. Efficient collusion in repeated auctions with communication. *Journal of Economic Theory*, 134(1):61–92, May 2007.

John Asker. A study of the internal organization of a bidding cartel. *American Economic Review*, 100(3): 724–762, June 2010.

Susan Athey and Kyle Bagwell. Optimal collusion with private information. *RAND Journal of Economics*, 32(3):428–465, Autumn 2001.

Susan Athey and Kyle Bagwell. Collusion with persistent cost shocks. *Econometrica*, 76(3):493–540, April 2008.

Susan Athey and David A. Miller. Efficiency in repeated trade with hidden valuations. *Theoretical Economics*, 2(3):299–354, September 2007.

Susan Athey and Ilya Segal. An efficient dynamic mechanism. Working paper, February 2007. URL http://kuznets.harvard.edu/~athey/EfficientDynamic.pdf.

Susan Athey, Kyle Bagwell, and Chris Sanchirico. Collusion and price rigidity. *Review of Economic Studies*, 71(2):317–349, April 2004.

Erik J. Balder. On the existence of optimal contract mechanisms for incomplete information principal-agent models. *Journal of Economic Theory*, 68(1):133–148, January 1996.

Salvador Barberà and Matthew O. Jackson. Strategy-proof exchange. *Econometrica*, 63(1):51–87, January 1995.

Pierpaolo Battigalli. Strategic independence and perfect Bayesian equilibria. *Journal of Economic Theory*, 70(1):201–234, July 1996.

Dirk Bergemann and Stephen Morris. Robust mechanism design. *Econometrica*, 73(6):1771–1813, November 2005.

Dirk Bergemann and Juuso Välimäki. The dynamic pivot mechanism. *Econometrica*, 78(2):771–789, march 2010.

Sushil Bikhchandani. The limits of ex post implementation revisited. Working paper, January 2005. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=605862.

Andreas Blume and Paul Heidhues. Private monitoring in auctions. *Journal of Economic Theory*, 131(1): 179–211, November 2006.

Timothy F. Bresnahan. Competition and collusion in the American automobile industry: The 1955 price war. *Journal of Industrial Economics*, 35(4):457–482, June 1987.

Meghan Busse. Firm financial condition and airline price wars. *RAND Journal of Economics*, 33(2):298–318, Summer 2002.

Kim-Sau Chung and Jeffrey C. Ely. Ex-post incentive compatible mechanism design. Working paper, May 2006. URL http://sites.google.com/site/kimsauchung/expost.pdf.

Kim-Sau Chung and Jeffrey C. Ely. Foundations of dominant strategy mechanisms. *Review of Economic Studies*, 74(2):447–476, April 2007. Working paper.

Claude d'Aspremont, Jacques Crémer, and Louis-André Gérard-Varet. Balanced Bayesian mechanisms. *Journal of Economic Theory*, 115(2):385–396, April 2004.

Natalia Fabra and Juan Toro. Price wars and collusion in the Spanish electricity market. *International Journal of Industrial Organization*, 23(3–4):155–181, April 2005.

Chaim Fershtman and Ariel Pakes. A dynamic oligopoly with collusion and price wars. *RAND Journal of Economics*, 31(2):207–236, Summer 2000.

Drew Fudenberg and David K. Levine. Efficiency and observability with long-run and short-run players. *Journal of Economic Theory*, 62(1):103–135, February 1994.

Drew Fudenberg and Jean Tirole. Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260, April 1991.

Drew Fudenberg, David K. Levine, and Eric S. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1039, September 1994.

Dennis Gärtner and Armin Schmutzler. Merger negotiations and ex-post regret. *Journal of Economic Theory*, 144(4):1636–1664, July 2009.

Edward J. Green and Robert H. Porter. Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):87–100, January 1984.

Joseph Greenberg, Robert Mackay, and Nicolaus Tideman. Some limitations of the Groves-Ledyard optimal mechanism. *Public Choice*, 29(2):129–137, March 1977.

Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, July 1973.

Theodore Groves and John O. Ledyard. Some limitations of demand revealing processes. *Public Choice*, 29 (2):107–124, March 1977a.

Theodore Groves and John O. Ledyard. Reply to comments by Tideman and Tullock and Greenberg, Mackay and Tideman on some limitations of demand revealing processes. *Public Choice*, 29(2):139–143, March 1977b.

Kathleen M. Hagerty and William P. Rogerson. Robust trading mechanisms. *Journal of Economic Theory*, 42(1):94–107, June 1987.

Joseph E. Harrington and Andrzej Skrzypacz. Collusion under monitoring of sales. *RAND Journal of Economics*, 38(2):314–331, Summer 2007.

Johannes Hörner and Julian S. Jamison. Collusion with (almost) no information. *RAND Journal of Economics*, 38(3):804–822, January 2008.

Leonid Hurwicz and Mark Walker. On the generic nonoptimality of dominant-strategy allocation mechanisms: A general theorem that includes pure exchange economies. *Econometrica*, 58(3):683–704, May 1990.

Philippe Jehiel, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William R. Zame. The limits of ex-post implementation. *Econometrica*, 74(3):585–610, May 2006.

Michihiro Kandori and Ichiro Obara. Efficiency in repeated games revisited: The role of private strategies. *Econometrica*, 74(2):499–519, March 2006.

John E. Kwoka, Jr. The price effects of bidding consipiracies: Evidence from real estate auction 'knockouts'. *Antitrust Bulletin*, 42(2):503–516, Summer 1997.

Jean-Jacques Laffont and Eric S. Maskin. A differential approach to dominant strategy mechanisms. *Econometrica*, 48(48):1507–1520, September 1980.

Margaret C. Levenstein. Price wars and the stability of collusion: A study of the pre-World War I bromine industry. *Journal of Industrial Economics*, 45(2):117–137, June 1997.

Liqun Liu and Guoqiang Tian. A characterization of the existence of optimal dominant strategy mechanisms. *Review of Economic Design*, 4:205–218, 1999.

Giuseppe Lopomo, Robert C. Marshall, and Leslie M. Marx. Inefficiency of collution at English auctions. *Contributions to Theoretical Economics*, 5(1):Article 4, 2005.

R. Preston McAfee and John McMillan. Bidding rings. *American Economic Review*, 82(3):579–599, June 1992.

Claudio Mezzetti. Mechanism design with interdependent valuations: efficiency. *Econometrica*, 72(5): 1617–1626, September 2004.

Paul R. Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, March 2002.

David A. Miller. Attainable payoffs in repeated games with interdependent private information. Working paper, January 2009. URL http://dss.ucsd.edu/~d9miller/papers/d9miller_Folk.pdf.

Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trade. *Journal of Economic Theory*, 29(2):265–281, April 1983.

David C. Parkes and Satinder Singh. An MDP-based approach to online mechanism design. In *Proceedings of the 17th Annual Conference on Neural Information Processing Systems (NIPS'03)*, 2003.

Alessandro Pavan, Ilya Segal, and Juuso Toikka. Dynamic mechanism design: Revenue equivalence, profit maximization and information disclosure. Working paper, August 2009. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1620662.

Martin Pesendorfer. A study of collusion in first-price auctions. *Review of Economic Studies*, 67(3):381–411, July 2000.

Robert H. Porter. A study of cartel stability: The Joint Executive Committee, 1880–1886. *Bell Journal of Economics*, 14(2):301–314, Autumn 1983.

Julio J. Rotemberg and Garth Saloner. A supergame-theoretic model of price wars during booms. *American Economic Review*, 76(3):390–407, June 1986.

Yuliy Sannikov and Andrzej Skrzypacz. Impossibility of collusion under imperfect monitoring with flexible production. *American Economic Review*, 97(5):1794–1823, December 2007.

Ran Shao and Lin Zhou. Optimal allocation of an indivisible good. Working paper, May 2008. URL http://ranshao.info/Documents/Optimal%20Allocation%20of%20an%20Indivisible%20Good%20%20%28May%202008%29.pdf.

Andrzej Skrzypacz and Hugo A. Hopenhayn. Tacit collusion in repeated auctions. *Journal of Economic Theory*, 114(1):153–169, January 2004.

Margaret E. Slade. Vancouver's gasoline-price wars: An empirical exercise in uncovering supergame strategies. *Review of Economic Studies*, 59(2):257–276, April 1992.

Jonathan Weinstein and Muhamet Yildiz. A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, 75(2):365–400, March 2007.