Confining the Coase Theorem: Contracting, Ownership, and Free-riding*

Tore Ellingsen†

Stockholm School of Economics and Norwegian School of Economics

Elena Paltseva‡

Stockholm School of Economics and New Economic School

This version: July 2015

Forthcoming in *The Review of Economic Studies*

Abstract

If individuals own the right to take any action that they please, and are free to contract about behavior, will outcomes be efficient in all situations? That is, does the Coase theorem hold? We study this classic question through the lens of a non-cooperative model of contract negotiations, considering both compulsory and voluntary participation in negotiations. In either case, we find that all consistent equilibria of the contracting game are efficient in the case of two players. But if participation is voluntary, and there are more than two players, there are situations in which all consistent equilibria are inefficient. Specifically, the provision of public goods tends to be inefficiently low due to strategic abstention from contracting. Free-riding on others' agreements can be avoided if individuals do not own all their actions. When actions involve the use of assets, efficient action ownership may correspond to collective rather than individual asset ownership.

JEL Codes: C78, D23

Keywords: Contracting, Renegotiation, Ownership, Property rights

*A previous version of this paper was circulated under the title "Non-cooperative Contracting." Financial support from the Torsten and Ragnar Söderberg Foundation is gratefully acknowledged. Thanks to Geir Asheim, Vincent Crawford, Avinash Dixit, Bård Harstad, Matthew Jackson, Eirik Gaard Kristiansen, Margaret Meyer, Paul Milgrom, Joel Watson, and Peyton Young for helpful conversations. Marco Ottaviani and four anonymous referees provided extremely valuable comments on earlier drafts. Thanks also to many seminar participants. We are responsible for remaining errors.

†Address: Department of Economics, Stockholm School of Economics, Box 6501, S—11383 Stockholm, Sweden. Email: gte@hhs.se

‡Address: SITE, Stockholm School of Economics, Box 6501, S—11383 Stockholm, Sweden. Email: elena.paltseva@hhs.se.

1 Introduction

If rational and fully informed individuals are free to contract, property rights are completely specified, and there are no transaction costs, will they always be able to reach an agreement to behave efficiently? According to Coase (1960), they will.[1] By and large, formal contract theory embraces Coase's view (e.g., Bolton and Dewatripont, 2005, page 7). However, the so-called Coase theorem is an informal argument based on a few stylized examples rather than a precise mathematical result, and there are at least two reasons to doubt it. First, Coase's stylized examples are special; they are all concerned with unilateral externalities among two parties, such as straying cattle that destroy crops on a neighbor's land. In general, externalities can be multilateral and non-additive. Second, Coase does not consider the detailed process of proposing and accepting contracts. In essence, he assumes that in the absence of transaction costs any efficiency gains will be realized through appropriate transfer payments. For there to be a theorem rather than merely a presumption, the efficiency of the agreement should be deduced from an analysis of the contracting process. A modern formulation of the question could therefore be: *If the contracting process is adequately formulated as a non-cooperative game, will players arrive at efficient agreements in all the (appropriately selected) equilibria of this game?*

In a nutshell, our answer is that two individuals will always be able to arrive at an efficient outcome, but more than two individuals will not always be able to do so. The number of individuals matters because it affects the scope for free-riding on others' agreements. With only two individuals, there obviously cannot be any agreement if one player fails to negotiate. In this case, we show that it is always possible to construct a suitable equilibrium in which willingness to negotiate is being sufficiently rewarded to bring both parties to the table. With more than two individuals, we find that it can be more profitable to let others agree and to catch a free ride. By emphasizing the distinction between participating in contracting and accepting a contract, we provide a new foundation for a variety of literatures on free-riding. We also demonstrate that the extent of the inefficiency is affected by the initial allocation of property rights. Indeed, for any situation, there is always some allocation of property rights, possibly involving collective asset ownership, that ensures an efficient outcome.

To introduce our approach, let us begin by explaining what we mean by a contract and a contract negotiation. According to *Encyclopaedia Britannica*, a contract is

> *a promise enforceable by law. The promise may be to do something or to refrain from doing something. The making of a contract requires the mutual assent of two or more persons, one of them ordinarily making an offer and*

[1] Coase did not think that the case of zero transaction costs is realistic; he used it as a benchmark to emphasize the need to study positive transaction costs.

another accepting. If one of the parties fails to keep the promise, the other is entitled to legal recourse.

Formally, we thus define a contract as a mutually agreed mapping from action profiles to monetary transfers. These monetary transfers regulate compensation both when parties comply with the contract's intention and when they deviate from it. Note that our definition is sufficiently general to encompass a wide range of commonly observed contracts, such as: (a) One party may promise to take a specific action in return for monetary payment, accepting to pay a penalty if any other action is taken. (b) One party may agree to work for the other, in return for a payment that depends on the amount of effort that is exerted – such as a piece rate or a bonus scheme. (c) Several parties may agree on how to divide up an income as a function of how each of them contributes to generating it. Even more precisely, a contract specifies monetary transfers as a function of actions taken in some *situation*, or action game, G. In the parlance of mechanism design theory, a contract is an *indirect mechanism* that all parties have agreed to. However, following the lead of Jackson and Wilkie (2005), we do not admit an outside planner to shape the mechanism. Contract formulation is left entirely to the players themselves, and no player has a privileged position in this respect.

Contract negotiations have four stages. At Stage 0, players learn the situation G and decide whether to take part in negotiations or not. At Stage 1, players make contract proposals. At Stage 2 players decide which contract proposal(s) to sign. A contract proposal becomes a valid contract if and only if it is signed by all players who may pay or receive transfers under the contract. At Stage 3, upon seeing the outcome of the negotiations, players play G modified by the agreed transfers. This four-stage game is denoted $\Gamma^V(G)$. If participation in negotiations is mandatory, as is implicitly assumed in parts of the contracting literature, there are only the last three stages. We denote this three-stage game $\Gamma(G)$.

A central observation is that the signing stage involves a coordination game, in which each concerned player may become pivotal. Thus, if the no-contract outcome is expected to be less favorable for each player, signing by all is incentive compatible. At first blush, it thus seems almost trivially easy to sustain cooperation. But upon closer inspection, there are two problems with this logic. One problem is that there could also be other equilibria, in which players fail to propose or sign desirable contracts because they expect others not to sign them. That is, while the set of equilibria might include efficient outcomes, inefficient outcomes are included too. We address this problem by invoking a powerful solution concept. Specifically, rather than basing our assessment on the large set of subgame-perfect equilibria, we judge Coase by the set of *consistent* equilibria (Bernheim and Ray, 1989). Consistency applies Pareto-dominant selection recursively and thus promotes efficiency to the maximal extent compatible with sequential rationality and equilibrium, respecting the spirit of Coase's argument. The second problem is that the logic takes for granted that players are willing to participate in negotiations to begin with. As it turns out, this assumption is sometimes unwarranted, and strategic non-participation could prevent cooperation.

In the benchmark case where each player owns her full set of actions, our main results can be summarized as follows: (i) Under mandatory participation ($\Gamma(G)$), the negotiation game always admits a large range of strategy profiles of G to be played in a subgame-perfect Nash equilibrium. Some of the equilibrium outcomes are efficient, but others are inefficient (Theorem 1). (ii) If we additionally insist that equilibria should be consistent, the solution set shrinks. Many inefficient outcomes fail the consistency criterion, and in the case of two players ($n = 2$) only efficient outcomes remain (Theorem 2). (iii) Under voluntary participation ($\Gamma^V(G)$), all consistent equilibria are also efficient in the case of $n = 2$ (Theorem 5). That is, with two players, the Coase theorem holds. (iv) However, for $n > 2$, there is a large and relevant class of situations G for which *all* consistent equilibrium outcomes of $\Gamma^V(G)$ are inefficient (Theorem 6). That is, under voluntary participation, the Coase theorem does not hold in the case of more than two players. The reason is that it can be too tempting to free-ride on others' agreements, since those others cannot credibly threaten to behave inefficiently upon such strategic non-participation. To sum up, the Coase theorem is sound in two-player situations, but not otherwise.

The model admits straightforward definitions of ownership of actions as well as assets. For actions taken by oneself, ownership is the right to take the action for free. For actions taken by another party, ownership is the right to release that party from penalties that are otherwise associated with the action. For assets, we say that the asset individually owned if an individual owns all asset-relevant actions, except a default (no-use) action profile. For example, the owner of a plot of land may own all actions that involve the land in any way, such as the rights to farm the land or walk across it. Conversely, the land is said to be collectively owned if more than one individual must consent to any specific action, or if more than one individual has land-use rights. In the latter case, actions may be individually owned, yet the asset is collectively owned.

Exploiting these definitions, we obtain two further insights. (v) In all situations G there is some allocation of *action ownership*, individual or collective, that guarantees efficient outcomes (Theorem 7). (vi) However, there is not always any *individual asset ownership* regime that guarantees efficiency (Theorem 8). Collective asset ownership might be required to prevent socially harmful actions – either by the owner or by others that the owner grants the right to behave harmfully. Again, the ultimate problem concerns voluntary participation: Under individual asset ownership, there will sometimes be free-riding with respect to entering negotiations about the regulation of externalities. Here, then, is a possible rationale for why assets are sometimes owned by legal entities such as governments or firms rather than by individuals.

We also investigate whether our results are robust to various changes of the contracting game. In particular, we address the objection that non-participation in negotiation here means that a player can insulate herself fully against all transfers off and on the equilibrium path. In reality, it may be difficult not to be the subject of a promise. We therefore also study the impact of allowing unilateral promises in addition to multilateral contracts. Overall, the insights remain broadly intact (Theorems 9 - 11). Moreover, our analysis of unilateral promises suggests a potential justification for why courts enforce such promises in case of reliance by the promisee – the promissory estoppel doctrine – but do not necessarily enforce frivolous promises.

The paper is organized as follows. Section 2 provides a simple introductory example. Section 3 sets up our formal apparatus. Section 4 contains our main results for the case in which players have to negotiate – that is, participation in negotiations is mandatory. Section 5 contains the corresponding set of results under voluntary participation in negotiations. Section 6 relaxes the assumption that each player has the right to all her available actions. Section 7 considers the robustness of our results to the possibility that players may make unilateral promises in addition to their agreements. Section 8 describes our contribution's relation to the existing literature. Section 9 concludes.

2 Example

The following simple example, involving multilateral and non-additive externalities, suggests why a Coase theorem holds in the case of two players and provides a stepping stone for the remainder of the paper. Two ranchers let their cattle graze on the same field. The animals sometimes stray. Suppose both ranchers suffer equally when cattle stray, and that either or both of them can take action to contain the cattle, let's say by herding them (take action H). The alternative is to be lazy and do nothing (take action L). The private cost of herding is 5. If only one of them engage in herding, the benefit to each is 4 (total benefit is 8). If both engage, the benefit to each is 7, so the total benefit is 14. Clearly, the best outcome arises when both engage, since this yields a total net gain of $2 \cdot 7 - 2 \cdot 5 = 4$ as compared to the net gain of $2 \cdot 4 - 5 = 3$ in case only one of them engages. However, without a binding contract, neither will perform the herding; the private cost of herding, 5, is always larger than the private gain, which is either 4 or 3. The example boils down to the Prisoners' dilemma game of Jackson and Wilkie's (2005) Example 1, reproduced in Figure 1.

Now, suppose the ranchers are at liberty to propose and accept contracts. A contract specifies four transfers – positive if rancher 1 pays, negative if rancher 2 pays – one transfer for each of the four cells in Figure 1. Let there be a first stage at which both simultaneously propose contracts and a second stage at which each is at liberty to sign

	H	L
H	$2, 2$	$-1, 4$
L	$4, -1$	$0, 0$

Figure 1: A Prisoners' Dilemma

any one of the proposals, or none.[2] A court will costlessly enforce all agreed transfers. Suppose one of the ranchers make the proposal $(0, -(2+\varepsilon), 2+\varepsilon, 0)$ and both sign. Then, the original situation (Figure 1) is transformed into the game in Figure 2.

	H	L
H	$2, 2$	$1+\varepsilon, 2-\varepsilon$
L	$2-\varepsilon, 1+\varepsilon$	$0, 0$

Figure 2: The modified game

If $\varepsilon > 0$, it is easy to check that (H,H) is a unique and strict equilibrium.

Let us sketch why (H,H) can also be sustained in equilibrium when both ranchers make simultaneous contract proposals. Consider the following pair of strategies. At the proposal stage, each player makes the proposal $(0, -(2 + \varepsilon), 2 + \varepsilon, 0)$. At the signing stage, each rancher signs rancher 1's proposal if and only if it is this expected proposal. Conversely, each rancher signs rancher 2's proposal if and only if it is the expected proposal and rancher 1 *fails* to make the expected proposal. This strategy profile forms a subgame-perfect equilibrium, as is easily checked.[3]

The above contracting game also has inefficient equilibria. One inefficient equilibrium is that no contract is ever signed. (If each player expects the opponent not to sign any contract, there is nothing to gain by signing oneself.) However, if we impose consistency (a form of renegotiation-proofness), such inefficient equilibria vanish.

On the other hand, we shall show that inefficient continuation equilibria would constitute useful threats in the case of more than two players and endogenous participation. Due to such threats, efficient outcomes are attainable in subgame-perfect equilibrium even if players are free not to participate in negotiations. But since inefficient continuation equilibria are often not consistent, the very argument that supported efficiency in the case of two players undermines it in the case of more players.[4]

[2] Note that the proposer also has to sign in order to be bound by the contract.

[3] Recall that if no contract is signed by both ranchers, each rancher ends up with a payoff of 0. Consider unilateral deviations from the posited strategy profiles, beginning at the signing stage. (i) If proposals are as expected, we have already argued that rancher 2 is better off signing rancher 1's proposal (expecting that rancher 1 will do so). Likewise, rancher 1 is better off signing the own proposal if expecting that rancher 2 will sign. (ii) If rancher 1 deviates at the proposal stage, the expectation is that rancher 2's proposal will be signed, so this does not benefit rancher 1. If only rancher 2 deviates at the proposal stage, it does not matter, as rancher 1's contract proposal will still be signed.

[4] Another problem emphasized by Jackson and Wilkie (2005) arises if players can make unilateral

Finally, in Section 6 we shall consider situations that correspond to cases in which one rancher has the right to deny or allow some of the other rancher's actions, for example through property rights to the field.

3 Definitions

Let G be a finite n-player game, called "the action game."The set of players is denoted $N = \{1, ..., n\}$.

Actions: Player i's set of pure strategies is denoted X_i, and the set of all pure strategy profiles is $X = \times_i X_i$. Player i's set of mixed strategies is $\Delta(X_i)$, and the set of all mixed strategy profiles is $\Delta = \times_i \Delta(X_i)$. Generic elements of X_i, X, $\Delta(X_i)$, and Δ are denoted x_i, x, μ_i and μ respectively.

Preferences: Players' preferences are given by a von Neumann-Morgenstern utility function $U_i : \Delta \to \mathbb{R}$. As usual, we let $U_i(x)$ denote Player i's utility under the mixed strategy profile putting all the probability on the pure strategy profile x.

Ownership: Initially, we assume that players own their entire action sets, in the sense that there are no penalties imposed on any action. We relax this "universal self-ownership" assumption in Section 6.

Contracts: Before choosing their strategies from Δ, the players engage in contracting. A contract t specifies for each pure strategy profile x a vector of net transfers. That is, a contract is an n-dimensional function $t : X \to \mathbb{R}^n$, where the i'th component $t_i(x)$ denotes the transfer from Player i (which may be positive or negative). In this way, Player i's final payoff under contract t and strategy profile x is denoted

$$\pi_i(x) = U_i(x) - t_i(x).$$

Let $\pi(x) = \{\pi_1(x), ..., \pi_n(x)\}$. For each strategy profile x, a contract must satisfy the budget balance constraint $\sum_{i=1}^{n} t_i(x) = 0$.[5] Let F be the set of feasible (i.e., budget-balanced) contracts. Thus, for each x, the set of payoff profiles that may be induced by some contract $t \in F$ is

$$\Pi(x) = \left\{ \pi(x) : \sum_{i \in N} \pi_i(x) = \sum_{i \in N} U_i(x) \right\}.$$

promises in addition to their standard contract proposals. Specifically, suppose that Player 1 proposes the agreement described above (Figure 2), and that Player 2 merely makes a unilateral promise to transfer slightly above 1 in case Player 1 plays H. At the signing stage, Player 2 will then not sign Player 1's proposal, and go on to earn 3 instead of 0. The only way for Player 1 to avoid this outcome is to also make unilateral promises, and making them conditional on Player 2 *not* signing 1's contract. Section 7 explores this issue.

[5] Of course, budget balance implies that the range of C_i could alternatively be expressed as $\mathbb{R}^{n-1}$. Not imposing budget balance, allowing the presence of a "source" or a "sink", would beg the question why these players are not modelled explicitly.

That is, a contract induces a re-distribution of the payoffs associated with each strategy profile. Similarly, for a mixed strategy profile μ, the set of feasible payoffs is

$$\Pi(\mu) = \left\{ \pi(\mu) : \sum_{i \in N} \pi_i(\hat{\mu}) = \sum_{i \in N} U_i(\mu) = \sum_{i \in N} \left[\sum_{x \in \text{supp}(\mu)} \mu(x) U_i(x) \right] \right\}.$$

We say that a strategy profile x is *efficient* if it maximizes total payoff, that is, if it belongs to the set $\{\arg\max \sum_{i \in N} U_i(x).\}$ We shall often be interested in efficient outcomes that yield each player at least what the player can guarantee herself. For this purpose, let

$$v_i(G) = \max_{\mu_i} \min_{\mu_{-i}} U_i(\mu_i, \mu_{-i})$$

denote Player i's maximin payoff, let

$$u_i(G) = \min_{\mu \in NE(G)} U_i(\mu),$$

(where $NE(G)$ denotes the set of Nash equilibria of G) be Player i's lowest payoff in any Nash equilibrium of G, and let $u_i^{ud}(G)$ denote Player i's lowest payoff in any Pareto-undominated equilibrium of G.

Finally, the following definition will prove useful.

Definition 1 *Player i is said to be affected by contract t if and only if $t_i(x) \neq 0$ for some x.*

Let N^t denote the set of players that are affected by contract t. Conversely, N/N^t are the players who are unaffected by t. Finally, let $\widetilde{G}(t)$ denote the game G modified by contract t – in other words, $\widetilde{G}(t)$ has G's strategy set X, but is played with utilities $\pi_i(x)$ instead of $U_i(x)$.

3.1 The contract negotiation game

To begin with, we assume that participation in negotiations is mandatory. The contract negotiation game has three stages.

At the proposal stage, henceforth called Stage 1, all players simultaneously make a contract proposal from the set F. (Without affecting the results, we could admit a no-proposal option too.) Let Player i's proposal be denoted τ_i.

At the signature stage, henceforth called Stage 2, each player first observes all the proposals.[6] Then, each player individually chooses at most one contract proposal to

[6] While this assumption is uncontroversial if $n = 2$, and is also realistic for negotiations in which all the negotiations occur around a single table, there are clearly many real negotiations that do not satisfy this particular aspect of zero transaction costs.

sign.[7] We realistically assume that players have no say about contracts that do not directly affect them.

Definition 2 *A contract t is said to be signed if and only if all affected players $j \in N^t$ have signed it.*

Any contract t is legally binding if and only if it is signed. Let N^S be the set of of players whose proposals were signed, and let $t^S = \sum_{i \in N^S} \tau_i$ be the effective transfers. In case no contract is signed, we say that $t^S = \varnothing$.

At the action stage, henceforth called Stage 3, each player observes all the signature decisions. Players go on to play $\widetilde{G}(t^S)$. Note that $\widetilde{G}(\varnothing) = G$.

We refer to the whole three stage game as $\Gamma(G)$. We initially focus on subgame-perfect Nash equilibria of $\Gamma(G)$.

4 Outcomes under mandatory participation

We begin by proving that any strategy profile, and any division of the associated surplus that gives each player i at least her worst equilibrium payoff of G, can be sustained in a subgame-perfect equilibrium of $\Gamma(G)$. The basic idea is that a player who is affected by a contract proposal may be able, by not signing it, to veto all outcomes that are worse for the player than the worst no-contract outcome; however, the player can be couched to sign any better contract proposal through the credible threat of a poor no-contract outcome in case the proposal is not signed.

Theorem 1 *A strategy profile $\hat{\mu} \in \Delta X$ and payoff profile $\pi(\hat{\mu}) \in \Pi(\hat{\mu})$ can be sustained in a subgame-perfect Nash equilibrium of $\Gamma(G)$ if $\pi_i(\hat{\mu}) \geq u_i(G)$ for all i.*

For expositional clarity we concentrate on the case of sustaining a pure strategy profile $\hat{x}$. The proof in case of a mixed strategy profile is relegated to the Appendix.[8]

Proof. Proceed in three steps: (i) Consider some candidate payoff vector. (ii) Show that there exists a contract t that implements these payoffs in an equilibrium of $\widetilde{G}(t)$. (iii) Show that $\Gamma(G)$ has a subgame-perfect equilibrium in which t is proposed and signed.

[7] As shown below, this is not a restrictive assumption. However, it is worth emphasizing the assumption that a player's own proposal does not in any way affect the player's ability to sign opponents' proposals. As indicated by Ellingsen and Miettinen (2008, 2014), it may be impossible to sustain efficient outcomes in stable equilibria if players become committed to not accepting outcomes that are worse than their own proposal.

[8] The only difference in the proof for a mixed strategy profile $\hat{\mu}$ is establishing existence of a system of feasible transfers defined on the support of $\hat{\mu}$ such that each player gets exactly the payoff $\pi_j(\hat{\mu})$ by playing any strategy in the support of $\hat{\mu}_j$ given that the other players play $\hat{\mu}_{-j}$. The definition of the transfers off the support of $\hat{\mu}$ and the rest of the proof are exactly similar to the case of sustaining a pure strategy profile.

Steps (i) and (ii) are easy. Consider any feasible profile of payoffs $\pi(\hat{x})$, such that $\pi_j(\hat{x}) \geq u_j(G)$. This profile can be implemented through the contract $\hat{t}$ specifying the following net transfers from each Player j,

$$\hat{t}_j(x) = \begin{cases} U_j(x) - \pi_j(x) & \text{if } x = \hat{x}; \\ (n-1)h & \text{if } x_j \neq \hat{x} \text{ and } x_{-j} = \hat{x}_{-j}; \\ -h & \text{if } x_j = \hat{x}_j \text{ and } |\{k : x_k \neq \hat{x}_k\}| = 1; \\ 0 & \text{otherwise,} \end{cases}$$

where

$$h = 1 + \max_{i,x',x''} \left[U_i(x') - U_i(x'') \right].$$

Note that $\sum_{j=1}^{N} \hat{t}_j(x) = 0$ for all strategy profiles x. To see that the contract ensures $\hat{x}$ to be an equilibrium of $\widetilde{G}(\hat{t})$, observe that, for all players j and strategies x_j,

$$U_j(\hat{x}) - \hat{t}_j(\hat{x}) \geq U_j(x_j, \hat{x}_{-j}) - \hat{t}_j(x_j, \hat{x}_{-j}),$$

or equivalently

$$\begin{aligned} \pi_j(\hat{x}) &\geq U_j(x_j, \hat{x}_{-j}) - (n-1)h \\ &> u_j(G), \end{aligned}$$

where the first inequality uses the definition of $\hat{t}_j(x)$ and the second uses the definition of h together with the facts that $n \geq 2$, $U_j(x_j, \hat{x}_{-j}) \leq \max_x U_j(x)$, and $u_j(G) \geq \min_x U_j(x)$.

It remains to show that $\hat{t}$ may be proposed and signed in an equilibrium of $\Gamma(G)$. This is slightly tedious, as we have to specify complete strategy profiles, and there are many kinds of off-equilibrium nodes. Consider the profile of strategies:

Stage 1: Each player i makes the proposal $\tau_i = \hat{t}$.[9]

Stage 2. If $\tau_1 = \hat{t}$, each player signs τ_1. If $\tau_1 \neq \tau_2 = \hat{t}$, each player signs τ_2, and so on. If no player offers $\hat{t}$, let players' signing decisions cohere with some subgame-perfect equilibrium strategy of the continuation game. If there are several such signing profiles, choose the one that is best for Player 1. If there are several of those, choose the one that is best for Player 2, etc. (As each continuation game is finite, such a signing profile always exists.)

Stage 3. (i) If some proposal $\tau_i = \hat{t}$ was signed, play $\hat{x}$. (ii) if $\tau_i = \hat{t}$ is signed by all but Player k, play Player k'th worst Nash equilibrium in the resulting subgame $\widetilde{G} = G$.[10] (iii) In all other situations, play the worst Nash equilibrium of $\widetilde{G} = G$ from the perspective of Player 1 (if there are multiple such equilibria, play the worst of them from the perspective

[9] As discussed below, we use this construction in which all players' contract proposals are identical for simplicity only.

[10] Caveat: There could be several such equilibria, but since it does not matter which of them is played, we skip devising a selection.

of Player 2, etc.).

Let us show that no player will ever find it optimal to deviate from the suggested equilibrium path, starting with Stage 3 and moving forwards.

At Stage 3, if $t^S = \hat{t}$, $\hat{x}$ represents a Nash equilibrium, as is already shown. In all other situations, the suggested behavior likewise forms a Nash equilibrium of $\widetilde{G}(t^S)$, so no player has any incentive to unilaterally deviate.

At Stage 2, consider first the branch along which $\tau_1 = \hat{t}$. Then, a unilateral deviation by Player k (not to sign τ_1) entails $t^S = \varnothing$, and Player k's worst equilibrium of G being played at Stage 3. Since $u_k(G) \leq \pi_k$, the deviation is not profitable. This takes care of deviations on the equilibrium path. Off the path, an analogous argument applies along the branch $\tau_1 \neq \tau_2 = \hat{t}$. Finally, along all other off-equilibrium branches, play accords with a subgame-perfect equilibrium, so no unilateral deviation is profitable.

At Stage 1, only Player 1 deviations $\tau_1 \neq \hat{t}$ affect the subsequent play. After such a deviation, $\tau_2 = \hat{t}$ will be signed instead, entailing exactly the same outcome as if Player 1 does not deviate. Thus, this is not a profitable deviation, concluding the proof. ■

Three observations are in order. First, there are many other strategy profiles of $\Gamma(G)$ that can sustain the same set of equilibrium outcomes. Specifically, it is not necessary to have players propose identical contracts. The important feature is that the player whose contract is supposed to be signed in equilibrium (here, Player 1) is deterred from deviating by the expectation that the deviation will be punished through coordination on some other contract proposal that is no more attractive to the player. Second, the transfers that are used to sustain the equilibria are of the same order of magnitude as are the payoffs in G. Each Player i can be induced to take the desired action through an incentive that does not exceed the difference between the highest and lowest payoff in G (as is clear from the definition of h). Third, the set of strategy profiles that may be sustained includes all efficient profiles, or more formally:

Remark 1 *Any efficient strategy profile $x^* \in X$ of any game G can be sustained in a subgame-perfect Nash equilibrium of $\Gamma(G)$.*

The remark follows immediately from Theorem 1 and the fact that for any efficient strategy profile x^* and any Nash equilibrium profile $\mu^{NE} \in NE(G)$,

$$\sum_{j=1}^{N} \pi_j(x^*) \geq \sum_{j=1}^{N} U_j(\mu^{NE}) \geq \sum_{j=1}^{N} u_j(G).$$

That is, the sum of payoffs at an efficient strategy profile x^* weakly exceeds the sum of payoffs at any Nash equilibrium of G, which in turn weakly exceeds the sum of the lowest payoffs players can earn in any Nash equilibrium of G. Hence, there is always a way to redistribute the sum of the payoffs at x^* to satisfy the conditions of Theorem 1.

The basic logic of Theorem 1 does not depend on the number of contracts that players are allowed to sign. A bit more precisely, suppose each player still proposes only one contract, but may sign as many proposals as she likes. Let $\Gamma^+(G)$ denote the corresponding contracting game.

Remark 2 *A strategy profile $\hat{x} \in X$ and payoff profile $\pi(\hat{x}) \in \Pi(\hat{x})$ such that $\pi_i(\hat{x}) \geq u_i(G)$ for each player i, can be sustained in a subgame-perfect Nash equilibrium of $\Gamma^+(G)$.*

The proof is closely analogous to the proof of Theorem 1 and hence omitted (but available on request). The intuition is plain enough: If all other players sign only a particular contract proposal $\tau_i = t$, a single player cannot expand the set of signed contracts beyond t. Note that this remark also implies that any efficient outcome that yields each player a payoff weakly above $u_i(G)$ can be sustained in a SPNE of $\Gamma^+(G)$.[11]

Above, we have identified a large set of equilibrium outcomes. Are there any others? For $n = 2$ the answer is negative.

Remark 3 *If $n = 2$, a strategy profile $\hat{x} \in X$ and payoff profile $\pi(\hat{x}) \in \Pi(\hat{x})$ can be sustained in a subgame-perfect Nash equilibrium of $\Gamma(G)$ if and only if $\pi_i(\hat{x}) \geq u_i(G)$ for all i.*

This extension of Theorem 1 is obvious, as each player can veto any contract by not signing it, forcing the play of G unmodified by transfers.

However, in some games with $n > 2$, it is possible to sustain equilibria of $\Gamma(G)$ in which some players' payoff is smaller than their worst equilibrium payoff of G. The reason is that the implementation in $\Gamma(G)$ of large-payoff non-equilibrium cells in G may involve contracts in which some players are not involved in transfers. Since their signature is not required, they cannot veto the agreement. Even if they do their best at Stage 3, they may be unable to sustain more than their maximin payoff. This observation will matter below.

4.1 Refinement: Consistency

In the context of contract negotiations with universal participation, it seems reasonable that the players should be able to coordinate on desirable equilibria; any inefficient contracts ought to be renegotiated. One way to capture this intuition is to impose the requirement that equilibria are consistent (Bernheim and Ray, 1989). In a one-stage

[11]Evidently, there are many other constellations of assumptions that one might consider. For example, may we still sustain efficient outcomes if contracts cannot specify negative transfers for anyone but the contract proposer, as in Jackson and Wilkie (2005)? If each player may only sign one contract, we can demonstrate that such a constraint on transfers precludes efficiency in some games. However, if players can sign multiple contracts, efficiency can be restored, albeit at the cost of some complexity (proof available on request).

game, the set of consistent equilibria coincides with the set of Pareto-undominated equilibria (i.e., the Pareto-frontier of the equilibrium set). In a finite multi-stage game, a consistent equilibrium is characterized recursively: any consistent equilibrium involves Pareto-undominated equilibria in all subgames, both on and off the equilibrium path. Conversely, any subgame-perfect equilibrium relying on the threat that deviations are punished through an inefficient continuation equilibrium fails the consistency criterion.

Applying the consistency refinement to the set of equilibria described in Theorem 1 sometimes gets rid of all the inefficient equilibria, while leaving a subset of the efficient equilibria intact. Specifically, as $u_i^{ud}(G)$ is Player i's lowest payoff in any consistent equilibrium of G (i.e., in any Pareto-undominated equilibrium of the one-shot game G), we have the following result.

Theorem 2 *Suppose $n = 2$. Then, a strategy profile $\hat{x} \in X$ and payoff profile $\pi(\hat{x}) \in \Pi(\hat{x})$ are supportable in a consistent equilibrium of $\Gamma(G)$ if and only if $\hat{x}$ is efficient and $\pi_i(\hat{x}) \geq u_i^{ud}(G)$ for both players.*

This is one part of our two-player Coase theorem (the second part concerns voluntary participation).

Proof. (This is merely a sketch, omitting many details that are analogous to those in the proof of Theorem 1.) Proceed recursively. At the last stage, if no contract has been signed, players will coordinate on an undominated equilibrium of G. Thus, each player i gets at least $u_i^{ud}(G)$ in this case. Hence, at the signing stage, it cannot be part of a consistent equilibrium that players sign a contract that is expected to entail a payoff below $u_i^{ud}(G)$ to any participant i in the ensuing subgame $\widetilde{G}$. That is, consistent equilibrium payoffs must exceed $u_i^{ud}(G)$. Conversely, any contract t that is expected to entail a payoff above $u_i^{ud}(G)$ to each participant i in the ensuing subgame $\widetilde{G}$ would only be signed in a consistent equilibrium if there is no alternative contract proposal t' with an induced consistent equilibrium profile which Pareto-dominates the equilibrium profile induced by t. In the latter case t' must be signed instead. We say that a proposal is efficient and consistent if it admits an efficient outcome in a consistent equilibrium. Finally, then, at the proposal stage, at least one player must make an efficient and consistent contract proposal yielding each player i at least $u_i^{ud}(G)$; otherwise, the deviation to such a contract would be profitable, as it would be signed in the consistent continuation. This proves that all efficient outcomes yielding at least $u_i^{ud}(G)$ to each player i can be sustained, while other outcomes cannot be sustained. ■

With more than two players, consistency does not imply efficiency. That is, in some situations G there are inefficient consistent equilibrium outcomes of $\Gamma(G)$(in addition to the efficient ones). Consider for example the three-player situation in Figure 3.

If there are no transfers, the game $\widetilde{G} = G$ has a single Nash equilibrium, namely (L,L,L) with a payoff of $(0, 0, 0)$. The efficient strategy profile is (H,H,H), and it is straight-

	H	L
H	$2,2,2$	$1,6,-10$
L	$6,1,-10$	$2,2,-3$
Player 3: H		

	H	L
H	$-1,-1,3$	$-1,0,-11$
L	$0,-1,-11$	$0,0,0$
Player 3:L		

Figure 3: Situation G admitting an inefficient consistent equilibrium of $\Gamma(G)$

forward to prove that this profile can be supported in a consistent equilibrium. Our claim here is that (H,L,H), despite being inefficient, is also a consistent equilibrium outcome of $\Gamma(G)$. For example, suppose that the associated contract between players 1 and 2 yields each of them 3.5, whereas Player 3 obtains -10. Why is it that Player 3 cannot propose a contract inducing (H,H,H) with associated Pareto-improving payoffs of, say, (4,4,-2)? The reason is that if players 1 and 2 were to sign that contract, Player 3's best response would be not to sign it, obtaining instead the Nash equilibrium payoff of G, namely $(0, 0, 0)$. Essentially, off the equilibrium path, Player 3 is better off by disrupting the contracting process than by facilitating a multilateral contract that players 1 and 2 prefer to their bilateral contract. Foreseeing this, players 1 and 2 prefer to stick to their bilateral contract.

This is not an isolated example; inefficient consistent equilibria frequently exists when the worst Pareto-undominated Nash equilibrium of G yields more than the maximin payoff for some player. Should we see these examples as violations of the Coase theorem? We prefer not to take a stand. First, it is conceivable that a further refinement of the solution concept would leave only the efficient outcomes. Second, we think that the case of mandatory participation in negotiations is not the most relevant for assessing the Coase theorem.

The above reasoning also helps to identify a sufficient condition for the efficiency of consistent equilibria.

Theorem 3 *Suppose $n > 2$, and $v_i(G) = u_i^{ud}(G)$ for all players i. Then all consistent equilibria of $\Gamma(G)$ are efficient.*

Proof. No player i would sign a contract entailing an equilibrium payoff $\pi_i < v_i(G)$. Thus, when $v_i(G) = u_i^{ud}(G)$ for all i, any contractually sustainable strategy profile must yield $\pi_i \geq u_i^{ud}(G)$ to each player i. But if so, there can be no inefficient consistent equilibrium; there would always be a profitable deviation at the proposal stage that would be signed by all the parties, by exactly same logic as in Theorem 2. ■

5 Voluntary Participation in Negotiations

So far, we have assumed that all players have to participate in negotiations. We now investigate how voluntary participation affects our analysis.

Suppose that each player, before the contract proposal stage, may decide whether or not to participate in contracting. That is, the player may commit to neither give nor receive transfers. We assume that the participation decisions are observed before proposals are made. Moreover, all players, also those that decide not to participate in contracting, observe the contracting process and learn about any ensuing agreement.[12] Let $\Gamma^V(G)$ denote the corresponding full game. When $n = 2$, non-participation by any player implies that the two players will be playing G unmodified by any transfer. With $n > 2$, it is possible to have contracts between a strict subset of the players.

The possibility to refrain from participation in contracting does not affect the set of subgame-perfect equilibrium outcomes.

Theorem 4 *A strategy profile $\hat{x} \in X$ and payoff profile $\pi(\hat{x}) \in \Pi(\hat{x})$ can be sustained in a subgame-perfect Nash equilibrium of $\Gamma^V(G)$ if $\pi_i(\hat{x}) \geq u_i(G)$ for all i.*

We know from Theorem 1 that all these outcomes are attainable in subgame-perfect equilibrium under mandatory participation, so here we merely need to check that there exist credible off-path threats such that a non-participation deviation is unprofitable. This is a straightforward extension of the proof of Theorem 1: In the two-player case, a non-participation deviation by Player i implies that G is played unmodified. If the expectation is that Player i's worst equilibrium of G will be played in response to this deviation, the deviation is thus unprofitable. With $n > 2$ players, if Player i chooses to exit from negotiations, the remaining $N-1$ players are left to contract among themselves. Just as $N-1$ opponents can always credibly keep Player i's payoff down to $\pi_i = u_i(G)$ following a proposal deviation by Player i in $\Gamma(G)$ they can do so now (indeed, they have extra flexibility, as the proposals are not yet made when the non-participation deviation is observed).

Let us next impose consistency again. In a two-player game, the possibility to refrain from participation does not affect the solution set; we have a perfect analogue (and complement) to Theorem 2.

Theorem 5 *Suppose $n = 2$. Then, a strategy profile $\hat{x} \in X$ and payoff profile $\pi(\hat{x}) \in \Pi(\hat{x})$ are supportable in a consistent equilibrium of $\Gamma^V(G)$ if and only if $\hat{x}$ is efficient and $\pi_i(\hat{x}) \geq u_i^{ud}(G)$ for both players.*

[12] An extension of the analysis would be to consider the case in which only the participation decisions are observable. Even if contracts are then secret to non-participants, they might still matter as the participants would now seek to maximize their joint payoff.

Since the proof is almost identical to the proof of Theorem 2, we omit it.[13]

However, with more than two players, the set of consistent equilibria can be heavily affected by potential non-participation. The reason is that negotiation participants may not be able to credibly threaten to punish a non-participant. Once they know that a player is unable to participate in contracting, the participants may desire to punish the non-participant, but by consistency they will not play a strategy profile that is payoff-dominated by any other subgame-perfect equilibrium in the ensuing contracting subgame. And if all such consistent strategy profiles are sufficiently beneficial for a set of non-participants, full participation cannot be supported. Our next result formalizes this intuition by characterizing a class of games G for which efficiency is unattainable due to free-riding.

In the characterization, an important number will be the lowest payoff that a single non-participant i can obtain in a consistent equilibrium of the subgame that starts at Stage 1 (the proposal stage). Denote by ΔX^{BR_i} a subset of strategy profiles in G where Player i plays a best response. Choose a strategy profile $\widetilde{\mu}^{BR_i} \in \Delta X^{BR_i}$ that maximizes the *joint* payoff to players $1, 2, ...i-1, i+1, ...n$. That is,

$$\widetilde{\mu}^{BR_i} = \arg \max_{\mu \in \Delta X^{BR_i}} \sum_{j \neq i} U_j(\mu).$$

If there are multiple such strategies, choose the one that is the worst for Player i. Denote Player i's payoff at this strategy profile by

$$f_i(G) = U_i(\widetilde{\mu}^{BR_i}).$$

Theorem 6 *Suppose $n > 2$. Moreover, for all players i suppose (i) $x_i^* \notin BR_i(x_{-i}^*)$ and (ii) $v_i(G) = u_i^{ud}(G)$. Then, $\Gamma^V(G)$ has an efficient consistent equilibrium if and only if*

$$\sum_{i \in N} U_i(x^*) \geq \sum_{i \in N} f_i(G).$$

The proof is in the Appendix; here is some interpretation. In these games there is a tension between efficiency and equilibrium (condition (i)). Also, the worst equilibrium payoff is as low as it gets, with each player obtaining her maximin payoff (condition (ii)). The class includes many games of economic interest, especially public goods provision games. Intuitively, the theorem rests on the following set of observations. First, the non-participating player is bound to best-repond to the anticipated actions of contracting coalition (these actions will typically be easy to predict from their contract).

[13] The most powerful consistent punishment threat is to play the worst (for the deviator) undominated Nash equilibrium of G, and a non-participation deviation is no different in payoff terms from a no-signing deviation.

Second, in this class of games, any consistent equilibrium in which only players $j \neq i$ can contract implies an action profile that cannot be Pareto-improved. Consequently, the worst consistent punishment for Player i's non-participation is achieved exactly at $\widetilde{\mu}^{BR_i}$. Indeed, were there another consistent equilibrium with worse non-participation punishment, consistency would require a higher joint payoff to players $j \neq i$ (as they can negotiate), which contradicts the definition of $f_i(G)$. Now, as $x_i^* \notin BR_i(x_{-i}^*)$, supporting it requires participation of all players. Ensuring universal participation is thus possible if and only if x^* provides sufficient resources for each Player i to overcome the incentive to unilaterally deviate, as given by $f_i(G)$.

One instance of G that has been studied in numerous experiments has four players, each with an endowment of M money units, individually choosing how much of the endowment to contribute to a public good. Each contribution x_i is multiplied by 1.6 and the resulting sum is divided equally among all four players. That is, Player i's payoff is

$$1.6\frac{\sum_j x_j}{4} + M - x_i.$$

Suppose the payers are selfish and risk neutral, so that money corresponds to utility. If players could decide whether to participate or not, i.e., if they were playing $\Gamma^V(G)$ instead of G, then Theorem 6 says that an efficient outcome, the full contribution profile (M, M, M, M) with payoffs $(1.6M, 1.6M, 1.6M, 1.6M)$, cannot be supported as a consistent equilibrium. To see this, note that $\Gamma^V(G)$ is covered by Theorem 6 (in G each player's unique dominant strategy is to contribute nothing, and the unique equilibrium payoff profile of G is thus (M,M,M,M), which coincides with the maximin payoff for each player. Let us illustrate the logic of the theorem: Player i's best response is always to contribute nothing. Thus, ΔX^{BR_i} includes all strategy profiles in which Player i contributes nothing. The payoff of the remaining three players $j \neq i$ is given by

$$\sum_{j \neq i} \left[1.6\frac{\sum_{j \neq i} x_j + 0}{4} + M - x_j \right] = 0.2 \sum_{j \neq i} x_j + 3M,$$

and is maximized by full contribution of players $j \neq i$. The payoff of Player i associated with maximum credible punishment by collaborating opponents is thus

$$1.6\frac{3M}{4} + M - 0 = 2.2M.$$

The sum of unilateral free-rider payoff for all 4 players thus equals $8.8M$, which exceeds the total payoff at the efficient outcome, $6.4M$. Hence, full cooperation is not sustainable in a consistent equilibrium of the contracting game.

In view of this inefficiency result, the natural next question is whether there is some different allocation of ownership rights that would admit efficient outcomes.

6 The Allocation of Ownership

Up until now, the assumption has been that each player i is entitled to all actions in X_i. Hence, there is no sense in which a player can violate another's ownership rights. Once an action that Player i can take is owned by another player $j \neq i$, that ownership right can be violated. We assume that there is an exogenous enforcement agency that penalizes such violation.[14] As will become clear, our use of the ownership concept is closely in line with that of *Encyclopaedia Britannica:* "Ownership of property probably means at a minimum that one's government or society will help to exclude others from the use or enjoyment of one's possession without one's consent, which may be withheld except at a price." Specifically, let $p_i(x)$ denote the exogenously imposed penalty paid by player i given the action profile x.

We are now ready to define precisely our notion of ownership.

Definition 3 (Self-ownership.) *Player i owns a personal action $x_i \in X_i$ if, in the absence of any contract, Player i has the right to take the action at no cost: $p_i(x_i, x_{-i}) = 0$ for all x_{-i}.*

In other words, a player owns a personal action if the player can take the action with impunity. Let *universal self-ownership* describe the case in which all actions are self-owned.

Definition 4 (External ownership.) *Player i owns another player's action $x_j \in X_j$ if, in the absence of any contract, (a) Player j must pay penalties $p_j(x_j, x_{-j}) > 0$, and (b) Player i has the right to relieve Player j of these penalties.*

That is, Player i owns Player j's action x_j if Player j incurs a penalty by taking the action *unless* Player i relieves Player j of this penalty, typically at a price. It is useful to make a distinction between dictatorial and non-dictatorial ownership.

Definition 5 (Non-dictatorial ownership.) *Ownership is non-dictatorial if each player $i \in N$ has self-ownership of at least one action x_i.*

It is also useful to distinguish between individual and collective action ownership.

Definition 6 (Individual action ownership.) *Ownership to an action x_i is individual if x_i has a single owner and collective otherwise.*

Under collective action ownership, it takes a collective decision by the action owners to relieve an actor of the stipulated penalties (for example, decision-making could be consensual or by majority vote).

[14] Since we shall focus on the case in which penalties are prohibitively large, it will typically not matter how the penalties are distributed. Assume for example that the penalties are paid to the owner of the violated right.

In order to go from action ownership to asset ownership, we must recognize that some assets can be used by multiple persons, often simultaneously yet in different ways. For example, think of a fisherman and a manufacturer who both want to use a river or a lake. In this case, we shall take individual asset ownership to mean that one person can regulate the other's asset use, but not vice versa.

To be more precise, let us focus attention on asset usage games G in which each player owns a default action (think of it as "no use"), and all other actions involve the asset in some way. We may then define individual asset ownership as follows.

Definition 7 (Individual asset ownership.) *In an asset usage game, asset ownership is said to be individual if one player owns all actions except the opponents' default actions.*

Conversely, collective asset ownership comprises either of two different departures from individual ownership: (i) collective ownership of specific actions; (ii) cases in which several players own multiple action rights.[15]

We may now analyze the contracting game Γ^V exactly as before, except for two changes. First, the game that is played when no contract is agreed is G modified by the penalties, call it $G(p)$. Second, the contracts now might include clauses whereby action owners relieve actors of such penalties.

Note that the magnitude of the penalty is a natural measure of the strength of the protection of rights. From now on, assume for simplicity that penalties are either zero, in the case of no violation, or "prohibitively large," in the case of violation. Formally, $p_i(x) \in \{0, \bar{p}\}$, and $\bar{p}$ satisfies the following condition.

Definition 8 (Strong property rights protection.) *Property rights protection is strong if the penalty $\bar{p}$ associated with all violations exceeds $\max_x U_i(x) - \min_x U_i(x)$ for all i.*

A simple asset usage game

To fix ideas, consider the classical case of a manufacturer and a fisherman who both desire to take actions that involve a river. The manufacturer either pollutes (plays P) or abates (plays A), whereas the fisherman either catches fish (plays F) or not (plays N). Specifically, let the manufacturer be the row player and let the fisherman be the column player of the games in Figure 4.

Universal self-ownership: When both players own their personal actions, the no-contract outcome is that the manufacturer pollutes, and the fisherman catches fish; i.e., the inefficient outcome (P,F) is the unique Nash equilibrium of the game G. Under general assumptions, our previous analysis shows that this problem can be solved contractually: In all consistent equilibria of the contracting game Γ (or Γ^V), the profile (A,F) is played,

[15] Observe that this notion of asset ownership is related to but somewhat richer than that of Hart and Moore (1990), since here a single individual can take more than one action that involves the asset.

	F	N
P	$1,1$	$1,0$
A	$0,3$	$0,0$

Self-ownership

	F	N
P	$1+\bar{p},1-\bar{p}$	$1,0$
A	$\bar{p},3-\bar{p}$	$0,0$

Manufacturer ownership

	F	N
P	$1-\bar{p},1+\bar{p}$	$1-\bar{p},\bar{p}$
A	$0,3$	$0,0$

Fisherman ownership

Figure 4: The Irrelevance of River Ownership

and the total surplus of 3 is split in such a way as to give each player a payoff of at least 1.

Manufacturer ownership: Suppose now instead that the manufacturer has the property right to the river, i.e., owns the right to take action P as well as the right to allow or deny the fisherman's action F. Suppose for simplicity the stipulated penalty for fishing is $\bar{p}$ regardless of the manufacturer's action. Without a contract, the fisherman now gets utility $1-\bar{p}$ or $3-\bar{p}$ from fishing. If property rights protection is strong, so $\bar{p}>3$, he will thus not fish under any circumstance, unless there is a contract. Therefore, in a consistent equilibrium, contract negotiations will involve an explicit clause whereby the manufacturer allows fishing against a transfer that depends on both players' actions (so as not to be tempted to pollute). For example, the contract might stipulate transfers from the fisherman of less than 1/2 in case (P,F) and, say, 3/2 in case (A,F). The contract thus simultaneously sells the right to catch fish and regulates pollution.

Fisherman ownership: If, on the other hand, the fisherman owns the river – i.e., the own action set as well as P – then it is the manufacturer who is potentially subject to penalties. Without a contract, the manufacturer now gets utility $1-\bar{p}$ from polluting, and if $\bar{p}>1$ there will be no pollution. In this case, there is no need for contracting in order to attain the efficient outcome (A,F).

An efficiency result

In the case of two players, we knew already that efficient outcomes are attainable. Our simple example above merely served to illustrate how property rights are defined in our model and how they affect the allocation of surplus.[16]

What about the case of many players? Is there always some allocation of action ownership that ensures efficiency? The answer is affirmative.

Theorem 7 *Suppose property rights protection is strong. For any G there is a non-dictatorial allocation of action ownership such that $\Gamma^V(G(p))$ has an efficient consistent equilibrium.*

[16]If contracts are incomplete, surplus allocation might impact efficiency, as was first pointed out by Grossman and Hart (1986). For example, if the fisherman could make non-contractible investments that would enhance the payoff from fishing, then it would typically be more efficient that the fisherman holds the property right over the river (so as to prevent underinvestment due to hold-up by the manufacturer).

The full proof is in the Appendix, but we sketch it here. Consider an efficient action profile x^*. Suppose the ownership structure is such that each player i has individual ownership of only one of her actions, $x_i^0 = x_i^*$. Then, since penalties for violation are large, x^* is an equilibrium of $G(p)$. So, we merely need to show that x^* can be supported in an efficient consistent equilibrium of $\Gamma^V(G(p))$. As x^* is an efficient equilibrium, it is also an undominated equilibrium. Thus, $u_i^{ud} \leq U_i(x^*)$, and Player i's worst undominated equilibrium can be used as a deterrent (non-renegotiable) threat against a deviation by Player i at any stage of the contracting game $\Gamma^V(G(p))$.

An inefficiency result

The crucial idea of the proof of Theorem 7 is that efficiency can be obtained if each players' self-ownership is confined to the player's component of the efficient action profile. But note that this action ownership allocation is inconsistent with our notion of individual asset ownership. For example, with individual asset ownership, the owner of any others' (asset-relevant) actions will by definition also have ownership of all the own (asset-relevant) actions. As we now show, inefficiency may then be unavoidable.

Theorem 8 *There are asset usage games G such that there exists no individual asset ownership allocation admitting an efficient consistent equilibrium of $\Gamma^V(G(p))$.*

Proof. Consider the following 3-player public good game G. Each player has three levels of contribution $x^L = 0$, $x^M = 1$, and $x^H = 25/16$. The payoff of player i is

$$U_i(x_i, x_{-i}) = \sum_{k=1}^{3} \sqrt{x_k} - x_i - \frac{7}{8} I_{x_i > 0},$$

where I is an indicator function. The privately optimal choice is always x^L (due to the high fixed cost, 7/8, of a positive contribution). Thus G has a unique Nash equilibrium $\left(x^L, x^L, x^L\right)$ with payoffs $(0, 0, 0)$. On the other hand, the efficient allocation is $\left(x^H, x^H, x^H\right)$ with payoffs $\left(\frac{21}{16}, \frac{21}{16}, \frac{21}{16}\right)$. Finally, notice that x_i^M maximizes the joint payoff of two players (i, j) regardless of x_j and x_k.

Now, without loss of generality assume that Player 1 is the asset owner, and thus owns all actions of the other two players except for a pair of default actions x_2^0 and x_3^0.

To prove inefficiency, we need to show that the joint payoff under $\left(x^H, x^H, x^H\right)$ is smaller than the sum of the individual non-participation payoffs f_i, $i = 1, 2, 3$ for different constellations of x_2^0, x_3^0. The key observation is that, regardless of the default action profile, pairwise negotiations entail the following outcomes: (i) If the asset owner, Player 1, is not part of the negotiations, $x_1 = x_L$. (ii) If Player 1 and Player j negotiate, $x_1 = x_j = x^M$.

The non-participation payoff of Player 1 is thus

$$f_1\left(x_2^0, x_3^0\right) = \sqrt{x^L} + \sqrt{x_2^0} + \sqrt{x_3^0} - x^L = \sqrt{x_2^0} + \sqrt{x_3^0},$$

and that of Players $i = 2, 3$ is

$$f_i\left(x_2^0, x_3^0\right) = 2\sqrt{x^M} + \sqrt{x_i^0} - x_i^0 - \frac{7}{8} I_{x_i^0 > 0} = 2 + \sqrt{x_i^0} - x_i^0 - \frac{7}{8} I_{x_i^0 > 0}.$$

Hence, the sum of non-participation payoffs is

$$\sum_{k=1}^{3} f_k\left(x_2^0, x_3^0\right) = 4 + 2\sqrt{x_3^0} - x_2^0 - \frac{7}{8} I\{x_2^0 > 0\} + 2\sqrt{x_3^0} - x_3^0 - \frac{7}{8} I_{x_3^0 > 0}.$$

It is easy to check that this expression is minimized for $x_2^0 = x_3^0 = x^L$ and even at these contribution levels it exceeds the sum of the payoffs at the efficient allocation

$$\sum_{k=1}^{3} f_k\left(x^L, x^L\right) = 4 > 3 \cdot \frac{21}{16} = \frac{63}{16},$$

which completes the proof. ■

Together, Theorems 7 and 8 say that there is always some individual allocation of *action ownership* that sustains efficiency, but that this allocation of action rights may have to involve collective *asset ownership*. The reason is that an individual asset owner may use the ownership privilege to free ride – or even to sell out the right to take a free ride at others' expense. While the inefficient outcome would be avoided if all players were to negotiate, in consistent equilibria it is not incentive compatible for all players to enter negotiations.

7 Robustness to Unilateral Promises

So far, we have studied contracts that take the form of agreements. In this section, we examine to what extent our results are robust when we also allow unilateral promises. We first consider the robustness of our inefficiency results (the case of voluntary participation and consistent equilibria), focusing exclusively on a class of promises that tend to be efficiency enhancing and is supported by legal doctrine. We next consider the robustness of our efficiency results (the case of mandatory participation) to the case in which all possible promises are allowed, whether supported by legal doctrine or not. Throughout, we revert to our initial assumption that players own their action sets.

7.1 Will restricted promises restore efficiency?

A possible objection to our inefficiency results is that our contracting process does not admit realistic unilateral promises. In reality, when a buyer unilaterally promises to pay money in return for some action taken by a potential seller, then this procurement contract may create a legal obligation for the buyer to pay once the seller has taken the requested action. That is, the buyer's obligation can arise regardless of whether the seller first agrees to the buyer's promise or the seller only holds the buyer to the announced promise afterwards. In fact, the legal doctrine of "promissory estoppel" calls for the enforcement of unilateral promises as long as the promisee's action reasonably relied on the promise.

We shall now demonstrate that such unilateral promises do not generally suffice to restore efficiency in the case of voluntary participation.

Consider the three-player game in Figure 5, with Player 1 choosing rows, Player 2 choosing columns, and Player 3 choosing matrices.

	H	L
H	$2, 2, 2$	$0, 5, 0$
L	$5, 0, 0$	$3, 3, -6$
Player 3: H		

	H	L
H	$0, 0, 5$	$-6, 3, 3$
L	$3, -6, 3$	$-1, -1, -1$
Player 3:L		

Figure 5: A three-player social dilemma.

This game has a unique Nash equilibrium, (L, L, L). Under voluntary contracting better outcomes can be sustained. Specifically, there are equilibria of $\Gamma^V(G)$ in which two players enter negotiations, while the third refrains: For example, if Player 3 (the matrix player) refrains from taking part in contracting, Player 3 will play the dominant strategy L. Thus, players 1 and 2 face contract negotiations about the situation depicted in Figure 6:

	H	L
H	$0, 0$	$-6, 3$
L	$3, -6$	$-1, -1$

Figure 6: Situation facing the two remaining players if the third stays out.

It is thus consistent for Player 1 and Player 2 to enter and negotiate (H,H), which in turn rationalizes the decision of Player 3 to not participate in negotiations (as Player 3 gets the payoff 5 by refraining). However, for exactly this reason, the efficient allocation (H,H,H) will not be sustainable, as it could not yield all players their non-participation payoff.

However, can the efficient allocation (H,H,H) be sustained in equilibrium if, in addition to agreements between participants, promises could be made (at Stage 2.5) by the participants to the non-participants? Call such contracting games $\Gamma^{VP}(G)$.

To start with, notice that supporting (H,H,H) with all three players participating is again unattainable. Indeed, were all three to enter, at least one of them would get a payoff not exceeding 2. Without loss of generality, say it is Player 3. If Player 3 instead chooses not to participate, the worst consistent non-participation payoff that Players 1 and 2 could keep Player 3 down to 3; this payoff realized by them offering (the non-participating) Player 3 a payment slightly above 3 in either (H,L,H) or (L,H,H). So, Player 3 would indeed prefer not to participate. An analogous argument ensures that there could be no consistent equilibrium with all three players participating.

Now then, can the outcome (H,H,H) be supported with two players participating only? It turns out the answer is negative. In fact, no consistent equilibrium can entail participation by two players. To see this, consider what will happen if only one player (say, Player 1) participates. Then, Player 1's consistent equilibrium strategy is to make promises to players 2 and 3 in order to implement the strategy profile $\mu \in \Delta$ in the modified game $\tilde{G}$ for which the resulting payoff of Player 1 is maximized (recall that Players 2 and 3 cannot themselves promise anything). In case there are multiple such strategy profiles, consistency selects any Pareto-undominated payoff profile. It is easy to see that Player 1 would thus implement strategy profile (L, H, H), offering players 2 and 3 a payment slightly above 3 each to persuade them to choose H over L. The ensuing action profile (L,H,H) leaves Player 1 with a payoff of -1, and yields a payoff of 3 to both Players 2 and 3 (see Figure 7).[17]

	H	L
H	$2-6, 2+3, 2+3$	$0-3, 5, 0+3$
L	$5-6, 0+3, 0+3$	$3-5, 3, -6+5$
Player 3: H		

	H	L
H	$0-3, 0+3, 5$	$-6, 3, 3$
L	$3-5, -6+5, 3$	$-1, -1, -1$
Player 3:L		

Figure 7: Payoffs if only Player 1 enters.

Consequently, either of players 2 and 3 would need to get at least 3 to enter a two-player coalition. More generally, any consistent equilibrium with two participating players must yield at least 3 to each of them. For each pair of players, there is only one strategy profile in G with this property (e.g., for Players 1 and 2 it is (L,L,H)), and such a profile cannot be sustained in equilibrium as the third player would deviate. Thus, a consistent equilibrium with two participating players is unattainable in the game $\Gamma^{VP}(G)$.

With one participant only, we already know that the outcome is (L,H,H) rather than

[17]It is straightforward to specify appropriate specification of off-path actions for this outcome to form a consistent equilibrium of $\Gamma^{VP}(G)$.

the efficient allocation (H,H,H). Thus, the example demonstrates that the promissory estoppel doctrine is not sufficient to restore efficiency under voluntary participation.

We have chosen the example in such a way as to make total efficiency insensitive to the availability of promises. An interesting question for future research is to understand whether the promissory estoppel doctrine could ever harm efficiency, or whether it is generally benign. In the latter case, our analysis of the participation problem would be a potential explanation (or at least a justification) for why this legal doctrine exists in the first place.

7.2 Will unrestricted promises destroy efficiency?

In the context of mandatory participation, we know that efficiency is attainable. Thus, in this case the relevant question is whether promises can undermine efficiency. Jackson and Wilkie (2005, 561) conjecture that they might, at least if sufficiently general promises are allowed. In Jackson and Wilkie, a promise by Player i can be any function $T_i^{JW} : X \to \mathbb{R}_+^{n-1}$ specifying non-negative transfers from Player i to each of the $n-1$ opponents as a function of the pure strategy profile x.

Since it does not matter whether opponents rely on the promise, not all such promises are covered by the promissory estoppel doctrine. As an illustration, consider the game in Figure 1 again. If rancher 1 promises to pay rancher 2 a transfer in case rancher 2 takes action H instead of the privately more desirable action L, the promise is clearly covered by the doctrine. But in addition to such ordinary "promises to exchange," Jackson and Wilkie allow purely "donative promises" according to which rancher 1 promises to pay a transfer to rancher 2 in case rancher 1 takes action H. That is, rancher 1 may use rancher 2 as a sink, with the implicit purpose of making credible the statement that "I, rancher 1, do not intend to play H".[18] It is these strategic donative promises that occupy center stage in Jackson and Wilkie (2005) and that play a crucial role below.

In the context of mandatory participation, Jackson and Wilkie (2005) ask and answer the following question: What if players can make unrestricted promises, but not multilateral contracts? That is promise by Player i can be any function $T_i^{JW} : X \to \mathbb{R}_+^{n-1}$ specifying non-negative transfers from Player i to each of the $n-1$ opponents as a function of the pure strategy profile x. Thus, Jackson and Wilkie (2005) study the two-stage game, denote it $\Gamma^{JW}(G)$, in which all players first simultaneously make promises T_i^{JW} and

[18] It is not clear that contract law supports such a transfer. In legal parlance, rancher 1 here issues a "gratuitous promise" which is ordinarily seen as a promise that "lacks consideration." (Consideration is defined as the price that the promisee pays in return for the promisor's promised action. For a discussion of consideration in the context of gratuitous promises, see for example Gordley, 1995.) Thus, if rancher 1 eventually were to play H, rancher 2 – who already benefits from rancher 2's action – might not be able to have the court enforce rancher 1's promise. With the recent growth of firms that offer "legally binding" commitment contracts, such as the company StickK, we may soon learn to what extent courts will uphold purely donative promises issued for self-control purposes.

then play G modified by these promises. Among other things, they showed that $\Gamma^{JW}(G)$ does not always admit efficient equilibrium outcomes.[19] One of their conjectures is that such unrestricted promises could also destroy opportunities for efficient agreements.

However, we shall show that efficiency is often attainable despite the availability of unilateral donative promises. Indeed, there are even games G for which the possibility of making promises helps to expand the set of efficient outcomes that could be sustained in an equilibrium of $\Gamma(G)$. The reason is that a promise can be used as a pure threat, forcing the opponent down to her maximin outcome of G, which in some games is below her worst Nash equilibrium payoff.[20]

Additional definitions

The key difference between promises and agreements is that agreements require joint consent by all affected parties, while promises require unilateral consent only by promisors, i.e., the players who make positive transfers. We shall assume that an agreement can, if the parties so wish, invalidate any of their unilateral promises. If so, promises are executed if and only if parties fail to reach an agreement.

Specifically, a contract proposal by Player i is now written $(\tau_i, \mathcal{T}_i) = (t, T)$, where τ_i denotes the agreement clause of the proposed contract, and $\mathcal{T}_i$ the promise clause, to be described in the next paragraph. At the signature stage, the players affected by the agreement clause, τ_i, make their signing decisions. If a proposed agreement τ_i is signed by all affected players N^t, t is binding and T becomes irrelevant. If τ_i is not signed by all affected players, the promises T are in play.

We focus attention on promise clauses that are activated if some affected player does not sign the proposed agreement. Hence, let M^t denote the set of all possible proper subsets of N^t and let $T_{jk} : X \times M^t \to \mathbb{R}_+$ denote a promised transfer from Player j to Player $k \neq j$ conditional on the set of signatures.[21] Since Player j has $n-1$ opponents, and their consent is not required to enact a promise, Player j's unilateral obligations according to some promise clause can be written $T_j : X \times M^t \to \mathbb{R}_+^{n-1}$. A promise clause thus specifies unilateral obligations $T = \{T_j\}_{j \in N^t}$ for each player affected by the agreement clause. Finally, denote the set of players who signed the contract by S^t, where $S^t \subset N^t$.

Let $\Gamma^R(G)$ be a contracting game with the same structure as before, but in which contracts are allowed to be of the form (t, T). Suppose finally that each player may sign

[19] This result is not quite as strong as it may seem, since Jackson and Wilkie do not prove that an equilibrium exists in this case.

[20] Of course, a key condition for this mechanism to work is that the promise is "turned off" by the signing of an agreement.

[21] If we would not allow the promises to differ depending on who signs the contract, the set of efficient payoff profiles that can be sustained would shrink, but not be empty. Intuitively, promise clauses serve to punish players who fail to sign an agreement. Promise clauses that do not target specifically the deviating player constitute less powerful threats.

at most one contract. (Alternatively, we could reach identical conclusions by allowing players to impose an entire-agreements clause in their contract proposals.)

Result

We want to demonstrate that that promise clauses can serve as threats, i.e., they can be used to minimize reluctant signatories' payoff. To see what that minimum is, suppose hypothetically that all Player i's opponents could commit to a pure strategy profile in case Player i does not sign an agreement. Then they could potentially keep Player i down to her pure strategy maximin payoff in G,

$$v_i^p(G) = \max_{x_i} \min_{x_{-i}} U_i(x_i, x_{-i}),$$

but no lower. (Since G is a finite game, v_i^p is well-defined.) Thus $v_i^p(G)$ is the harshest threat that Player i could ever face. As it turns out, such threats can be implemented through promise contracts, and that possibility in turn defines the range of sustainable efficient outcomes.

Theorem 9 *Suppose $n > 2$. Then any strategy profile $\hat{x} \in X$ and payoff profile $\pi(\hat{x}) \in \Pi(\hat{x})$ that yields each player i a payoff weakly above $v_i^p(G)$ can be sustained in a subgame-perfect equilibrium of $\Gamma^R(G)$.*

The main difference as compared to $\Gamma(G)$ is that the outside option of Player j (in case she unilaterally deviates by not signing the contract of Player i) is now affected by unilateral transfer clauses rather than merely by the payoffs in G. The crucial step is to show that it is still possible for Player i to make Player j sign the agreement despite Player j being able to affect the disagreement payoff by making promises.

The proof of Theorem 9 rests on the following Lemma.

Lemma 1 *Suppose $n > 2$. Then, for any i there exists a profile of promise functions $T^{-i}(x) = \left(T_1^{-i}(x), ..., T_{i-1}^{-i}(x), T_{i+1}^{-i}(x), ..., T_n^{-i}(x)\right)$ which bounds the payoff of Player i in $\Gamma^{JW}(G)$ to $v_i^p(G)$.*

The lemma's proof is similar to the proof of Proposition 4 of Jackson and Wilkie (2005) and is relegated to the Appendix. In short, each of Player i's opponents commits to choose the strategy that supports the maximin of Player i in game G by making a large transfer to all remaining players in case of any deviation from the intended maximin strategy profile. Hence, it becomes too costly for Player i to pay any opponent to deviate.

Having established the existence of such promises, it remains to show how they can be used as a credible threat in order to induce Player i to sign an agreement. Specifically, let each player i offer an identical contract $(\tau_i, \mathcal{T}_i) = (\hat{t}, \hat{T})$ covering all the players, where the agreement clause specifies the following net transfers from each Player j,

$$\hat{t}_j(x) = \begin{cases} U_j(x) - \pi_j(x) & \text{if } x = \hat{x}; \\ (n-1)h & \text{if } x_j \neq \hat{x} \text{ and } x_{-j} = \hat{x}_{-j}; \\ -h & \text{if } x_j = \hat{x}_j \text{ and } |\{k : x_k \neq \hat{x}_k\}| = 1; \\ 0 & \text{otherwise,} \end{cases}$$

and the promise clause is

$$\hat{T}_j(x, S^t) = \begin{cases} T_j^{-m}(x) & \text{if } S^t = N^t / \{m\}; \\ 0 & \text{otherwise.} \end{cases}$$

In case of agreement, this contract clearly entails the payoffs $(\pi_1(\hat{x}), \pi_2(\hat{x}), ..., \pi_N(\hat{x}))$.

The remainder of the proof proceeds along the lines of the proof of Theorem 1 to show that everyone signing the contract proposed by Player 1 is indeed an equilibrium of $\Gamma^R(G)$. For Stage 3, along the equilibrium path, the argument is identical. At Stage 2, no player m would want to deviate from signing Player 1's contract, as everyone else's promises T^{-m} guarantee that $v_i^p(G)$ is the highest payoff Player m can get, and $v_i^p(G) \leq \pi_m(\hat{x})$. Similarly, there is no way that Player m can improve her payoff at the contract proposal stage, as $v_i^p(G)$ is the highest payoff Player m can get by unilaterally altering the own promise clause.

Just as Theorem 1, Theorem 9 ensures the existence of efficient equilibria in the game $\Gamma^R(G)$. However, while Theorem 9 covers many games, it does not cover two-player games. Indeed, the method of proof does not generalize to this case; when $n = 2$, it is no longer true that all maximin outcomes of G can be supported as equilibria of $\Gamma^R(G)$ (exactly like not all equilibria of G remain equilibria of $\Gamma^{JW}(G)$, as shown by Jackson and Wilkie.) While we can show that efficient outcomes are also sustainable in 2×2 games with at least one PSNE, such as the game in Figure 1, we have neither a proof nor a counterexample for the entire class of two-player games. We conjecture that the result generalizes, but must leave the question open.

Conditioning promises on strategies

We next identify additional features of the contracting technology that are sufficient for sustaining efficient outcomes in all finite normal-form games.

Suppose now that contracts may condition transfers on mixed strategies in G, that is on Δ, rather than just on the realized profile of actions (pure strategies) X. Of course, this assumption is strong, arguably unduly so.[22] Furthermore, allow transfers to be non-

[22]While it is not unreasonable to assume that players have access to public randomization devices, and may agree ex ante to condition transfers on the realization of such a device, it is an entirely different

deterministic (when transfers are non-deterministic, they cannot straightforwardly be neutralized by an offsetting transfer in the other direction). Formally, a non-deterministic contract specifies a probability distribution over transfers, so let Ψ denote the set of all probability distributions on the space of functions from Δ into $\mathbb{R}^{n-1}$, and Ψ^+ denote the set of all probability distributions on the space of functions from Δ into $\mathbb{R}_+^{n-1}$. Then we can express Player i's contract proposal as a pair of functions $(\tau_i^E, \mathcal{T}_i^E) = (t^E, T^E)$, where the agreement clause $t^E \in \Psi$, and the promise clause specifies unilateral obligations $T_j^E : M^t \to \Psi^+$ for all players $j \in N^t$.[23] Let $\Gamma^E(G)$ denote the full game with this extensive set of contracting opportunities.

Let

$$v_i(G) = \max_{\mu_i} \min_{\mu_j} U_i(\mu_i, \mu_j)$$

be Player i's mixed-strategy maximin payoff in G. Since G is a finite game, v_i is well-defined.

Theorem 10 *For any two-player game G, $\Gamma^E(G)$ has an efficient subgame-perfect equilibrium.*

Let us here give the core of the proof, while relegating details to the Appendix when noted.

As in the previous subsection, the outside option of Player 2 (in case she does not sign the contract of Player 1) is now affected by unilateral transfer clauses rather than merely by the payoffs in G. The crucial step is to show that, also in the case of two players, Player i can choose her promise clause to put an upper bound on what Player j can achieve. Moreover, this bound can be made so low that there is a division of the efficient surplus that both players prefer.

It is convenient to start by analyzing the simpler game in which agreements cannot be signed, that is, $\Gamma^{JW}(G)$ extended to allow conditioning on mixed strategies – call this game $\Gamma^{JW,E}(G)$.

Lemma 2 *For each $d > 0$ and $i = 1, 2$ there exists a bounded promise of Player $j \neq i$, $T_{j,d} : \Delta \to \mathbb{R}_+$, never exceeding $h\,(4h/d - 1)$, such that Player i's payoff in $\Gamma^{JW,E}(G)$ is at most $v_i(G) + d$.*

matter to verify ex post which mixed strategy a player has been following. To allow a court to enforce the transfers specified in such a contract, the following seems necessary. First, a randomizing player must, when playing G, publicly announce the randomization device before taking any action. Second, and more controversially, there must be a credible link between the device's realization and the action. As there is no guarantee that the prescribed action is ex post incentive compatible, the player must effectively have delegated the (contingent) action execution. In a partial way, this brings into the model the undesirable feature that contracts directly restrict actions, which is the feature of some other approaches to contracting that we least like. (Note however that the choice of randomization device is still not directly controlled by the contract.)

[23]Jackson and Wilkie mention randomization over the set of transfers, but for technical reasons do not explicitly allow it (Jackson and Wilkie, 2005, page 549).

Proof. See Appendix. ■

The logic behind Lemma 2 runs roughly as follows. First, Player 1 cannot make a unilateral promise such that Player 2's resulting payoff is below $v_2(G)$. The reason is that Player 2 can always choose not to promise any unilateral transfers, in which case the transfers of Player 1 can only improve Player 2's minimum payoff. Second, Player 2 can be held sufficiently close to $v_2(G)$ by a promise $T_{1,d}(\mu)$ that mixes between two sets of unilateral promise transfers. The first promise, call it $\underline{T}_{1,d}(\mu)$, marginally ensures the dominance of the strategy μ_1^m which causes $v_2(G)$ under the provision that Player 2 does not offer a contract. The second promise, call it $\overline{T}_{1,d}(\mu)$, makes μ_1^m much more strongly dominant; it induces Player 1 to play μ_1^m unless Player 2 offers some large transfer l in return for a different strategy. Suppose Player 1 plays the first clause with a sufficiently small probability $p(l)$. Then, any clause that is part of a (mixed-strategy) best response of Player 2 yields a finite payoff to Player 2. Moreover, the best response can either yield Player 2's minimax in G with high probability $1 - p(l)$ (if the best response transfer is too small to counteract the dominance of μ_1^m produced by $\overline{T}_{1,d}(\mu)$), or yield a loss of an order of magnitude of l with relatively small probability $p(l)$. (In this latter case Player 2's transfer should be sufficiently large to counteract $\overline{T}_{1,d}(\mu)$, which implies that Player 2 loses at least an extra l when playing this transfer against $\underline{T}_{1,d}(\mu)$.) It turns out that l does not need to be very large to ensure that Player 2's payoff does not exceed $v_2(G) + d$; in fact, $l = 2h\left(\frac{2h}{d} - 1\right)$ would already be sufficiently high. Finally, notice that, by making μ_1^m just dominant, $\underline{T}_{1,d}(\mu)$ clearly does not exceed h, and thereby $\overline{T}_{1,d}(\mu)$ does not exceed $h + l$.

Given our previous analysis of $\Gamma(G)$, it is now straightforward to formulate the equilibrium strategies of $\Gamma^E(G)$. For example, consider an efficient strategy profile x^*, pick

$$d = \frac{U_1(x^*) + U_2(x^*) - (v_1 + v_2)}{2}.$$

If $d > 0$, consider equal division of the efficient surplus:

$$\begin{aligned} \pi_1(x^*) &= v_1 + d, \\ \pi_2(x^*) &= v_2 + d. \end{aligned}$$

Let us show that x^* and the allocation $(\pi_1(x^*), \pi_2(x^*))$ can be supported by as an efficient subgame-perfect equilibrium of $\Gamma^E(G)$.

At Stage 1, players $i = 1, 2$ offer the following "identical" contracts $(\tau_i^E, \mathcal{T}_i^E) =$

(t^{E*}, T^{E*}), where the agreement clause is

$$t_j^{E*}(x) = \begin{cases} U_j(x) - \pi_j(x) & \text{if} \quad x = x^*; \\ h & \text{if} \quad x_j \neq x_j^* \text{ and } x_{-j} = x_{-j}^*; \\ -h & \text{if} \quad x_j = x_j^* \text{ and } x_{-j} \neq x_{-j}^*; \\ 0 & \quad\quad \text{otherwise,} \end{cases}$$

and the unilateral promise clause is

$$T_j^{E*}(x, S_{t^*}) = \begin{cases} T_{j,d} & \text{if} \quad j \text{ signs the contract, while } i \text{ does not;} \\ 0 & \quad\quad \text{otherwise.} \end{cases}$$

That is, as in Theorem 9, each player uses the promise clause to impose an upper bound on the payoff of a non-signing opponent; the only difference is that the promise clauses are here non-deterministic.

Stages 2 and 3 are analogous to the proof of the best-reply properties in Theorem 9. Here, by deviating at the proposal stage the players can manipulate the outcome of the branches where one party fails to sign the contract by altering unilateral clauses. However, since any deviation by Player i to a different promise cannot yield her more than $v_i + d = \pi_i(x^*)$, it follows that neither player can profitably deviate by manipulating the promise.

We are left with the case of $d = 0$, that is, when the the sum of the maximin payoffs is efficient. In this case all Nash equilibria of G are efficient, and with payoffs equal to the players' maximins. Such an equilibrium can be straightforwardly supported as a SPNE of $\Gamma^E(G)$; let each player offer zero transfers both for the agreement clause and for the promise clause. A unilateral deviation to positive transfers can now only improve the situation for the opponent (who is always assured the maximin) at the expense of the deviating player (for details, see the Appendix).

The only difference between the efficient equilibria of $\Gamma^E(G)$ and those of $\Gamma(G)$ is that we now depend on public randomization. The unilateral clauses $T_{j,d}(\mu)$ involve mixing, and the mixing probabilities must be observable; otherwise, Player 1 would potentially choose a different strategy (depending on whether the maximin strategy is a best reply or not).

Having demonstrated that the negotiation game $\Gamma^E(G)$ always has an efficient outcome, let us finally show that the range of efficient outcomes is no smaller than for $\Gamma(G)$ and could be larger. In particular, any efficient outcome that yields each player more than the maximin payoff may be sustained (whereas in $\Gamma(G)$, with two players, each player must obtain at least their smallest Nash equilibrium payoff).

Theorem 11 *Any efficient outcome that yields each player i a payoff strictly above $v_i(G)$ can be sustained in a subgame-perfect equilibrium of $\Gamma^E(G)$.*

Lemma 2 already demonstrates that Player i can bring Player j's payoff arbitrarily close to v_j (though it could require larger and larger transfers). Now, to sustain the efficient equilibrium x^* with payoffs $(\pi_1(x^*), \pi_2(x^*))$, where $\pi_i(x^*) > v_i(G)$, choose

$$d = \min_{i=1,2}(\pi_i(x^*) - v_i(G)) > 0, \tag{1}$$

and let players $i = 1, 2$ offer "identical" contracts $(\tau_i^E, \mathcal{T}_i^E) = (t^{E*}, T^{E*})$ with

$$t_j^{E*}(x) = \begin{cases} U_j(x) - \pi_j(x) & \text{if } x = x^*; \\ h & \text{if } x_j \neq x_j^* \text{ and } x_{-j} = x_{-j}^*; \\ -h & \text{if } x_j = x_j^* \text{ and } x_{-j} \neq x_{-j}^*; \\ 0 & \text{otherwise,} \end{cases}$$

and

$$T_j^{E*}(x, S_{t^*}) = \begin{cases} T_{j,d} & \text{if } j \text{ signs the contract, while } i \text{ does not;} \\ 0 & \text{otherwise,} \end{cases}$$

where $T_{j,d}(\mu)$ is as set by Lemma 2 for d defined in (1). Since any Stage 1 deviation of Player i cannot yield more than $v_i + d \leq \pi_i(x^*)$, it follows that neither player can profitably deviate by manipulating her/his promise. The rest of the proof is analogous to the proof of Theorem 9. For the case of more than two players, the proof relies on the same construction as the proof of Theorem 9, except allowing conditioning on mixes.

Since we can never have any equilibrium of $\Gamma^E(G)$ in which a player gets less than the maximin payoff, Theorem 11 provides a complete characterization of attainable efficient payoff profiles, with the exception of the measure zero set in which some player gets exactly the maximin payoff.[24]

8 Related literature

Economists have long understood that well-designed transfer schemes can alleviate conflicts of interest and sustain cooperation. However, much of the early literature presupposes that at least the general features of these transfer schemes are imposed by a third party – a "social planner". For example, this is the approach taken by Varian (1994b), building on previous work by Guttman (1978).[25] Among other things, Varian shows that the Prisoners' Dilemma is resolved if players can announce binding promises of payments in return for cooperation by the opponent. More generally, we know from mechanism design theory that, under symmetric information, a wide range of efficient outcomes can

[24] As a curiosity, it can be shown that such extreme payoff profiles are attainable in a subgame-perfect $\varepsilon-$equilibrium.

[25] See also Guttman (1987), Danziger and Schnytzer (1991), Guttman and Schnytzer (1992), and Varian (1994a).

often be attained if players are confined to resolve their conflicts through appropriately designed mechanisms – even if the mechanism designer does not know people's preferences – see Moore (1992) for a survey.

When the designer knows the players' preferences, and is the first mover, efficiency is particularly easy to attain. For example, in Myerson (1991, Chapter 6), an outside mediator proposes a contract, and the players' decisions are to accept or reject the contract proposal. If all players accept, they are forced to play the associated action profile. Myerson shows that a large set of outcomes, some of which are efficient, can typically be sustained as equilibria of such contract-signing games.

Our approach is somewhat similar to Myerson's in adopting the view that contract-signing is usefully seen as a coordination game in which each player can become pivotal. However, our analysis differs from Myerson's in three respects. First, we do not assume that a signed contract directly forces players' actions. Instead, we assume that if the players accept a contract, they have to make the transfers that the contract specifies for the specific action profile that is subsequently played – which may or may not be the profile that the contract "intends".[26] Second, we assume that contracts are proposed non-cooperatively by the players themselves. Indeed, we insist that all players have the opportunity to propose contracts rather than presuming that some player or players – the principals – are granted exclusive proposal rights.[27] These two modifications do not harm efficiency. Third, Myerson only considers the case in which all players have to participate. We also consider the case in which players may refrain from taking part in contract negotiations. This modification does threaten efficiency. From a mechanism design perspective, our analysis thus highlights the importance of the standard assumption that the planner moves first. If players may individually decide whether to participate in contracting before the planner designs the contract, and if the planner cannot directly penalize non-participants, we conjecture that some version of our inefficiency result (Theorem 6) carries over to the mechanism design framework.[28]

With respect to the modelling approach – especially in disallowing an outside planner or mechanism designer, but also in terms of concepts and notation – our analysis is most closely related to Jackson and Wilkie (2005) and Yamada (2003). Jackson and Wilkie and Yamada consider non-cooperative promise games, where players themselves are quite free to decide on the shape of transfers. That is, players are not even confined to making

[26]This is the standard assumption of the mechanism design literature, and obviously Myerson only makes the stronger assumption for simplicity.

[27]In a related contribution, Watson (2011) takes the mechanism-design approach, but relaxes the amount of centralization that is required of the contracting process. His major insight is that efficient outcomes are often attainable even if parties are arranged in a network structure and are confined to contract only with parties that they are linked to, provided there are enough rounds of (sequential) negotiations and the network indirectly connects everyone.

[28]If this conjecture is correct, the upshot is that non-cooperative contract negotiations of the form we consider suffice to attain the (constrained) efficiency bound.

a specific sort of promise, as in Varian (1994b). However, Jackson and Wilkie do not admit contracts in our sense of the word. Instead, their model assumes that players issue unilateral promises to pay non-negative transfers to opponents depending on realized action profiles. Equilibria of this promise game have rather different properties from those of our contract-proposal game. For example, efficient outcomes are not always attainable in the case of $n = 2$.

While our model takes a more optimistic view of contracting opportunities than Jackson and Wilkie (2005), it arguably takes a less optimistic view than the related literature about "contracting on contracts"; see, e.g., Katz (2006), Kalai et al. (2010), Yamashita (2010), and Peters and Szentes (2012). This literature also dispenses with the social planner, but instead allows that each player may use a unilateral contract (or promise) to commit to their own action as a function of the contracts of other players. By construction, there is no explicit mutual assent by all players, but cooperation may arise in equilibrium nonetheless, as the commitments intertwine. In Yamashita (2010), intertwining is accomplished by letting promises depend on messages that are sent after all promises are observed (messages effectively report on these promises). The contracting-on-contracts approach is extremely powerful in its ability to sustain cooperative outcomes; indeed it can attain efficiency despite disallowing explicit transfer payments (i.e., utility is non-transferable).[29] However, it also makes great demands on contract enforcers. In effect, the enforcers are asked to verify the contracting process, including any ultimately "unsuccessful" promises (or at least messages about these). By contrast, our set-up merely requires that enforcers observe a single final agreement linking transfers to behavior in G. However, in the extension of our model where players are free to make promises, some form of intertwining seems necessary in order to sustain efficiency. Specifically, there is an affinity between Theorem 1 of Yamashita (2010) and of our Theorem 9.[30]

Our work is also related to the principal-agent literature, most closely with the work on contracting with externalities, and especially Holmström (1982) and Segal (1999,2003). In Segal's model, a principal trades with several agents, the trade with one agent may impose externalities on other agents, and there is an undesirable no-trade outcome in case of no contract. In our vocabulary, this setting translates to a game G with a unique and inefficient Nash equilibrium, and with a particular asymmetric payoff structure; one player, the principal, is not taking any action, but is potentially affected by actions of others (the agents). But there is a major difference with respect to commitment possibilities. In Segal's model, only one player can propose a contract, and this proposal can be made before any participation decision by the other players. Segal (1999, Proposition 10) shows that, if there is no restriction on the space of available contracts a principal can

[29]Yamada (2003) may be seen as a link between the two approaches, as his contracts specify conditional transfers, rather than actions, while also conditioning on opponents' contracts.

[30]In the proof of Theorem 9, players utilize unilateral promise clauses that are de-activated in case of signing as powerful threats to induce everyone to sign the desirable contract.

implement the efficient outcome and extract all the surplus over and above the no-trade outcome. In the case of positive externalities, the contract threatens to suspend trade with all agents if any single agent refuses the contract's terms. Thus, each agent becomes pivotal, as in Myerson (1991), whereas the principal serves as a residual claimant. (In a sense, Holmström's famous result is stronger and more surprising; it demonstrates that an efficient outcome can be attained even when it is only possible to make the contract contingent on an aggregate measure of all the agents' actions.) This outcome corresponds to one of the extreme points in the set of equilibria described in our Theorem 1. Since Segal does not admit strategic non-participation decisions by agents prior to the principal's proposal, inefficient free-riding does not occur when contracts are unrestricted.

The model of common agency, due to Bernheim and Whinston (1986), is more distantly related. There, several principals promise conditional payments to a single agent, who does not offer payments. In terms of the game G, it is again quite special, as only one player (the single agent) is taking an action. In terms of contracting, Bernheim and Whinston's model involves unilateral promises rather than multilateral agreements (but note that the promised payments are not restricted to be non-negative, unlike the promises studied by Jackson and Wilkie, 2005). Finally, there is a default action, called non-participation, which never involves payments. If the agent participates, i.e., takes any other action, all the principals' promises must be honored. Translated to our model, it is as if the interaction starts at Stage 2.5, with no contracts having been signed, and each principal then exercising her property rights to price those actions that she owns. Much of Bernheim and Whinston's analysis is devoted to the problems that are posed by non-observability of the agent's action, but it also provides powerful results for the complete information environment. In particular, the common agency model then has equilibrium outcomes in which the contracts jointly create an incentive for the agent to take the efficient action. In fact, all strong equilibrium outcomes have this feature whenever non-observability does not pose a problem; see Bernheim and Whinston's Theorem 2.

Let us turn now to the issue of participation. Participation constraints are at the center of many models of second-best contracting. For example, Mailath and Postlewaite (1990) show that it can be impossible to attain agreement to fund a socially desirable project when agents have private information about their willingness to pay and have the right to refrain from participation. Under symmetric information, as we assume here, this problem would vanish if only all the players could be brought to the negotiation table. In our case, the problem arises precisely because players find it in their interest not to take part in negotiations at all. Therefore, the more relevant comparison is with the work on coalition formation, to which we now turn.

Several strands of literature have investigated the possible inefficiencies that may arise in coalition formation. All strands share the cooperative (i.e., axiomatic) approach to what will be achieved in negotiations, but some of them treat the participation decision in a non-cooperative fashion. Thus, it is natural to consider how our fully non-cooperative approach relates to these partly cooperative approaches.

An early strand of cooperative analysis had revealed that many cooperative games have an empty core. This finding was interpreted to mean that efficiency is not generally implied by voluntary negotiations; see especially Aivazian and Callen (1981). However, as pointed out by Coase (1981) a non-prediction is different from a failure prediction. Only if the prediction is that the outcome might be inefficient will the Coase theorem be overturned.

In a second strand, Aumann and Myerson (1988) study a non-cooperative game in which players' strategies are what bilateral links to form, with the resulting graph representing a cooperation structure, in the sense of Myerson (1977).[31] They show that all stable cooperation structures may involve inefficient outcomes. Intuitively, a subset of players may prefer to cooperate exclusively among themselves, because if additional players are included, the per player payoff is likely to be reduced. This inefficiency result is quite different from our inefficiency results, and depends on the sequential set-up that precludes the players from discussing the terms of trade simultaneously with the formation of links. Moreover, in the Aumann-Myerson framework, the payoff of a player with no links is unaffected by the links that form between other players, whereas in our framework, it is the positive or negative externalities from a contract on outsiders that threaten efficiency.

A third strand of analysis, which is more closely related to our approach, is the literature on non-cooperative coalition formation that emphasizes externalities from coalition members on outsiders. Seminal contributions include d'Aspremont et al (1983), in the context of cartels, Kamien and Zhang (1990) in the context of mergers, and Carraro and Siniscalco (1993) and Barrett (1994) in the context of international environmental agreements. Each of these contributions relates to quite specific situations. With the exception of Kamien and Zhang (1990), they study coalitional stability without a detailed analysis of the contracts between coalition members.

These contributions were followed by a more abstract "second-generation" analysis of coalition-formation and externalities due to, among others, Chatterjee et al (1993), Ray and Vohra (1997), and Seidmann and Winter (1998).[32] In these models, the environment is usually more general and the coalition-formation process more detailed. On the other hand, the game G is suppressed, and in place of our modified game $\widetilde{G}$ there are reduced-form payoffs associated with each coalition structure. Importantly, it is an axiom that coalitions are internally efficient. That is, this literature takes for granted

[31] The Myerson-value is a generalization of the Shapley-value in the sense that it coincides with the Shapley value when all players are connected.

[32] For an extension of this line of work to the setting of ongoing interactions, see Konishi and Ray (2003), Gomes and Jehiel (2005), Bloch and Gomes (2006), and Hyndman and Ray (2007).

that players who eventually engage in contracting will be able to maximize their joint payoff. Any inefficiencies arise because of externalities on players that do not participate in contracting. By contrast, our non-cooperative model of contracting does not make any assumptions about outcomes, and it does not generally imply internal efficiency, or even a unique equilibrium payoff for the players who engage in contracting. The multiplicity of outcomes among negotiators in turn can have consequences for participation in negotiations in our setting. Thus, although some insights about the role of externalities are related, our entirely non-cooperative analysis does not offer a justification for the cooperative elements of these papers.

Finally, our findings are related to the literature on "folk theorems" in repeated games. It has been known at least since Aumann (1959) that in infinitely repeated games without discounting, all feasible and individually rational outcomes can be sustained as Nash equilibria of the supergame. Closely related folk theorems are now known to hold in the limit as discounting goes to zero, and even if equilibria are required to be subgame-perfect; see Fudenberg and Maskin (1986). Aumann (1959) also showed that the solution set is much smaller, and is often empty, if equilibria are required to be strong. However, strong equilibrium is a very restrictive criterion, and several authors have subsequently proposed weaker notions of coalition-proofness. Among those concepts, we focus on consistent equilibrium, because it combines two desirable features; the solution set is never empty, and efficient outcomes are favored.[33] Although our setting is quite different, it shares with the supergame literature the notion that the multiplicity of equilibrium payoffs in future subgames (e.g., the signing stage) can help support cooperation now (e.g., the proposal stage).[34] Likewise, our set of subgame-perfect equilibria has the flavor of a folk theorem, and we can use the same (or similar) notions of multilateral deviation to refine this large equilibrium set. We think that these close analogies are interesting. However, just as our analysis abstracts from enforcement problems – which are the primary focus of the supergames literature – that literature largely abstracts from the contracting process and participation issues that are our main focus.

Concerning the interpretation of our results, perhaps the main question is: *Does free-riding on others' contracts constitute a violation of the Coase theorem?* In the coalition-formation literature described above, several authors have interpreted their results in this way; see Ray and Vohra (1997) for a particularly clear statement.[35] However, others

[33]The first feature is universally desirable. The second feature is desirable for us, because we seek to give the Coase theorem its best possible shot.

[34]Benoît and Krishna (1985) develop this theme in the context of finitely repeated games.

[35]For related discussions, see also Dixit and Olson (2000) and Ellingsen and Paltseva (2012). However, in neither of these two cases is there a watertight argument that outcomes will be inefficient. Dixit and Olson's model has an efficient equilibrium, whereas Ellingsen and Paltseva basically assume that non-participation in contracting also entails a commitment to a particular action in G. While such commitment technologies are plausible in some of the settings that they study, the assumption is clearly restrictive.

argue instead that such free-riding is due to an incomplete specification of property rights; see, e.g., Libecap (2014). Resolving this question requires a clear definition of what it means for property rights to be completely specified. Under our definition, which we believe to be consistent with legal practice as well as with existing definitions in the economics literature, property rights are completely specified whenever (i) it is clear, for each action, who has the right to take or charge for it, and (ii) all actions are perfectly contractible. Since we find that inefficiencies can occur despite unrestricted contractibility (and also maintaining the assumptions complete information and zero transactions costs), and the magnitude of such inefficiency depends on who has what rights, we conclude that Coase was wrong: Well-specified property rights do not suffice to produce efficient final outcomes, and the distribution of property rights across agents does affect the degree of efficiency.

The idea that the allocation of property rights across agents might affect economic efficiency is the centerpiece of the property-rights theory of the firm, due to Grossman and Hart (1986) and Hart and Moore (1990). But in that literature, the distortions are caused either by non-contractible ex ante investments (as in Grossman and Hart, 1986, and Hart and Moore, 1990) or by ex post negotiation failures – due either to disagreement or resentment (as in Hart and Moore, 2008, and Hart, 2009). Here, instead, the distortions are caused by strategic non-participation in negotiations. We see these approaches as fully compatible and complementary.

It is remarkable that in later writings Coase (1974, page 375) himself implicitly came to touch on the impact of the distribution of property rights. In his detailed analysis of British lighthouses, which for many years were largely privately owned, Coase summarizes the smooth workings of private contracting: "In those days, shipowners and shippers could petition the Crown to allow a private individual to construct a lighthouse and to levy a (specified) toll on ships benefitting from it." Notice how, on this occasion, Coase seems to take for granted that it is better for the property rights to lighthouse services to rest with a single party, rather than with all the potential users. Our analysis suggests a simple explanation, not involving transaction costs: The problem of free-riding on others' contracts is smaller under the former ownership allocation than under the latter. Specifically, if all shipowners were to have the right to take their ships wherever they want at no cost, some shipowners or shippers would be better off not participating in negotiations – benefitting instead from the lighthouse services provided by others – even if these services are inferior to those that would have been provided if everyone were to participate in negotiations.

9 Conclusion

We have proposed and analyzed a non-cooperative model of contract negotiations under complete information. Our findings reconcile two apparently conflicting intuitions. The first intuition is that costless contracting ought to admit efficient outcomes if people are willing and able to negotiate. The second intuition is that efficiency may be incompatible with voluntary participation in contract negotiations.

If there are two players, as in all the examples of Coase (1960), we find that the first intuition applies, and the Coase theorem holds. Previous reasons for why the Coase theorem may fail in the case of $n = 2$ are thus argued to rely on restrictive assumptions concerning contract negotiations. For example, Ellingsen and Paltseva (2012) implicitly assume that non-participation in negotiations automatically entails commitment to a particular action in G, whereas Jackson and Wilkie (2005) admit only unilateral promises and not bilateral agreements.

With more than two players, who all own the rights to all their actions, it can be individually rational not to participate in negotiations. When the situation is one of public goods provision, some players will typically be catching a free ride on other players' agreements. These agreements, struck between a subset of the players, in turn will fail to implement efficient outcomes.

The implication is that, except in the case of two players, the Coase theorem should not be invoked as a starting point for the analysis of law and organization. Even in the hypothetical world without transaction costs, the ownership of actions and assets would frequently have efficiency consequences.

A Appendix

A.1 Proof of Theorem 1

To extend the proof to mixed strategies, we need to establish existence of a system of feasible transfers defined on the support of $\hat{\mu}$ such that each player gets exactly the payoff $\pi_j(\hat{\mu})$ by playing any strategy in the support of $\hat{\mu}_j$ given that the other players play $\hat{\mu}_{-j}$. (The definition of the transfers outside the support of $\hat{\mu}$ and the rest of the proof are exactly similar to the pure strategy case.)

Consider any feasible profile of payoffs $\pi(\hat{\mu})$, such that $\pi_j(\hat{\mu}) \geq u_j(G)$. Denote by $n_j \geq 1$ the number of Player $j's$ strategies in the support of $\hat{\mu}_j$, and denote these strategies by $\hat{x}_j^1$, $\hat{x}_j^2$,..., $\hat{x}_j^{n_j}$ respectively. For any pure strategy profile $\hat{x} = (\hat{x}_j, \hat{x}_{-j})$ in the support of $\hat{\mu}$, denote by $\mu_{-j}(\hat{x}_{-j})$ the probability players $k \neq j$ are playing $\hat{x}_{-j}$. That is,

$$\mu_{-j}(\hat{x}_{-j}) = \prod_{k \neq j} \mu_k(\hat{x}_k).$$

Define a system of transfers $t_j(x_j, \hat{x}_{-j})$ on the support of $\hat{\mu}$ in the following way:

(A) For each player j and each $\hat{x}_j^k \in \text{supp}(\hat{\mu}_\text{j})$, $k = 1, ..., n_j$, the expected payoff of Player j is exactly equal to $\pi_j(\hat{\mu})$:

$$\sum_{\hat{x}_{-j} \in \text{supp}(\hat{\mu}_{-j})} \mu_{-j}(\hat{x}_{-j}) \left[U_j(\hat{x}_j^k, \hat{x}_{-j}) - t_j\left(\hat{x}_j^k, \hat{x}_{-j}\right) \right] = \pi_j(\hat{\mu}). \tag{2}$$

These equations ensure that Player j is indifferent between any of her strategies $\hat{x}_j^k$ as long as the other players stick to playing $\hat{\mu}_{-j}$. There are n_j such equations for Player j and $\sum_{j=1}^N n_j$ such equations in total. For notational convenience, denote each of the equations entering subsystem (2) by the strategy whose payoff it represents. For example, we refer to equation

$$\sum_{\hat{x}_{-1} \in \text{supp}(\hat{\mu}_{-1})} \mu_{-1}(\hat{x}_{-1}) \left[U_1(\hat{x}_1^k, \hat{x}_{-1}) - t_1\left(\hat{x}_1^k, \hat{x}_{-1}\right) \right] = \pi_1(\hat{\mu})$$

as equation $(2.\hat{x}_1^k)$.

(B) The transfers are balanced for each pure-strategy profile $\hat{x}$ in the support of mixed-strategy profile $\hat{\mu}$. That is,

$$\sum_{j=1}^N t_j(\hat{x}) = 0. \tag{3}$$

There are $\Pi_{j=1}^N n_j$ such equations. Again, denote each of the equations entering subsystem (2) by the strategy profile it corresponds to (e.g., $(2.\hat{x})$).

If the transfers of each Player $j = 1, ..., N$ solve system (2)-(3) on the support of $\hat{\mu}$, and are defined as follows outside the support of $\hat{\mu}$:

$$\hat{t}_j(x) = \begin{cases} (n-1)h & \text{if } x_j \notin \text{supp}(\hat{\mu}_j) \text{ and } x_{-j} \in \text{supp}(\hat{\mu}_{-j}) \ ; \\ -h & \text{if } x_j \in \text{supp}(\hat{\mu}_j) \text{ and } |\{k : x_{-j} \notin \text{supp}(\hat{\mu}_{-j})\}| = 1; \\ 0 & \text{otherwise,} \end{cases}$$

then the rest of the proof is exactly the same as in case of sustaining a pure strategy.

Lemma A1 *System (2)-(3) always has a solution.*

Proof. We will show that the system (2)-(3) is indeterminate (i.e., that there are more unknowns than equations) and consistent, and thus always has a solution.

There are

$$\prod_{j=1}^{N} n_j + \sum_{j=1}^{N} n_j$$

linear equations in the system of equations (2)-(3), and $N \prod_j n_j$ unknown transfer values $t_j(\hat{x}_j, \hat{x}_{-j})$. However, the system (2)-(3) is not linearly independent. Indeed, the feasibility of payoffs combined with the balanced transfer scheme (3) imply that the weighted sum of equations (2) with weights equal to $\mu_j(\hat{x}_j^k)$ respectively, and taken across all players $j = 1, ..., N$ is equal to zero:

$$\begin{aligned} & \sum_j \left[\sum_{k=1,...,n_j} \mu_j(\hat{x}_j^k) \sum_{\hat{x}_{-j} \in \text{supp}(\hat{\mu}_{-j})} \mu_{-j}(\hat{x}_{-j}) \left[U_j(\hat{x}_j^k, \hat{x}_{-j}) - t_j\left(\hat{x}_j^k, \hat{x}_{-j}\right) \right] \right] \\ = & \sum_j \left[\sum_{k=1,...,n_j} \sum_{\hat{x}_{-j} \in \text{supp}(\hat{\mu}_{-j})} \left[\mu_j(\hat{x}_j^k) \mu_{-j}(\hat{x}_{-j}) \right] \left[U_j(\hat{x}_j^k, \hat{x}_{-j}) - t_j\left(\hat{x}_j^k, \hat{x}_{-j}\right) \right] \right] \\ = & \sum_j \left[\sum_{\hat{x} \in \text{supp}(\hat{\mu})} \mu(\hat{x}) \left[U_j(\hat{x}) - t_j(\hat{x}) \right] \right] \\ = & \sum_j \pi_j(\hat{\mu}) - \left[\sum_{\hat{x} \in \text{supp}(\hat{\mu})} \mu(\hat{x}) \sum_j t_j(\hat{x}) \right] = \sum_j \pi_j(\hat{\mu}). \end{aligned}$$

Thus, let us drop equation $(2.\hat{x}_1^1)$ – the equation corresponding to Player 1 playing strategy $\hat{x}_1^1$ - from consideration, and prove that the remaining system of equations has full rank of $\prod_j n_j + \sum_j n_j - 1$. For each of the pure strategy profiles $\hat{x}$ in the support of $\hat{\mu}$, consider transfer of Player 1, $t_1(\hat{x})$. This transfer enters one (and only one) equation $(3.\hat{x})$ from subsystem (3). Moreover, for $\hat{x} = \left(\hat{x}_1^k, \hat{x}_{-1}\right)$, $k = 2, ..., n_1$, the respective transfer

$t_1\left(\hat{x}_1^k, \hat{x}_{-1}\right)$ also enters equation $(2.\hat{x}_1^k)$ of subsystem (2),

$$\sum_{\hat{x}_{-1}\in \text{supp}(\hat{\mu}_{-1})} \mu_{-1}(\hat{x}_{-1}) \left[U_1(\hat{x}_1^k, \hat{x}_{-1}) - t_1\left(\hat{x}_1^k, \hat{x}_{-1}\right)\right] = \pi_1(\hat{\mu}).$$

Notice that this is only true for $k = 2, ..., n_1$, as we just dropped the equation $(2.\hat{x}_1^1)$ from our system. Using equations $(3.\left(\hat{x}_1^k, \hat{x}_{-1}\right))$ for all possible $\hat{x}_{-1}$ to exclude unknown $t_1\left(\hat{x}_1^k, \hat{x}_{-1}\right)$ from equation $(2.\hat{x}_1^k)$ transforms $(2.\hat{x}_1^k)$ into

$$\sum_{\hat{x}_{-1}\in \text{supp}(\hat{\mu}_{-1})} \mu_{-1}(\hat{x}_{-1}) \left(U_1(\hat{x}_1^k, \hat{x}_{-1}) + \left[\sum_{j>1} t_j\left(\hat{x}_1^k, \hat{x}_{-1}\right)\right]\right) = \pi_1(\hat{\mu})$$

for $k = 2, ..., n_1$. Rewrite the above as

$$\sum_{\hat{x}_{-1}\in \text{supp}(\hat{\mu}_{-1})} \mu_{-1}(\hat{x}_{-1}) \left[\sum_{j>1} t_j\left(\hat{x}_1^k, \hat{x}_{-1}\right)\right] = \pi_1(\hat{\mu}) - U_1(\hat{x}_1^k, \hat{\mu}_{-1}),$$

and denote this transformed equation $(2.\hat{x}_1^k)'$. Notice that the new linear system of equations obtained from the system (2)-(3) by replacing equations $(2.\hat{x}_1^k)'$ by equations $(2.\hat{x}_1^k)$ for $k = 2, ..., n_1$ is equivalent to the the original system (2)-(3). In this new system, the transfers $t_1(\hat{x})$ of Player 1 enter only the subsystem (3), one per equation. Thus, the equations in this subsystem are linearly independent and independent of the rest of the system. Thus, subsystem (3) contributes $\prod_j n_j$ to the system (2)-(3) rank. Also, in determining the rank of the (2)-(3) system, we can continue with the analysis of its remaining part, the new version of subsystem (2) - the version in which equations $(2.\hat{x}_1^k)$ are replaced by equations $(2.\hat{x}_1^k)'$.

Consider the new version of subsystem (2). Notice that Player 2's transfer $t_2\left(\hat{x}_1^1, \hat{x}_2^k, ..., \hat{x}_N^1\right)$ enters only the equation $(2.\hat{x}_2^k)$, $k = 1, ..., n_2$ (they would also enter the transformed version of equation $(2.\hat{x}_1^1)$, but recall that it is dropped from the system). Thereby, equations $(2.\hat{x}_2^k)$, $k = 1, ..., n_2$ are linearly independent among themselves, and with the remaining equations of the subsystem. They contribute n_2 to the rank of the system. Similarly to above, we can continue with the analysis of the rank of system's remaining part - subsystem (2) in which equations $(2.\hat{x}_1^k)$ are replaced by equations $(2.\hat{x}_1^k)'$ for $k = 2, ..., n_1$, and equations $(2.\hat{x}_2^k)$, $k = 1, ..., n_2$ are dropped.

Now, inspect equations $(2.\hat{x}_1^k)'$, $k = 2, ..., n_1$. Transfer $t_2\left(\hat{x}_1^2, \hat{x}_2^k, ...\right)$ enters only equation $(2.\hat{x}_1^2)'$, transfer $t_2\left(\hat{x}_1^3, \hat{x}_2^k, ...\right)$ enters only equation $(2.\hat{x}_1^3)'$, etc. In other words, these equations are also linearly independent, and independent of the remaining subsystem. They contribute $n_1 - 1$ to the rank of the system, and can be dropped from consideration.

Finally, consider any of the remaining $\sum_{j=3}^N n_j$ equations, e.g., $(2.\hat{x}_j^k)$. It contains a term $t_j\left(\hat{x}_1^1, \hat{x}_2^1, ..., \hat{x}_j^k, ..., \hat{x}_N^1\right)$ that does not enter any of the remaining $\sum_{j=3}^N n_j - 1$ equations. Thus, the remaining subsystem is also linearly independent.

Hence, we have shown that the rank of the considered system is equal to

$$\prod_{j=1}^N n_j + n_2 + n_1 - 1 + \sum_{j=3}^N n_j = \prod_{j=1}^N n_j + \sum_{j=1}^N n_j - 1.$$

It remains to show that the number of variables exceeds the rank of the system, that is

$$N \prod_{j=1}^N n_j \geq \prod_{j=1}^N n_j + \sum_{j=1}^N n_j - 1$$

for any $n_j \geq 1$ and $N \geq 2$. Indeed

$$\begin{aligned} & N \prod_{j=1}^N n_j - \left[\prod_{j=1}^N n_j + \sum_{j=1}^N n_j - 1\right] = (N-1) \prod_{j=1}^N n_j - \sum_{j=1}^N n_j + 1 \\ = & \prod_{j=1}^N n_j - (n_{N-1} + n_N) + 1 + \sum_{j=1}^{N-2} \left(\prod_{k=1}^N n_k - n_j\right) \\ \geq & n_{N-1} n_N - (n_{N-1} + n_N) + 1 = (n_{N-1} - 1)(n_N - 1) \geq 0. \end{aligned}$$

Thereby, our system (2)-(3) is indeterminate. It is clearly consistent, as the rank of its augmented matrix is also equal to $\prod_{j=1}^N n_j + \sum_{j=1}^N n_j - 1$. Thereby, it always has a solution. ■

A.2 Proof of Theorem 6

We begin by proving a sequence of lemmas.

Lemma A2 *A "punishment" strategy* $\widetilde{\mu}^{BR_i} = \arg\max_{\mu \in \Delta X^{BR_i}} \sum_{j \neq i} U_j(\mu)$ *exists and belongs to* ΔX^{BR_i}.

Proof. Consider some strategy μ_0 in ΔX^{BR_i}. Such a strategy exists (e.g., any undominated NE would be an example of such an strategy profile). If the joint payoff of players $j \neq i$ at this strategy profile μ_0 cannot be exceeded by their joint payoff in any other strategy μ_1 in ΔX^{BR_i}, then $\mu_0 = \arg\max_{\mu \in \Delta X^{BR_i}} \sum_{j \neq i} U_j(\mu) = \widetilde{\mu}^{BR_i}$ and the result is proven.

Otherwise, consider the subset $D_i(\mu_0) \in \Delta X$ of all strategy profiles that give higher joint payoff of players $j \neq i$ than at strategy profile μ_0. Consider the set $\mathbb{D} \subset \mathbb{R}$ of values of *joint* payoff of players $j \neq i$ for the strategy profiles in $D_i(\mu_0)$. By the completeness axiom, there exists a supremum of this set

$$\overline{d} = \sup \mathbb{D} = \sup_{\mu \in D_i(\mu_0)} \left[\Sigma_{j \neq i} U_j(\mu) \right],$$

Let us show that this supremum can be achieved at some strategy profile that also belongs to $D_i(\mu_0)$ (i.e., the set of strategy profiles that maximize the joint payoff of players $j \neq i$ is non-empty).

By the definition of supremum, there exists a sequence $\{\mu_n\} \in D_i(\mu_0)$ such that

$$\overline{d} - 1/n < \Sigma_{j \neq i} U_j(\mu_n) \leq \overline{d}.$$

This sequence $\{\mu_n\}$ is bounded in the metric space of all strategies $\mathbb{R}^{\times_i X_i}$, since a (mixed) strategy of any player k can be represented as a vector of weights in $|X_k|$- dimensional space of pure strategies of Player k, with weights between 0 and 1 (and all weights summing to 1). Thus, it contains a converging subsequence $\{\mu_{n_k}\}$. Denote the limit $\hat{\mu}$. Recall that the non-participating Player i is always playing a best response at ΔX^{BR_i}. As the best response correspondence in G – a finite game with continuous payoffs – is upper hemi-continuous, Player i's limit strategy $\hat{\mu}_i$ belongs to her best responses to $\hat{\mu}_{-i}$ in G. That is, $\hat{\mu}$ also belongs to ΔX^{BR_i}. Further, by continuity of the payoff functions, the vector of payoffs of players converges to their payoff at $\hat{\mu}$,

$$\lim_{k \to \infty} \Sigma_{j \neq i} U_j(\mu_{n_k}) = \Sigma_{j \neq i} U_j(\hat{\mu}).$$

Now, if the set of strategy profiles that maximize the joint payoff of players $j \neq i$ over all profiles in ΔX^{BR_i} is not a singleton, repeat the above argument to select a subset of it that maximizes the payoff of Player i. ■

Lemma A3 *The payoff $f_i(G)$ constitutes the worst punishment for Player i's non-participation that can be implemented by participating players $j \neq i$ in a consistent equilibrium.*

Proof. First, notice that the non-participating Player i would choose to play a best response to any action of the participating coalition. That is, a strategy profile in a "punishing" consistent equilibrium should belong to ΔX^{BR_i}.

Second, let's show that the strategy profile $\widetilde{\mu}^{BR_i} \in \Delta X^{BR_i}$ that yields the highest *joint* payoff to players $1, 2, ...i-1, i+1, ...n$, and if there are multiple such strategies, is the best for Player i

$$\widetilde{\mu}^{BR_i} = \arg \max_{\mu \in \Delta X^{BR_i}} \sum_{j \neq i} U_j(\mu),$$

can indeed be supported in a consistent equilibrium. Consider the following equilibrium construction (very similar to the one in Theorem 1):

Stage 1: Each player $j \neq i$ makes the proposal $\tau_j = t$ that supports the strategy profile $\widetilde{\mu}^{BR_i} \in \Delta X$ along the equilibrium path.

Stage 2. If $\tau_1 = t$, each player $j \neq i$ signs τ_1. If $\tau_1 \neq \tau_2 = t$, each player signs τ_2, etc. If no player offers t, players coordinate on some signing-profile that forms a consistent equilibrium of the respective subgame (and, thus, has a Pareto-undominated continuation). If there are several such profiles, choose the one that is the best for Player 1. If there are several of those, choose the one that is best for Player 2, etc. As each signing-stage subgame is finite, backward induction ensures existence of such signing profile.

Stage 3. (i) If some proposal $\tau_j = t$ was signed by all players $j \neq i$, play $\widetilde{\mu}^{BR_i}$. (ii) if $\tau_j = t$ is signed by all but Player k, play Player k's worst Pareto-undominated Nash equilibrium in the resulting game G. (iii) In all other situations, play the worst Pareto-undominated Nash equilibrium from the perspective of Player 1 (if there are multiple such equilibria, play the worst of them from the perspective of Player 2, etc.).

Let us show that this is indeed an equilibrium, and that it has Pareto-undominated continuations at each proper subgame. Start with Stage 3 and move forwards.

At Stage 3, if $t^S = t$, $\widetilde{\mu}^{BR_i}$ represents a Pareto-undominated Nash equilibrium of the game. Indeed, notice that any undominated Nash equilibrium of G belongs to ΔX^{BR_i}. By construction, the joint payoff of players $j \neq i$ at $\widetilde{\mu}^{BR_i}$ is at least as high as at any Pareto-undominated Nash equilibrium. Thus, there is a split of this joint payoff such that each player $j_i \in N'$ gets a payoff that weakly exceeds her payoff in her worst undominated Nash equilibrium, making unilateral deviation unprofitable. Also, as $\widetilde{\mu}^{BR_i}$ maximizes the *joint* payoff to players $j \neq i$ on ΔX^{BR_i}, $\widetilde{\mu}^{BR_i}$ is undominated (conditional on Player i's non-participation). In all other situations, the rule above prescribes an undominated Nash equilibrium, so no player has any incentive to unilaterally deviate, and no Pareto improvement is possible.

At Stage 2, consider first the branch along which $\tau_1 = t$. Then, a unilateral deviation by some Player $k \neq 1, i$ (not to sign τ_1) entails $t^S = \varnothing$, and Player k's worst Pareto-undominated equilibrium of G being played at Stage 3, so the deviation is not profitable. This takes care of deviations on the equilibrium path. Off the path, an analogous argument applies along the branches $\tau_1 \neq \tau_2 = t'$, etc. Notice that in these cases the equilibria have undominated continuations at each proper Stage 3 subgame. Finally, along any other off-equilibrium branch, by definition the signing decisions are in line with consistent equilibria that have Pareto-undominated continuations at each proper Stage 3 subgame.

At Stage 1, only Player $1'$s deviations $\tau_{j1} \neq t$ affect the subsequent play. After such a deviation, $\tau_2 = t$ will be signed instead, entailing exactly the same outcome as if Player 1 does not deviate. Thus, this is not a profitable deviation, so our suggested strategy profile is an equilibrium. As argued above, there could be no other Pareto-improving equilibrium in this subgame of $\Gamma^V(G)$.

Finally, there could not be a worse punishment in a consistent equilibrium. Indeed, recall, that all "punishment" equilibria should belong to ΔX^{BR_i}. Thus, a consistent "punishment" equilibrium that would entail a payoff for Player i' below

$$f_i(G) = U_i(\widetilde{\mu}^{BR_i})$$

would imply that players $j \neq i$ obtain a higher joint payoff than at $\widetilde{\mu}^{BR_i}$, which contradicts the definition of $f_i(G)$. ■

With these lemmas in hand, we may complete the proof of Theorem 6.

(a) Assume that condition

$$\sum_{i \in N} U_i(x^*) \geq \sum_{i \in N} f_i(G) \tag{4}$$

is met. We can then demonstrate that the following is an efficient consistent equilibrium of $\Gamma^V(G)$.

Participation stage: All players choose to participate.

Contract proposal stage: All players who decide to participate propose the same contract t_{N^P}; the specification of this contract depends on the set of participating players. Specifically, (i) If all players participate, then each player proposes a contract t that specifies for each player $j \in N$ the following transfers at x^*:

$$f_i(G) - U_j(x^*) + \frac{1}{|N|}\left(\sum_{i \in N} U_i(x^*) - \sum_{i \in N} f_i(G)\right), \tag{5}$$

and establishes punishments for unilateral deviations from x^*, so that x^* is supported as an equilibrium allocation. (ii) If there is a single non-participant i, then players $j \neq i$ choose the contract that supports the profile $\widetilde{\mu}^{BR_i}$. If the set of non-participants comprises two or more players, each participant (member of N^P) proposes a contract that supports the profile μ^{N_p} with the highest joint payoff to the members of N^P among all the action profiles in which players N/N^P play consistent best responses (which exists and can be supported by the arguments very similar to the above lemmas). Finally, if only one player participates, no contract is proposed.

Contract signing stage: Enumerate all players belonging to N^P as j_1, j_2,...,j_{N^P}. If $\tau_{j_1} = t_{N^P}$ (as proposed above), each player $j \in N^P$ signs τ_{j1}. If $\tau_{j_1} \neq \tau_{j_2} = t_{N^P}$, each player signs τ_{j_2}, etc. In case all participants deviate at the proposal stage, let their signing coordinate on some signing profile that is compatible with a consistent equilibrium continuation (such a continuation always exists, as consistent equilibria always exist, but its exact form depends on the deviation proposals).

Implementation stage: If some proposal $\tau_{j_i} = t_{N^P}$ (as described above) was signed by all players $j \in N^P$, play the profile suggested by this contract. If a single player $j_k \in N^P$ fails to sign this contract, play Player j_k's worst undominated Nash equilibrium of the resulting game. In all other situations, play the worst undominated Nash equilibrium from the perspective of Player j_1 (if there are multiple such equilibria, play the worst of them from the perspective of Player j_2, etc.). If only one player participates, play this player's worst undominated Nash equilibrium of G.

Let us show that this is indeed an efficient consistent equilibrium (in particular, that it has Pareto-undominated continuations at each proper subgame). Start with the implementation stage and move backwards.

Implementation stage: If no more than one person has chosen to participate, an undominated Nash equilibrium is chosen. Suppose now that $N^P \geq 2$ players participate. If $t^S = t_{N^P}$, the resulting strategy profile represents an undominated Nash equilibrium of the game. Indeed, if $N^P = N$, the resulting strategy profile is x^*, and the allocation rule (5) ensures that no unilateral deviation is profitable. As x^* is efficient, it cannot be renegotiated either. If $N^P \neq N$, then the contract implements an action profile that constitutes an non-renegotiable equilibrium of resulting game. Also, if $t^s \neq t_{N^P}$, the implemented profile is consistent by construction.

Contract signing stage: Consider first the branch along which $\tau = t_{N^P}$. Then, a unilateral deviation by some Player $j_2 \neq j_1$, $j_2 \in N^P$ (not to sign τ) entails $t^S = \varnothing$, and Player j_2's worst undominated equilibrium of G being played at Stage 3, so the deviation is not profitable. Along any other off-equilibrium branch, by definition the signing decisions are part of Pareto-undominated equilibria that have undominated continuations at each proper 3rd stage subgame, and, as above, they exist because each signing game is finite.

Contract proposal stage: Only Player j_1 deviations $\tau_1 \neq t_{N^P}$ affect the subsequent play. After such a deviation, $\tau_2 = t_{N^P}$ will be signed instead, entailing exactly the same outcome as if Player j_1 does not deviate. Thus, this is not a profitable deviation, so our suggested strategy profile is an equilibrium. In turn, there could be no another, Pareto-improving equilibrium in each subgame of $\Gamma^V(G)$ where players $j \in N^P$ choose to participate.

Participation stage: No player from N may find it profitable to deviate and non-participate, as it would decrease her payoff. Also, as x^* is efficient, there is no Pareto improving equilibrium in the entire game.

(b) Now assume that condition (4) does not hold. If a player k chooses to participate, she needs to obtain at least $f_i(G)$ in resulting consistent equilibrium. Indeed, if she refuses to take part, she cannot be punished more than by $f_i(G)$. So, condition (4) simply means that no contract supporting x^* can deliver sufficient payoff to all participating parties.

A.3 Proof of Theorem 7

We seek to demonstrate that x^* can be supported in an efficient consistent equilibrium of $\Gamma^V(G(p))$. Consider the following equilibrium:

- Participation stage: all players choose not to participate
- Contract proposal stage (off the equilibrium path): Let N^P denote the set of players that negotiate (participate). All players in N^P propose a zero transfer contract, call it t'.
- Contract signing stage (off the equilibrium path): Enumerate participants as j_1, j_2,...,j_{N^P}. If $\tau_{j_1} = t'$, each participant signs τ_{j1}. If $\tau_{j_1} \neq \tau_{j_2} = t'$, each participant signs τ_{j_2}, etc. In case of a N^P-lateral proposal deviation, participants sign in a consistent manner (more below).
- Implementation stage: If no more than one person has chosen to participate in negotiations, or if some proposal $\tau_{\ j_i} = t'$ was signed by all players $j \in N^P$, play x^*. If $\tau_{j_i} = t'$ is not signed by Player j_k only, play Player j_k's worst undominated Nash equilibrium in the resulting game. In all other situations, play the worst undominated Nash equilibrium from the perspective of Player j_1 (if there are multiple such equilibria, play the worst of them from the perspective of Player j_2, etc.).

We now check that this strategy profile indeed forms an efficient consistent equilibrium of $\Gamma^V(G(p))$ (in particular, that it has undominated continuations at each proper subgame). Start with the implementation stage and move backwards.

Implementation stage: If at most one player has participated in negotiations, players face $G(p)$ and x^* is trivially the unique Nash equilibrium. Since x^* is efficient, it is also a consistent equilibrium of $G(p)$. If $t^S = t'$, x^* is again the unique Nash equilibrium, as the renegotiated game remains $G(p)$. In all other situations $\tilde{G}(p)$, players are supposed to play an undominated Nash equilibrium, so no player has any incentive to unilaterally deviate, and no Pareto-improvement is possible.

Contract signing stage: Consider first the branch along which $\tau_{j_1} = t'$. Then, a unilateral deviation by some Player j_i, $j_i \in N^P$ (not to sign τ_{j_1}) entails $t^S = \varnothing$, and Player j_i's worst undominated equilibrium of $G(p)$ being played at Stage 3. This deviation is thus not profitable. An analogous argument applies along the branch $\tau_{j_1} \neq \tau_{j_2} = t'$. Notice that in both these cases the equilibria have undominated continuations at each proper 3-stage subgame. Finally, along any other off-equilibrium branch, by definition the signing decisions entail consistent equilibrium outcomes, which are supported by undominated continuations at each proper 3rd stage subgame. (Again, these exist because each signing game is finite. Thus, backward induction can be used to find them.)

Contract proposal stage: Given the anticipated signing decisions, only Player j_1 deviations $\tau_{j_1} \neq t'$ affect the subsequent play. After such a deviation, $\tau_{j_2} = t'$ will be signed instead, entailing exactly the same outcome as if Player j_1 does not deviate. Moreover, regardless of future signing decisions, there is no other contract that could entail an ultimate Pareto-improvement, as the profile x^* is efficient.

Participation stage: as all players choose not to participate along the equilibrium path, a unilateral deviation cannot be profitable. Also, as x^* is efficient, there is no Pareto improving equilibrium in the entire game, which concludes the proof.

A.4 Proof of Lemma 1

Denote the strategy profile that yields Player i's pure strategy maximin in G by $x^{i,m}$. Consider the transfer functions of players $j = 1, ..., i-1, i+1, ..., n$ to players $k \neq j$

$$T_{jk}^{-i}(x) = \begin{cases} 2h & \text{if } x_j \neq x_j^{i,m}; \\ 0 & \text{otherwise.} \end{cases}$$

Assuming that Player i does not promise any transfers, $T_{jk}^{-i}(x)$ ensure that $x_j^{i,m}$ is a dominant strategy for players $j = 1, ..., i-1, i+1, ..., n$. This, together with $x^{i,m}$ being the maximin strategy profile for Player i, implies that $x^{i,m}$ is a Nash equilibrium of the resulting game, with the payoff to Player i given by

$$U_i^G(x^{i,m}) = v_i^p(G) = \max_{x_i} \min_{x_j} U_i^G(x_i, x_j).$$

The rest of the proof effectively repeats the proof of Theorem 4 in Jackson and Wilkie (2005). Specifically, let us show that Player i cannot increase her payoff by offering some transfer function $T_i'(.) \neq 0$. This can only be improving if it leads to play of something other than $x_j^{i,m}$ by some player $j = 1, ..., i-1, i+1, ..., n$ (as $x^{i,m}$ is a maximin for Player i so i cannot do better by unilaterally changing her action). First, consider the case where a pure strategy Nash equilibrium $\hat{x}$ is played at the action stage, where $\hat{x}_j \neq x_j^{i,m}$ for some player $j \neq i$. Let there be $K \geq 1$ players $j \neq i$ such that $\hat{x}_j \neq x_j^{i,m}$, and consider some such j. Player j's pay-off from the profile $\hat{x}$ is

$$U_j^G(\hat{x}) - (n-1)2h + 2h(k-1) + T_{ij}'(\hat{x}).$$

By playing $x_j^{i,m}$ instead she gets

$$U_j^G(x_j^{i,m}, \hat{x}_{-j}) + 2h(k-1) + T_{ij}'(x_j^{i,m}, \hat{x}_{-j}).$$

As $\hat{x}$ is a Nash equilibrium,

$$T'_{ij}(\hat{x}) - T'_{ij}(x_j^{i,m}, \hat{x}_{-j}) \geq U_j^G(x_j^{i,m}, \hat{x}_{-j}) - U_j^G(\hat{x}) + (n-1)2h.$$

By the definition of h, and the fact that $n \geq 2$, it follows that

$$T'_{ij}(\hat{x}) > 3h + T'_{ij}(x_j^{i,m}, \hat{x}_{-j}) \geq 3h$$

for any j such that $\hat{x}_j \neq x_j^{i,m}$. So, by the definition of h and the fact that $K \geq 1$, Player i's payoff in $\hat{x}$ is at most

$$U_i^G(\hat{x}) - 3hK + 2hK \leq U_i^G(\hat{x}) - h < U_i^G(x^{i,m}) = v_i^p(G).$$

Hence, such a deviation cannot be profitable. When $\hat{x}$ is a mixed strategy equilibrium, the result is proved by using a similar argument for each strategy in the support of $\hat{x}_j$.

A.4.1 Proof of Lemma 2

Let us prove this result for $i = 1$ and $j = 2$ (the reverse is proved in exactly the same way). Denote by μ_1^m the strategy of Player 1 that ensures the maximin of Player 2 (if there are several such strategies, pick one). Assume that Player 1 mixes between two unilateral transfer promises. (i) One transfer promise, $\underline{T}_1(\mu)$, satisfies

$$\begin{aligned}
\underline{T}_1(\mu_1^m, \mu_2) &= 0, \ \forall\ \mu_2 \in \Delta(X_2); \\
\underline{T}_1(\mu_1, \mu_2) &= \max(U_1(\mu_1, \mu_2) - U_1(\mu_1^m, \mu_2) + \delta, 0), \ \forall\ \mu_1 \neq \mu_1^m,\ \mu_2 \in \Delta(X_2).
\end{aligned}$$

(where $0 < \delta \ll 1$). Here, if Player 2 proposes no contract, μ_1^m becomes the dominant strategy for Player 1. (ii) The other promise, $\overline{T}_1(\mu)$, satisfies

$$\begin{aligned}
\overline{T}_1(\mu_1^m, \mu_2) &= 0, \ \forall\ \mu_2 \in \Delta(X_2); \\
\overline{T}_1(\mu_1, \mu_2) &= \underline{T}_1(\mu_1, \mu_2) + l, \ \forall\ \mu_1 \neq \mu_1^m,\ \mu_2 \in \Delta(X_2).
\end{aligned}$$

That is, if Player 2 does not offer a contract, μ_1^m dominates all the other strategies available to Player 1 by at least some (large) $l > 0$.

To characterize the highest payoff Player 2 can achieve by responding to mixtures on this support, it is useful to first consider the payoffs from responding to $\underline{T}_1$ and $\overline{T}_1$ separately.

Lemma A4 *If Player 1 has made the proposal $\underline{T}_1$, an upper bound of the payoffs that*

Player 2 can achieve through any clause $T_2(\mu)$ is

$$V_2 = \max_{\mu_2 \in \Delta(X_2)} \left(\left[\max_{\mu_1 \in \Delta(X_1)} U_1(\mu_1, \mu_2) + U_2(\mu_1, \mu_2) \right] - U_1(\mu_1^m, \mu_2) \right).$$

Proof. The proof is by contradiction. Suppose that there is some proposal T_2 such that the game $\widetilde{G}(\underline{T}_1, T_2)$, has a (worst for Player 2) Nash Equilibrium $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ yielding a payoff to Player 2

$$U_2^G(\widetilde{\mu}_1, \widetilde{\mu}_2) + \underline{T}_1(\widetilde{\mu}_1, \widetilde{\mu}_2) - T_2(\widetilde{\mu}_1, \widetilde{\mu}_2) > V_2.$$

As Player 1 made no transfers for the strategy μ_1^m, Player 1 must get at least as much in this equilibrium as she would get by playing μ_1^m in the original situation G; otherwise Player 1 would choose to deviate. That is,

$$U_1^G(\widetilde{\mu}_1, \widetilde{\mu}_2) - \underline{T}_1(\widetilde{\mu}_1, \widetilde{\mu}_2) + T_2(\widetilde{\mu}_1, \widetilde{\mu}_2) \geq U_1^G(\mu_1^m, \widetilde{\mu}_2).$$

Summing up these two inequalities we get

$$U_1^G(\widetilde{\mu}_1, \widetilde{\mu}_2) + U_2^G(\widetilde{\mu}_1, \widetilde{\mu}_2) > U_1^G(\mu_1^m, \widetilde{\mu}_2) + V_2,$$

or equivalently,

$$U_1^G(\widetilde{\mu}_1, \widetilde{\mu}_2) + U_2^G(\widetilde{\mu}_1, \widetilde{\mu}_2) - U_1(\mu_1^m, \widetilde{\mu}_2) > V_2,$$

which contradicts the definition of V_2. ■

Lemma A5 *If Player 1 has made the proposal $\overline{T}_1$, an upper bound of the payoffs that Player 2 can achieve through any clause $T_2(\mu)$ is given by V_2.*

Proof. Completely analogous to the previous argument. ■

Assume now that Player 1 mixes between $\underline{T}_1$ and $\overline{T}_1$ with probabilities

$$p = \frac{V_2 - v_2}{V_2 - v_2 + l}, \quad 1 - p = \frac{l}{V_2 - v_2 + l}.$$

It turns out that there is an upper bound to how much Player 2 can gain in playing against this mix of $\underline{T}_1$ and $\overline{T}_1$.

Lemma A6 *If Player 1 randomizes over $\underline{T}_1$ and $\overline{T}_1$ with respective probabilities p and $1 - p$, Player 2 can not obtain an expected payoff above*

$$v_2 + \frac{(V_2 - v_2)^2}{(V_2 - v_2 + l)}.$$

Proof. Consider any contract $T_2'(\mu)$. There are two cases, depending on whether Player 1 plays the strategy μ_1^m in Player 2's worst Nash equilibrium of $\widetilde{G}(\overline{T}_1, T_2')$.

First, assume that Player 1 plays μ_1^m in this equilibrium. Then the payoff of Player 2 cannot exceed $U_2(\mu_1^m, \mu_2^m) = v_2$ (as Player 1 makes no transfers conditional on her playing μ_1^m). In turn, Lemma A4 implies that the payoff of Player 2 in $\widetilde{G}(\underline{T}_1, T_2')$ cannot exceed V_2. Therefore, the payoff of Player 2 from playing T_2' does not exceed

$$pV_2 + (1-p)v_2,$$

or equivalently

$$\frac{(V_2 - v_2)\, V_2}{V_2 - v_2 + l} + \frac{l v_2}{V_2 - v_2 + l} = v_2 + \frac{(V_2 - v_2)^2}{V_2 - v_2 + l}.$$

Now assume that Player 1 plays a (possibly mixed) strategy $\widetilde{\mu}_1 \neq \mu_1^m$ in the equilibrium of $\widetilde{G}(\overline{T}_1, T_2')$ (and Player 2 plays some $\widetilde{\mu}_2$). Then by Lemma A5 the payoff of Player 2 in this equilibrium cannot exceed V_2. Further, the strategy profile $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ is also an equilibrium in the situation $\widetilde{G}(\underline{T}_1, T_2')$. Indeed, for each strategy profile μ not involving μ_1^m, the contracts $\underline{T}_1(\mu)$ and $\overline{T}_1(\mu)$ differ by l (by definition of $\overline{T}_1$and $\underline{T}_1$). This implies that Player 2's payoff in $\widetilde{G}(\overline{T}_1, T_2')$ exceeds that in $\widetilde{G}(\underline{T}_1, T_2')$ by exactly l, and Player 1's payoff in $\widetilde{G}(\overline{T}_1, T_2')$ is below that in $\widetilde{G}(\underline{T}_1, T_2')$ by the same l. Since Players 1 and 2 do not have a profitable deviation from $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ in $\widetilde{G}(\overline{T}_1, T_2')$, they cannot have one in $\widetilde{G}(\underline{T}_1, T_2')$ either. This argument implies that the payoff of Player 2 associated with the equilibrium $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ of $\widetilde{G}(\underline{T}_1, T_2')$ is at most

$$V_2 - l.$$

Therefore, the payoff of Player 2 from promising $T_2'(\mu)$ does not exceed

$$p\,(V_2 - l) + (1-p)V_2,$$

which can be rewritten as

$$\frac{(V_2 - v_2)}{V_2 - v_2 + l}\,(V_2 - l) + \frac{l}{V_2 - v_2 + l} V_2 = v_2 + \frac{(V_2 - v_2)^2}{V_2 - v_2 + l}.$$

■

Now, it is enough notice that for any $d > 0$ one can choose

$$l(d) = \max\left[0, (V_2 - v_2)\left(\frac{(V_2 - v_2)}{d} - 1\right)\right].$$

Denoting by $T_{1,d}(\mu)$ the mix of $\underline{T}_1$ and $\overline{T}_1$ with respective probabilities p and $1-p$ corresponding to $l(d)$ (the transfer functions are denoted $\underline{T}_{1,d}$ and $\overline{T}_{1,d}$ respectively) completes the proof, except for the transfer bound.

To establish the upper bound on transfers, observe that

$$\max_{\mu}\left(\underline{T}_{1,d}(\mu),\overline{T}_{1,d}(\mu)\right)=\max_{\mu}\left(\underline{T}_{1,d}(\mu)\right)+l(d))<1+\max_{\mu',\mu''}(U_1(\mu')-U_1(\mu''))+l(d)\leq h+l(d).$$

Evaluate $l(d)$:

$$l(d)\leq(V_2-v_2)\left(\frac{(V_2-v_2)}{d}-1\right).$$

Notice that

$$V_2-v_2=\max_{\mu_2\in\Delta(X_2)}\left(\left[\max_{\mu_1\in\Delta(X_1)}U_1(\mu_1,\mu_2)+U_2(\mu_1,\mu_2)\right]-U_1(\mu_1^m,\mu_2)\right)-v_2<2h.$$

As a result,

$$l(d)\leq 2h\left(\frac{2h}{d}-1\right)$$

and

$$\max_{i,\mu}\left(\underline{T}_{i,d}(\mu),\overline{T}_{i,d}(\mu)\right)<h+l(d)\leq h\left(4\frac{h}{d}-1\right).$$

A.4.2 Proof of Theorem 10, case $d=0$

If $d=0$, $U_1(x^*)+U_2(x^*)=(v_1+v_2)$. Since $v_i\leq u_i$, all Nash equilibria of G are thus efficient, with payoffs $U_i(x^*)=v_i$. We seek to prove that each equilibrium x^* can be supported.

A degenerate version of our previous proofs applies in this case. Consider the following strategy profile: At the proposal stage (Stage 1), both players offer the "null" contract with zero transfers in all cells of G both for the agreement clause and for the promise clause. At the signing stage (Stage 2), no contract is signed by either player (so that the null promises are enacted). At the action stage (Stage 3), if the contract proposals and the signing behavior were as above, x^* is played in the resulting game G. If only one of the players deviated at any of the previous stages, the "worst" equilibrium (which is here of course no worse than any other equilibrium) for this player is played in the resulting game. Otherwise any equilibrium is played (again, this choice is irrelevant).

Let us check that the strategy profile forms an efficient SPNE of $\Gamma^E(G)$: There is no profitable deviation at Stage 2 for any of the players, as the opponent is expected not to sign any contracts. Similarly, at Stage 1, a deviation of Player i to another agreement clause does not change the outcome (as players are expected not to sign any contracts at Stage 2). The only remaining deviation is to another promise clause. However, any promise of a positive payment from Player i could only increase the (ultimate) payoff of Player j, as Player j is already assured the maximin payoff v_j. Since (v_1+v_2) is efficient, this increase would necessarily be at the expense of Player i.

References

Aivazian, V.A. and Callen, J.L. (1981), "The Coase Theorem and the Empty Core", *Journal of Law and Economics* 24, 175-181.

d'Aspremont, C.A., Jacquemin, J., Gabszewicz, and Weymark, J. (1983), "On the Stability of Price Leadership," *Canadian Journal of Economics* 16, 17-25.

Aumann R.J. (1959), "Acceptable Points in General Cooperative n-person Games", in R. D. Luce and A. W. Tucker (eds.), Contributions to the Theory of Games IV, *Annals of Mathematics Study* 40, 287-324, Princeton, NJ: Princeton University Press.

Aumann, R.J. and Myerson, R.M. (1988), "Endogenous Formation of Links between Players and of Coalitions: An Application of the Shapley Value", Chapter 12 in A.E. Roth (ed.), *The Shapley Value: Essays in Honor of Lloyd Shapley,* Cambridge: Cambridge University Press.

Barrett, S. (1994), "Self-Enforcing International Environmental Agreements", *Oxford Economic Papers* 46, Special Issue on Environmental Economics (Oct., 1994), 878-894.

Benoît J.-P. and V. Krishna (1985), "Finitely Repeated Games", *Econometrica* 53, 890-904.

Bernheim, B.D. and Whinston, M.D. (1986), "Common Agency", *Econometrica* 54, 923-942.

Bernheim, B.D. and Ray, D. (1989), "Collective Dynamic Consistency in Repeated Games", *Games and Economic Behavior* 1, 295-326.

Bloch, F. and Gomes, A. (2006), "Contracting with Externalities and Outside Options", *Journal of Economic Theory* 127, 172-201.

Chatterjee, K., Dutta, B., Ray, D., and Sengupta, K. (1993), "A Noncooperative Theory of Coalitional Bargaining", *Review of Economic Studies* 60, 463-477.

Coase, R.H. (1960), "The Problem of Social Cost", *Journal of Law and Economics* 3, 1-44.

Coase, R.H. (1974), "The Lighthouse in Economics", *Journal of Law and Economics* 17, 357-376.

Coase, R.H. (1981), "The Coase Theorem and the Empty Core: A Comment", *Journal of Law and Economics* 24, 183-187.

Carraro, C. and Siniscalco, D. (1993), “Strategies for the International Protection of the Environment”, *Journal of Public Economics* 52, 309-328.

Dixit A.K. and Olson, M. (2000), “Does Voluntary Participation Undermine the Coase Theorem?” *Journal of Public Economics* 76, 309-335.

Danziger, L. and Schnytzer, A. (1991), “Implementing the Lindahl Voluntary-Exchange System”, *European Journal of Political Economy* 7, 55-64.

Ellingsen, T. and Miettinen, T. (2008), “Commitment and Conflict in Bilateral Bargaining”, *American Economic Review* 98, 1629-1635.

Ellingsen, T. and Miettinen, T. (2014), “Tough Negotiations: Bilateral Bargaining with Durable Commitments”, *Games and Economic Behavior* 87, 353-366.

Ellingsen, T. and Paltseva, E. (2012). “The Private Provision of Excludable Public Goods: An Inefficiency Result”, *Journal of Public Economics* 96, 658-669.

Fudenberg D. and E. Maskin (1986), “The Folk Theorem for Repeated Games with Discounting and with Incomplete Information”, *Econometrica* 54, 533-554.

Gomes, A. and Jehiel, P. (2005), “Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies”, *Journal of Political Economy* 113, 626-667.

Grossman, S.J. and Hart, O.D. (1986), “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration”, *Journal of Political Economy* 94, 691-719.

Gordley, J. (1995), “Enforcing Promises”, *California Law Review* 83, 547-614.

Guttman, J.M. (1978), “Understanding Collective Action: Matching Behavior”, *American Economic Review* 68, 251-255.

Guttman, J.M. (1987), “A Non-Cournot Model of Voluntary Collective Action”, *Economica* 54, 1-19.

Guttman, J.M. and Schnytzer, A. (1992), “A Solution to the Externality Problem Using Strategic Matching”, *Social Choice and Welfare* 9, 73-88.

Hart, O.D. (2009), “Hold-up, Asset Ownership, and Reference Points”, *Quarterly Journal of Economics* 124, 267-300.

Hart, O.D. and Moore, J. (1990), “Property Rights and the Nature of the Firm”, *Journal of Political Economy* 98, 1119-1158.

Hart, O.D. and Moore, J. (2008), "Contracts as Reference Points", *Quarterly Journal of Economics* 123, 1-48.

Holmström, B, (1982), "Moral Hazard in Teams", *Bell Journal of Economics* 13, 324-340.

Hyndman, K. and Ray, D. (2007), "Coalition Formation with Binding Agreements", *Review of Economic Studies* 74, 1125–1147.

Jackson, M.O. and Wilkie, S. (2005), "Endogenous Games and Mechanisms: Side Payments among Players", *Review of Economic Studies* 72, 543-566.

Kalai, A.T., Kalai, E., Lehrer, E. and Samet, D. (2010), "A Commitment Folk Theorem", *Games and Economic Behavior* 69, 127-137.

Kamien, M.I. and Zang, I. (1990), "The Limits of Monopolization through Acquisition", *Quarterly Journal of Economics* 2, 465-499.

Katz, M.L. (2006), "Observable Contracts as Commitments: Interdependent Contracts and Moral Hazard", *Journal of Economics and Management Strategy* 15, 685-706.

Konishi, H. and Ray, D. (2003), "Coalition Formation as a Dynamic Process", *Journal of Economic Theory* 110, 1-41.

Libecap, G.D. (2014), "Addressing Global Environmental Externalities: Transaction Costs Considerations", *Journal of Economic Literature* 52, 424-479.

Mailath, G. and Postlewaite, A. (1990), "Asymmetric Information Bargaining Problems with Many Agents", *Review of Economic Studies* 57, 351-367.

Moore, J. (1992), "Implementation, Contracts, and Renegotiation in Environments with Complete Information", Chapter 5 of J.-J. Laffont (ed.), *Advances in Economic Theory, Sixth World Congress,* Volume I, Cambridge: Cambridge University Press.

Myerson, R.B. (1977), "Graphs and Cooperation in Games", *Mathematics of Operations Research* 2, 225-229.

Myerson, R.B. (1991), *Game Theory: Analysis of Conflict,* Cambridge MA: Harvard University Press.

Peters, M. and Szentes, B. (2012), "Definable and Contractible Contracts", *Econometrica* 80, 363-411.

Ray, D. and Vohra, R. (1997), "Equilibrium Binding Agreements", *Journal of Economic Theory* 73, 30-78.

SEGAL, I. (1999), "Contracting with Externalities", *Quarterly Journal of Economics* 114, 337-388.

SEGAL, I. (2003), "Coordination and Discrimination in Contracting with Externalities", *Journal of Economic Theory* 113, 147-181.

SEIDMANN, D.J. AND WINTER, E. (1998), "Exploring Gains from Trade in Multilateral Bargaining: A Theory of Gradual Coalition Formation", *Review of Economic Studies* 65, 793-815.

VARIAN, H. (1994a), "Sequential Provision of Public Goods", *Journal of Public Economics* 53, 165-186.

VARIAN, H. (1994b), "A Solution to the Problem of Externalities when Agents are Well-Informed", *American Economic Review* 84, 1278-1293.

WATSON, J. (2011), "Contractual Chains: Decentralized Contracting with Externalities", Manuscript, University of California San Diego.

YAMADA, A. (2003), "Efficient Equilibrium Side Contracts", *Economics Bulletin* 3(6), 1-7.

YAMASHITA, T. (2010), "Mechanism Games with Multiple Principals and Three or More Agents", *Econometrica* 78, 791-801.