Private Contracts and Social Inefficiency: Confining the Coase Theorem*

Tore Ellingsen[†]and Elena Paltseva[‡]

October 2014
Preliminary and Incomplete

Abstract

If people are free to contract, will outcomes be efficient? We study the 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. With mandatory participation, efficiency is attainable also in many-player situations. But if participation is voluntary, and there are more than two players, there is a large class of situations in which all consistent equilibria are inefficient due to free-riding. In these cases, efficient contracting would require a different initial distribution of property rights.

JEL Codes: C78, D23

Keywords: Endogenous mechanisms, Contracting, Renegotiation

*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, Eirik Gaard Kristiansen, Margaret Meyer, Paul Milgrom, Joel Watson, Peyton Young, and especially to Matthew Jackson 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 people 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 question the argument. 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 that game possess equilibria in which players arrive at efficient agreements?*

To motivate our choice of model, let us begin by explaining what we mean by a contract and a contract negotiation. According to the Academic Edition of 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 another accepting. If one of the parties fails to keep the promise, the other is entitled to legal recourse.*

Accordingly, we 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: (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.

More formally, a contract specifies monetary transfers as a function of actions taken in some *situation*, or action game, G. Contract negotiations are viewed as multi-stage process.

[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.

At Stage 0, players learn the situation G and decide whether to take part in negotiations or not. At Stage 1, players propose contracts. At Stage 2 players decide which contracts to sign. A contract becomes valid if and only if it is signed by all players who may pay or receive transfers under the contract.[2] At Stage 3, upon seeing the outcome of the negotiations, players play G modified by the agreed transfers. The whole interaction is thus a four-stage game, call it $\Gamma^V(G)$. If participation in negotiations is mandatory, as is implicitly assumed in parts of the 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 simple to sustain cooperation. But upon closer inspection, the issue is more intricate.

Our main results can be summarized as follows: (i) Under mandatory participation ($\Gamma(G)$), the model 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. Under this solution concept, $\Gamma(G)$ thus always admits a weak form of the Coase theorem. (ii) However, if we additionally insist that equilibria should be consistent (Bernheim and Ray, 1989), and there are no more than two players, only efficient outcomes are selected. That is, with two players, consistency implies a strong form of the Coase theorem for $\Gamma(G)$. (iii) Under voluntary participation ($\Gamma^V(G)$), on the other hand, there is a large and relevant class of situations G involving more than two players for which all consistent equilibrium outcomes of $\Gamma^V(G)$ are inefficient. That is, under voluntary participation, the Coase theorem does not hold in any form for this class of situations. The reason is that it is tempting to free-ride on others' agreements, since these others cannot credibly threaten to behave inefficiently upon such strategic non-participation.

Finally, our model invites a simple definition of property rights. A property right is the right to to put a price on an action. For actions taken by oneself, a property right is the right to take the action for free. For actions taken by another party, a property right is the right to insist that the action is not taken, except against compensation determined by the owner.[3] Our definition encompasses property rights over assets, but is much broader. Our above results describe outcomes for the case in which everyone initially owns all their personal actions, so property rights are widely dispersed. The more general version of the model has an additional stage between the signing stage and the action stage, at which action owners unilaterally determine prices for any actions that are not covered by a contract.

In all situations G we find that there is always some allocation of property rights that

[2]We shall also consider the case in which some players, at Stage 2.5, have the right unilaterally to put prices on actions that are not covered by a contract. See below.

[3]Insistence that the action is not taken effectively assumes a well-functioning legal system; there must be a deterrent threat of punishment in case the action is taken nonetheless.

guarantee efficient outcomes. Indeed, if one player owns all action rights – except for a single default action for each player – the outcome will be efficient. Since efficiency is not generally attainable under dispersed property rights, this result contradicts Coase's assertion that, in the absence of transaction costs, the allocation of property rights has only distributional consequences.

The paper is organized as follows. Section 2 describes our contribution's relationship to the existing literature. Section

2 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).[4] 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 be attained if people are confined to resolve their conflicts through appropriately designed mechanisms; see Moore (1992) for a survey.

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 similar to Myerson's in adopting the view that contract-signing is a coordination game in which each player becomes 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". 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; we do not presume that some player or players, the principals, are granted exclusive proposal rights. 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.

With respect to the modeling approach, our analysis is perhaps 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

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

shape of transfers. That is, players are not confined to making 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. We return to this issue in Section 7.

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 literature about "contracting on contracts"; see, e.g., Katz (2006), Kalai et al. (2010), Yamashita (2010), and Peters and Szentes (2012). There, 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).[5] 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.[6]

Our work is also related to the principal-agent literature, most closely with the work on contracting with externalities, and especially Segal (1999). There, 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 bad Nash equilibrium, and with a particular asymmetric payoff structure; one player, the principal, is not taking any action, but is heavily 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 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). This outcome corresponds to one of the extreme points in the set of equilibria described in our Theorem 1. Since

[5] 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.

[6] However, in Appendix C we discuss an extension of the model in which some form of intertwining seems necessary in order to sustain efficiency.

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.

Turning to the issue of participation, at least several previous strands of literature have investigated the possible inefficiencies that may arise in non-cooperative coalition formation. Since the formation of a coalition is typically interpreted as participation in contract negotiations, it is natural to consider how our fully non-cooperative approach relates to these partly cooperative approaches.

An early strand of cooperative analysis had argued that many cooperative games have an empty core, and thus 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 is inefficient will the Coase theorem be overturned.

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).[7] 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

[7] 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.

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 due to, among others, Chatterjee et al (1993), Ray and Vohra (1997), and Seidmann and Winter (1998).[8] In these "hybrid" 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. (The models are referred to as hybrids, because they comprise both non-cooperative and cooperative elements.) That is, this literature takes for granted 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, in our non-cooperative model 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. Thus, although some insights about the role of externalities are related, our non-cooperative analysis does not offer a general justification for the hybrid approach.

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 have chosen to focus on consistent equilibrium, partly because it has the virtue that the solution set is never empty. 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

[8] 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).

cooperation now (e.g., the proposal stage).[9] 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.[10] However, others 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 when all actions are 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.

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 the lighthouse services to rest with a single party, such as the provider of lighthouse services, rather than with all the potential users – among whom the free-rider problem might have been severe.[11]

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

[10] 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 effectively 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.

[11] It is perhaps tempting to think that outcomes would again turn efficient if such users were allowed to trade their property rights. However, there would instead be a free-rider problem in the market for property rights, as recognized in the analogous context of take-overs by Grossman and Hart (1980). For a relevant and striking real-world application, see the analysis of oil-field unitization by Libecap and Wiggins (1984).

3 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 Example 1, reproduced in Figure 1.

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

Figure 1: A Prisoners' Dilemma

Now, suppose the ranchers are at liberty to propose and accept contracts of the form $(t^{HH}, t^{HL}, t^{LH}, t^{LL})$, where t^{ij} denotes the transfer from rancher 1 to rancher 2 in case they end up playing the action profile (i, j). More precisely, let there be a first stage in which both simultaneously propose contracts and a second stage at which each is at liberty to sign any one of the proposals, or none.[12] 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

[12]Technically, we could allow each rancher to sign multiple contracts, but that raises a number of issues that are only distracting at this point. We will address them below.

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.[13] Section 3 shows how this insight generalizes.

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 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 not renegotiation-proof, the very argument that supported efficiency in the case of two players undermines it in the case of more players.[14]

4 Definitions

Let G be a finite n-player game, called "the action game."[15] The set of players is denoted $N = \{1, ..., n\}$. 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. 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.

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

[13] 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.

[14] Another problem emphasized by Jackson and Wilkie arises if players can make unilateral 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 5 explores this issue.

[15] It would be desirable to also consider games with a continuum of strategies, but due to technical complications we refrain from doing so here.

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 $\sum_{i=1}^{n} t_i(x) = 0$.[16] 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

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

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

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

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 yields 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 $U_i(x)$.

[16] 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.

4.1 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. Then, each player individually chooses at most one contract proposal to sign.[17] 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. They 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 $\Gamma(G)$.

5 Outcomes under mandatory participation

Our first aim is to establish whether there is a subgame-perfect equilibrium of $\Gamma(G)$ supporting the play of an efficient action profile. However, one might also wonder which other action profiles might be supported. We therefore begin by proving the more general result 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 other 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.*

[17] 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 shown by Ellingsen and Miettinen (2008, 2014), outcomes tend to be highly inefficient if players become committed to not accepting outcomes that are worse than their own proposal.

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.[18]

The proof proceeds as follows: (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.

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 $\tau_i = \hat{t}$.[19]

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 $\hat{t}$, players' signing decisions are done in a way that results in a Pareto-undominated outcome. If there are several such outcomes, 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

[18]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.

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

each signing-stage subgame is finite, backward induction ensures existence of such a signing profile.[20]

Stage 3. (i) If some proposal $\tau_i = \hat{t}$ was signed, play $\hat{x}$. (ii) if $\tau_1 = \hat{t}$ is signed by all but Player k, play Player k'th worst Nash equilibrium in the resulting subgame $\widetilde{G} = G$.[21] (iii) If $\tau_1 \neq \tau_2 = \hat{t}$ and τ_2 is signed by all but Player k, play Player k'th worst Nash equilibrium in the resulting subgame $\widetilde{G} = G$. (iv) 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 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 rule above prescribes 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 any other off-equilibrium branch, each player is content not to sign any contract proposal; as opponents are expected not to sign, $t^S = \varnothing$ regardless of the own signing decision.

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 $\Gamma(G)$.*

[20] This latter part of the contract allows us to focus on Pareto-dominant outcomes off-equilibrium path. While this feature it of no importance now, it will facilitate our subsequent analysis of consistent equilibria.

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

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.

5.1 Extensions

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 $\Gamma^+(G)$.[22]

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

[22]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).

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.

5.2 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 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). The proof proceeds 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.

	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 $\Gamma(G)$

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 straightforward 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. In fact, we can formulate 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.*

Intuitively, no player would ever sign a contract that yields her less than $v_i(G)$ on the equilibrium path. Thus, in these situations, any contractually sustainable strategy profile must yield at least $u_i^{ud}(G)$ to each player i. But if so, there can be no inefficient consistent equilibrium; there would always be a profitable contract deviation at the proposal stage that would be signed by all the parties, by exactly same logic as in Theorem 2.

6 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.[23] 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.[24]

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.*

Since the proof is almost identical to the proof of Theorem 2, we omit it.[25]

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

[23] 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.

[24] 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); see the proof of Theorem 1.

[25] 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.

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 is 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} player i's set of best-responses in G. 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.

In these games there is a tension between efficiency and equilibrium (condition (i)), and the worst equilibrium payoff is as bad 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). Second, in this class of games, any consistent equilibrium in which only players $j \neq i$ can contract implies an allocation 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 $j \neq i$, which contradicts the definiton 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 is multiplied by 1.6 and the resulting sum is divided equally among all four players. That is, player i's payoff is

$$1.6\frac{\sum_j c_j}{4} + M - c_i.$$

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 cannot be supported as a consistent equilibrium. To see this, note that (i) $\Gamma^V(G)$ is covered by Theorem 6 (each player's unique dominant strategy is to contribute nothing, and the unique equilibrium payoff profile of G is (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

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

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 is $6.4M$. Hence, full cooperation is not sustainable in a consistent equilibrium of the contracting game.

7 Promises as outside options

This section is not edited*******

Jackson and Wilkie (2005) conjectured that agreements might admit efficient equilibrium outcomes, but they also noted that the presence of unilateral promises could potentially threaten that efficiency. A promise by Player i is a 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. Recall that Jackson and Wilkie (2005) studied the two-stage game, denote it $\Gamma^{JW}(G)$, in which all players first simultaneously make promises T_i^{JW} and then play G modified by these promises. Among other things, they showed that $\Gamma^{JW}(G)$ does not always admit efficient equilibrium outcomes.

In this Section, we shall show that efficiency is often attainable despite the availability of unilateral 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.

Before providing a formal analysis of promises, let us clarify the connection to promises that are observed in practice. When a buyer unilaterally promises to pay an amount T 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. These promises represent a plausible extension of the agreements considered above. However, Jackson and Wilkie (2005) also admit more exotic promises. As an illustration, consider the game in Figure 1 again. The analogy to the procurement example is 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. 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".[26] It is these strategic donative promises that occupy center stage in Jackson and Wilkie (2005) and that play a crucial role below.

7.1 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 stands for the agreement clause of the proposed contract, and $\mathcal{T}_i$ for 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

[26]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.

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

7.2 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 7 *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 7 rests on the following Lemma.

[27] 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.

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 7 ensures the existence of efficient equilibria in the game $\Gamma^R(G)$. However, while Theorem 7 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. The next subsection identifies additional features of the contracting technology that are sufficient for sustaining efficient outcomes in all finite normal-form games.

7.3 Conditioning promises on strategies

Let us now allow contracts to 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.[28] Furthermore, allow transfers to be non-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 $j \in N^t$.[29] 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 8 *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 $\Gamma^{JW,E}(G)$.

[28] 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 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.)

[29] JW mention randomization over the set of transfers, but for technical reasons do not explicitly allow it (JW, page 549).

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 $v_i(G) + d$.*

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

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

and the unilateral promise clause is

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

That is, as in Theorem 7, 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 7. 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 9 *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 7. For the case of more than two players, the proof relies on the same construction as the proof of Theorem 7, 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 9 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.[30]

8 Conclusion

We have proposed and analyzed a non-cooperative model of contract negotiations. 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), only the first intuition applies. Efficiency is attainable whether participation is voluntary or not.

With more than two players, it can be individually rational not to participate in negotiations. Even in the best attainable equilibrium of the contract negotiation game, some player will be catching a free ride on others' agreements. These agreements in turn will fail to implement efficient outcomes. That is, with more than two players neither participation nor efficiency is ensured.

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

Appendices

A Proof of Theorem 1

As mentioned in the main text, 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 case of sustaining a pure strategy profile.

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 $t_j\left(x_j, \hat{x}_{-j}\right)$ on the support of $\hat{\mu}$ in the following way:

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

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

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, mark each of the equations entering subsystem (2) by the strategy whose payoff it represents. For example, we would refer to equation

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

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,

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

There are $\Pi_{j=1}^N n_j$ such equations. As above, mark 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 $\hat{\mu}$

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

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

Lemma 3 *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 $N \cdot \prod_j n_j$ unknown transfer values $t_j\left(\hat{x}_j, \hat{x}_{-j}\right)$. 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:

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

Thereby, let's drop equation $(2.\hat{x}_1^1)$ - the one 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, $t_1\left(\hat{x}\right)$. This transfers enter 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)

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

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 $(2.\hat{x}_1^k)$ will transform $(2.\hat{x}_1^k)$ into the following form

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

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

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

and denote this transformed equation by $(2.\hat{x}_1^k)'$. 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). At the same time, in this new system the transfers $t_1(\hat{x})$ of player 1 enter only the subsystem (3), one per each equation. Thereby, the equations in this subsystem are linearly independent both across themselves and with the rest of the system. In other words, 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 look closely at 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 across themselves and with the remaining subsystem, and we can again continue with the analysis of the remaining part of the system only.

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

In other words, 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

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

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

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

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. ■

B Proof of Theorem 6

Lemma 4 *"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.

If the above does not hold, then 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 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's show that this supremum can be achieved at some strategy profile that also belongs to $D_i(\mu_0)$, or, in other words, that the set of strategy profiles that maximize the joint payoff of players $j \neq i$ is non-empty.

By 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 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). Thereby, it contains a converging subsequence $\{\mu_{n_k}\}$, with the limit that we denote by $\hat{\mu}$. Recall that the non-participating player i is always playing best response at ΔX^{BR_i}. As the best response correspondence in G - a finite game with continuous payoffs - is upper hemi-continous, player i's limit strategy $\hat{\mu}_i$ belongs to her best response to $\hat{\mu}_{-i}$ in G. That is, $\hat{\mu}$ also belongs to ΔX^{BR_i}. Further, by continuity of 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 non a singleton, repeat the above argument to select a subset of it that maximizes the payoff of player i. ■

Lemma 5 *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 ensures allocation $\widetilde{\mu}^{BR_i}$ 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' signing decisions are done in a way that results in a Pareto-undominated outcome. If there are several such outcomes, 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'th worst Pareto-undominated Nash equilibrium in the resulting game. (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-efficient 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. This means, that 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 equilibria, making unilateral deviation non-profitable. As player i plays best responses, $\widetilde{\mu}^{BR_i}$can be supported as a Nash equilibrium. Also, as $\widetilde{\mu}^{BR_i}$yields the highest *joint* payoff to players $j \neq i$ on ΔX^{BR_i}, it would be impossible to increase the payoff of any of the players without lowering the payoff of at least one of the others and still being in ΔX^{BR_i} strategy set. That is, $\widetilde{\mu}^{BR_i}$ 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 Pareto-efficient continuations at each proper 3-stage subgame. Finally, along any other off-equilibrium branch, by definition the signing decisions are in line with Pareto-efficient equilibria that have Pareto-efficient continuation at each proper 3rd stage 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. In turn, there could be no another, Pareto-improving equilibrium in this subgame of $\Gamma^V(G)$, as argued above.

Finally, it is sufficient to notice that there could not be a worse punishment in a consistent equilibrium. Indeed, recall, that all "punishment" equilibria should belong to ΔX^{BR_i}. Then, a consistent "punishment" equilibrium resulting in Player i's payoff below

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

implies that players $j \neq i$ should jointly get more than at $\widetilde{\mu}^{BR_i}$, which contradicts the definition of $f_i(G)$. ■

Now we are ready to prove Theorem 6.

Proof. (a) Assume that condition

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

is met, and let's demonstrate existence of an efficient renegotiation-proof equilibrium of $\Gamma^V(G)$. Consider the following 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}; however, the specification of this contract will depend on the set of players who decides to participate. Specifically, if the set of players who decide to participate is universal, then all players propose a contract t that supports allocation x^* and provides players $j \in N'$ with the following transfers at x^*

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

If there exists a single non-participating player i , then players $j \neq i$ choose the above-mentioned contract that implements allocation $\widetilde{\mu}^{BR_i}$. If the non-participating set of players includes two or more players, all participating players N^P propose a consistent contract that yields the allocation μ^{N_p} with the highest joint payoff to the members of N^P among all the allocations in which players N/N^P play best responses (which exists and can be supported by the arguments very similar to the above lemmas). Finally, if only one player chooses to participate, no contract is proposed.

Contract signing stage: Number 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 of N^P-player contract proposal deviation, players sign in a way that makes their signing decisions Pareto-efficient equilibria which have Pareto-efficient continuation at each proper subgame.

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 allocation suggested by this contract. If $\tau_{j_i} = t_{N^P}$ is not signed by a single Player $j_k \in N^P$, play Player j_k'th 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.). If no more than one person has chosen to participate, play this player worst undominated Nash equilibrium of G.

Let us show that this is indeed an efficient renegotiation-proof equilibrium (in particular, that it has Pareto-efficient 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. Assume now that $N^P \geq 2$ players choose to participate. If $t^S = t_{N^P}$, the resulting allocation represents an undominated Nash equilibrium of the game. Indeed, if N^P includes N', the resulting allocation is x^*, and the allocation rule (5) ensures that no unilateral deviation is profitable. As x^* is efficient, it cannot be renegotiated either. If $N'/N^P \neq \varnothing$, then the contract implements an allocation on N^P-Pareto frontier, and constitutes an non-renegotiable equilibrium of resulting game.

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 in line with Pareto-efficient equilibria that have Pareto-efficient continuation 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. ■

Appendix B: Proof ofÉ

C 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 $\hat{x}_j$.

C.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 6 *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 7 *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 8 *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 6 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 7 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

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

C.2 Proof of Theorem 8, 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

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

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

[3] 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.

[4] 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.

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

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

[7] Bernheim, B.D. and Whinston, M.D. (1986): Common Agency, *Econometrica* 54, 923-942.

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

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

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

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

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

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

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

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

[16] Ellingsen, T. and Miettinen, T. (2008): Commitment and Conflict in Bilateral Bargaining, *American Economic Review* 98, 1629-1635.

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

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

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

[20] 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.

[21] Grossman, S.J. and Hart, O.D. (1980): Takeover Bids, The Free-Rider Problem, and the Theory of the Corporation, *Bell Journal of Economics* 11, 42-64.

[22] 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.

[23] Gordley, J. (1995): Enforcing Promises, *California Law Review* 83, 547-614.

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

[25] Guttman, J.M. (1987): A Non-Cournot Model of Voluntary Collective Action, *Economica* 54, 1-19.

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

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

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

[29] Jackson, M.O. and Wilkie, S. (2013): Erratum to "Endogenous Games and Mechanisms: Side Payments among Players," *Review of Economic Studies* 80, 840-843.

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

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

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

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

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

[35] 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.

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

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

[38] Peters, M. and Szentes, B. (2012): Definable and Contractible Contracts, *Econometrica* 80, 363-411.

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

[40] Segal, I. (1999): Contracting with Externalities, *Quarterly Journal of Economics* 114, 337-388.

[41] 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.

[42] Varian, H. (1994a): Sequential Provision of Public Goods, *Journal of Public Economics* 53, 165-186.

[43] Varian, H. (1994b): A Solution to the Problem of Externalities when Agents are Well-Informed, *American Economic Review* 84, 1278-1293.

[44] Yamada, A. (2003): Efficient Equilibrium Side Contracts, *Economics Bulletin* 3(6), 1-7.

[45] Yamashita, T. (2010): Mechanism Games with Multiple Principals and Three or More Agents, *Econometrica* 78, 791-801.