Coalitions in Repeated Games*

S. Nageeb Ali† Ce Liu‡

January 5, 2025

Abstract

This paper proposes a framework and solution concept for repeated coalitional behavior. We model history-dependent schemes in which coalitions refrain from blocking today given that their behavior affects future play. We evaluate the effectiveness of these schemes across a wide range of settings and apply our results to repeated matching and negotiations.

*We thank Dan Barron, Federico Echenique, Matt Elliott, Drew Fudenberg, Ben Golub, Yingni Guo, John Hatfield, Scott Kominers, Maciej Kotowski, Elliot Lipnowski, Francesco Nava, Alex Wolitzky, Nathan Yoder, and various conference and seminar audiences for helpful comments.

†Department of Economics, Pennsylvania State University. Email: nageeb@psu.edu.

‡Department of Economics, Michigan State University. Email: celiu@msu.edu.

1 Introduction

The literature on repeated games studies how history-dependent schemes support cooperation. Each player adheres to the scheme because she expects rewards from doing so and punishment should she deviate. This canonical approach models non-cooperative play in which actions and deviations are chosen only by individuals.

But, in many contexts, analysts have found it more useful to allow groups of players to act jointly. One then looks for schemes that are immune to group deviations. For instance, matching and network theory allow for pairwise deviations and study stable arrangements in which no pair can profitably deviate. Similarly, political economy models emphasize Condorcet winners in which decisive coalitions of players jointly select policy. Broadly, cooperative game theory studies stable solutions, exemplified by notions of the core, in which no group of players finds it profitable to deviate.

A natural question is whether and how one can marry these two powerful approaches to modeling group behavior. Doing so would elucidate the scope for carrots and sticks to create long-run stable coalitional behavior; such analyses might be helpful for the study of repeated matching, network formation, and voting. It could also help determine whether, and under what conditions, equilibria of repeated non-cooperative games are resistant to coalitional deviations. Our goal is to provide a simple and tractable framework that combines elements of each approach.

We illustrate our framework using the *Roommates Problem*. Ann, Bella, and Carol decide who will room together. The hitch is that only two people can share a room, leaving at least one person out. Each person prefers to have a roommate, and each has a favorite; Table 1 depicts their payoffs.

	Ann	Bella	Carol
Ann	1	3	2
Bella	2	1	3
Carol	3	2	1

Table 1. Payoffs of Row Player from matching with Column Player (or remaining unmatched).

No arrangement is stable: in each case, at least one pair of players would profitably deviate. For example, if the proposed arrangement pairs Ann and Bella as roommates leaving Carol out, Bella and Carol would each be better off if they defected from the arrangement and roomed together instead.

What if this problem is repeated? Suppose the trio make choices monthly, accruing

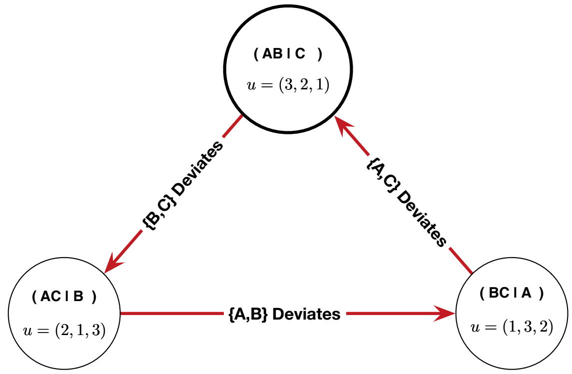


Figure 1. A perfect coalitional equilibrium for the roommates' problem if $\delta \geq 1/2$.

the flow payoffs above, each weighted by the discount factor δ. In each period, a pair of players can jointly defect. And, just like repeated games, no one can commit to her future behavior on- or off-path. Can carrots and sticks support a match?

Figure 1 depicts such a scheme. On the path of play, Ann and Bella room together each year, leaving Carol out. While Bella and Carol would myopically prefer to share a room, the scheme deters this deviation if δ exceeds 1/2. The reason is that Bella anticipates that after the deviation, starting from next month, Ann and Carol would room together and she would then be left out. Her short-term gain from deviating does not offset her long-term loss, since $(1-\delta)3+\delta(1) \leq 2$. Moreover, the punishment is itself credible because the prescription following every history is self-enforcing.

We model these schemes in general games. We study the repeated play of an abstract stage game in which individuals and coalitions can take actions, accommodating both normal-form and characteristic function games. In line with repeated games, we examine history-dependent schemes where no coalition profits from blocking at any history given how it affects continuation play. We dub such schemes *perfect coalitional equilibria* (PCE). In the context of repeated normal-form games, PCE refines subgame perfect equilibria, offering immunity to coalitional deviations. In repeated coalitional settings, PCE unsheathes the power of carrots and sticks, highlighting how dynamic incentives can deter coalitions from blocking.

Our main results identify the limits of PCE. We offer conditions under which history dependence thwarts coalitional deviations so that the set of PCE-supportable payoffs is large. Conversely, we also identify settings in which coalitional deviations choke the possibility for cooperation, resulting in "anti-folk" theorems. Underlying our results is a simple guiding principle: a coalition can withstand punishments if its members have aligned interests. In such cases, all members of the coalition enjoy a high minmax, considerably above their individual minmax, which constrains the power of carrots and sticks. However, if there is any wedge in members' incentives, player-specific punishments can splinter coalitions. Then, the set of PCE-supportable payoffs virtually coincide with that of subgame perfect equilibrium payoffs (if players are patient).

Building on this principle, we explore various features that can align coalitions' interests and identify how these features constrain the ability of PCE to enforce cooperation. One such feature is the use of *strongly symmetric* schemes, where players behave symmetrically after every history. These schemes feature often in the study of repeated games, such as grim-trigger punishments used to support cooperation in a repeated prisoner's dilemma or to sustain collusion among oligopolists. Although these schemes are subgame perfect equilibria, we show that they typically are not PCE. The reason is that these schemes effectively align players' incentives, which makes their punishments non-credible. By contrast, schemes that feature asymmetric play off-path can credibly deter blocking coalitions and support a larger payoff set.

Do transfers align incentives? One might expect the answer to be yes: if a coalition achieves a net gain in utility by blocking, it can distribute those gains among its members to ensure that each benefits. However, we show that if all transfers are publicly observed, a PCE can splinter coalitions by conditioning on who pays whom. On the other hand, if a coalition can make transfers "secretly"—that is, without the transfers being publicly observed—incentives within that coalition then can be entirely aligned. Such a coalition effectively functions as a unitary agent and secures a high payoff across all PCE. Therefore, secret side-payments limit what a PCE can enforce.

In addition to characterizing these limits, we show that PCEs are highly tractable. Identifying the set of PCE payoffs for a fixed δ can actually be simpler than standard self-generation methods: all PCE-supportable payoffs can be supported by stationary PCE whenever the stage game exhibits *default-independent power*. This property holds in the characteristic function games studied in cooperative game theory, matching models without externalities, and political economy models of voting.

Building on Kelso and Crawford (1982), we apply these results to repeated labor-market matching, in particular, to speak to the debate on wage transparency. In light of this debate, we vary whether past wage terms are publicly or privately observed. We find that a vast range of outcomes can be supported by public wages but private wages lead to a complete collapse of intertemporal incentives: the supportable payoff set reduces then to the core of the stage game. Who then benefits from wage transparency? Workers benefit if they are plentiful or their marginal returns fall quickly. Absent transparency, competition bids down their wages; by contrast, wage transparency enables them to use collective-bargaining schemes that enforce high wages. If workers are scarce or their marginal returns fall slowly, then firms profit from wages transparency because they can then collusively suppress wages.

We also study repeated multilateral negotiations in settings in which some players have veto power. The stage-game core here involves the veto players capturing the entire surplus. Against that backdrop, we show that history dependence can promote egalitarianism. Moreover, using our stationarity result described above, we characterize the set of supportable payoffs for fixed discount factors, which typically includes equal splits. However, these egalitarian schemes collapse if veto players can make secret side-payments to other coalition members; if every minimal winning coalition can make secret side-payments, veto players return to being de facto dictators.

We briefly discuss related work. Bernheim and Slavov (2009) propose the notion of a Dynamic Condorcet Winner for an infinitely repeated voting game. Our solution concept coincides with theirs when applied to majority-rule voting. While they describe some properties of their concept, they do not characterize its limits. Consequently, many issues central to our study—such as alignment or the role of transfers—do not feature in their work. Moreover, our primary interest is to go beyond voting games so as to describe the power of carrots and sticks broadly in coalitional settings.

Given the role of preference alignment, our work uses approaches developed by Fudenberg and Maskin (1986), Abreu, Dutta, and Smith (1994), and Wen (1994); in particular, our notion of alignment in non-transferable utility settings draws on Abreu, Dutta, and Smith's *non-equivalent utilities* condition. One of our results highlights how secret side-payments can align coalitional incentives and thereby undermine punishments. That secret side-payments can disrupt dynamic incentives features also in Barron and Guo (2021), who consider a relational contracting game between a long-run principal and a sequence of short-run agents.

Our work contributes to the discussion of wage transparency, a theme of recent interest; see Cullen (2024) for a survey. Cullen and Pakzad-Hurson (2023) model equilibrium effects of wage transparency in a bargaining model with incomplete information and show that it can disadvantage workers. We offer a complementary perspective, highlighting how the observability of past offers can help workers collectively bargain for higher wages. Beyond this application, we view incorporating long-run incentives in Kelso and Crawford (1982)'s workhorse model to be of independent interest. In the static setting, this framework has been enriched in various directions (e.g., Hatfield and Milgrom 2005; Hatfield, Kominers, Nichifor, Ostrovsky, and Westkamp 2013) but relatively little is known about how carrots and sticks affect these matching markets.

Several papers explore coalitional deviations in repeated games. Aumann (1959) and Rubinstein (1980) study Strong Nash and Strong Perfect Equilibria of infinitely repeated games using other payoff criteria. These solution concepts assume that players cannot commit to long-term plans on the equilibrium path but can do so when deviating. DeMarzo (1992) focuses on finite-horizon games and proposes an inductive solution concept that corresponds to a Strong Nash equilibrium of the reduced normal-form game.[1] A different strand of the literature models questions of renegotiation—see, for instance, Bernheim and Ray (1989), Farrell and Maskin (1989), Miller and Watson (2013), and Safronov and Strulovici (2018)—in which players can rewire their expectations of future play. Our work studies the complementary question of when coalitions refrain from profitably blocking *given* their rational expectations of future play.

This motivation connects our work to that on coalition formation, which studies related but distinct considerations. Konishi and Ray (2003) model a setting where payoffs accrue in real time and coalitions evaluate moves according to a recursive continuation value, similar to our approach here; Gomes and Jehiel (2005) consider similar dynamics through the lens of non-cooperative extensive-form games. This theme also features in the study of dynamic matching, e.g., Corbae, Temzelides, and Wright (2003), Kadam and Kotowski (2018a,b), Doval (2022), and Kotowski (2024), in which players account for future play when deciding with whom to match. Rostek and Yoder (2024) propose a stability concept for static matching problems in which similar considerations emerge from players thinking strategically about the contracts that others would accept.

Since our initial draft, several papers have applied our solution concept to analyze labor markets. Liu (2023) and Liu, Wang, and Zhang (2024) study repeated matching models between long-run firms and short-run workers, identifying the limits of self-enforcing matching processes with and without transfers. Bardhi, Guo, and Strulovici (2024) examine labor markets in which firms learn about workers' types, emphasizing how early-career discrimination can result in persistent wage gaps.

[1] An alternative way to model coalitional play is through repeated extensive-form games (Mailath, Nocke, and White 2017); Hatfield, Kominers, and Lowery (2020) and Hatfield, Kominers, Lowery, and Barry (2020) use such an approach to model collusion in brokered and syndicated markets.

This paper proceeds as follows. Section 2 describes the basic framework. Section 3 identifies structural properties of PCE and characterizes its payoff set. Section 4 studies the game augmented with transfers. Section 5 applies our results to matching and distribution problems. Section 6 concludes. All proofs are in appendices.

2 Model

Players $N := \{1, 2 \ldots, n\}$ interact repeatedly at $t = 0, 1, 2, \ldots$. A coalition is a nonempty subset of N, and we denote the set of coalitions by $\mathcal{C} := 2^N \backslash \{\emptyset\}$.

The Stage Game. Let A be the set of *alternatives*, which is a compact metrizable space. An alternative a generates a payoff vector $v(a) := (v_1(a), \ldots, v_n(a)) \in \mathbb{R}^n$ for players, where the mapping $v : A \to \mathbb{R}^n$ is continuous.

Each alternative, if prescribed, may be *blocked* by individuals and coalitions. If coalition C blocks alternative a, it can choose any alternative in $E_C(a)$; if no coalition blocks, then a is played. For each coalition C, $E_C : A \rightrightarrows A$ specifies C's *effectivity correspondence* (Rosenthal 1972; Chwe 1994); we assume that this correspondence is upper and lower hemicontinuous, compact-valued, and reflexive (i.e., $a \in E_C(a)$). It simplifies notation to also assume that effectivity correspondences are monotone in the coalition: for each alternative a, $E_{C'}(a) \subseteq E_C(a)$ for $C' \subseteq C$. Our results hold without this assumption but require adapting notation, which we detail in footnotes.

This abstract stage game captures commonly studied settings, as described below.

Example 1. *Consider a normal-form game in which player i's action set, A_i, is compact: A_i can be either the set of pure actions or the set of mixtures over finite actions. The set of alternatives is the set of action profiles $A := \times_{i=1}^n A_i$. The effectivity correspondence is $E_C(a) := \left\{a' \in A : a'_j = a_j \text{ for all } j \notin C\right\}$, modeling the possibility for a blocking coalition to choose action profiles in which players outside the coalition do not change their actions. This formulation extends the standard definition for individual deviations (used in Nash equilibria).*

Example 2. *Consider characteristic function games (N, U) where for each coalition $C \in \mathcal{C}$, the mapping $U(C) \subseteq \mathbb{R}^{|C|}$ specifies a set of feasible payoff vectors for coalition C if it forms. An alternative a is now a tuple (π, u), where π is a partition of N, and $u \in \mathbb{R}^n$ is a payoff vector satisfying $u_C \in U(C)$ for each coalition $C \in \pi$. The effectivity* *correspondence $E_C(a)$ specifies the set of alternatives to which coalition C may move, and the payoff function is $v((\pi, u)) = u$.*

Example 3. *Consider majority voting, as in Bernheim and Slavov (2009). Let $\mathcal{W}$ be the set of coalitions that have at least $\lceil \frac{n}{2} \rceil$ players. The effectivity correspondence specifies that for every a, $E_C(a) = A$ if $C \in \mathcal{W}$, and $E_C(a) = \{a\}$ otherwise.*

Outcomes, Histories, and Plans. At the end of each period, a stage-game outcome $o := (a, C)$ records the chosen alternative and the identity of the blocking coalition (if any), where $C = \emptyset$ if the recommendation in that period was unblocked.[2] The set of stage-game outcomes is $\mathcal{O} := A \times 2^N$. A t-period history, $h := (a^\tau, C^\tau)_{\tau=0}^{t-1}$, specifies alternatives and blocking coalitions for t periods. We denote the set of all t-period histories by $\mathcal{H}^t$ for $t \geq 1$, and $\mathcal{H}^0 = \{\emptyset\}$ for the singleton comprising the initial null history. The set of all histories is $\mathcal{H} := \bigcup_{t=0}^{\infty} \mathcal{H}^t$. A *plan* $\sigma : \mathcal{H} \to A$ recommends an alternative following each history.

Payoffs. A path $(a^t)_{t=0,1,2,\ldots}$ is an infinite sequence of alternatives; from that path, player i accrues a normalized discounted payoff of $(1 - \delta) \sum_{t=0}^{\infty} \delta^t v_i(a^t)$, in which δ in $[0, 1)$ is a common discount factor. After a history h, a plan σ results in the path $(\sigma(h), \sigma(h, \sigma(h), \emptyset), \ldots)$ recursively and $U_i(h|\sigma)$ denotes player i's payoff from that path.

Solution Concept. Before describing our solution concept, we restate the "static core" in the language of this model. In the stage game, coalition C profitably blocks alternative a if there exists $a' \in E_C(a)$ such that $v_i(a') > v_i(a)$ for all $i \in C$. An alternative a is a *core alternative* if it cannot be profitably blocked by any coalition. A payoff vector $\widetilde{v}$ is in the *core* if there exists a core-alternative a such that $\widetilde{v} = v(a)$.

Our solution concept elaborates on this notion, where each coalition thinks about not only today but also how choices today affect continuation play.

Definition 1. *Coalition C* ***profitably blocks*** *plan σ at history h if there exists $a' \in E_C(\sigma(h))$ such that*

$$(1 - \delta) v_i(a') + \delta U_i(\, h, (a', C) \mid \sigma) > U_i(h \mid \sigma) \quad \textit{for all } i \in C.$$

[2] We include the identity of the blocking coalition in the stage-game outcome to study settings in which behavior is perfectly observed. If the stage game is a normal-form game as in Example 1, then the identity of a blocking coalition can be inferred from the chosen alternative. However, in settings like matching, the chosen alternative would not always reveal who blocked.

Definition 2. *A plan σ is a* ***perfect coalitional equilibrium*** *(PCE) if it cannot be profitably blocked by any coalition at any history.*

A PCE is a plan such that, at every history, no coalition blocks the recommended alternative given how that affects continuation play. Our notion of profitable blocking requires every coalition member to gain strictly; modifying the definition so that each coalition member benefits weakly and at least one does so strictly would not affect our results. In the interests of parsimony, we do not model what happens when multiple coalitions block simultaneously; given that no coalition profitably blocks in a PCE, allowing for this possibility would not affect our results. We formalize this consideration in the Supplementary Appendix.

Observe that if $\delta = 0$, PCEs implement only core alternatives of the stage game. Furthermore, a plan that prescribes a core alternative a^* after every history is necessarily a PCE for every $\delta \geq 0$. Below, we characterize other outcomes that can be supported by PCE.

3 What Payoffs Are Supported by PCE?

3.1 Preference Alignment and Coalitional Minmaxes

In the introduction, we mentioned how if a coalition's interests are aligned, then its members enjoy a high minmax value in any PCE. We illustrate this phenomenon using the common-interest game depicted in Table 2. Here, we adopt the specification of individual and coalitional moves stipulated in Example 1: each player can adjust her own action and the pair can choose an action profile.

	L	R
U	$1,1$	$0,0$
D	$0,0$	$0,0$

Table 2. A Game with Perfectly Aligned Preferences

This game has a unique PCE. It prescribes (U, L) at every history guaranteeing each player a payoff of 1. To see why, consider an arbitrary PCE in which each player accrues a normalized discounted payoff of $w \in [0, 1]$, and let $\underline{w}$ denote the infinum of continuation payoffs across histories. In the first period, if the pair blocks the recommended alternative and chooses (U, L) instead, each player receives at least

$(1-\delta)+\delta\underline{w}$. The pair profits from the deviation unless $w \geq (1-\delta)+\delta\underline{w}$. Because this inequality must hold for w arbitrarily close to $\underline{w}$, it then follows that $\underline{w} \geq 1$.

This example illustrates how aligned preferences create a gap between PCE and subgame perfect equilibrium (henceforth SPE) payoffs. As (D, R) is a Nash equilibrium of the stage game, all payoffs in $[0, 1]$ can be supported in SPE of the repeated game.

To formalize the relevant notion of alignment, players i and j have *aligned preferences* if there exist $k > 0$ and $c \in \mathbb{R}$ such that $v_j(a) = kv_i(a) + c$ for all $a \in A$; otherwise, their preferences are misaligned.[3] We can partition the set of players such that those in the same cell of the partition have aligned preferences and those in different cells do not. For each player i, let $C(i)$ denote the set of players whose preferences are aligned with hers.

We use alignment to define player i's *coalitional minmax*, which is the lowest payoff that she can be pushed down to when the coalition $C(i)$ collectively best-responds.

$$\underline{v}_i^\circ := \min_{a\in A} \max_{a'\in E_{C(i)}(a)} v_i(a'). \qquad \text{(Player } i\text{'s coalitional minmax)}$$

This term is well-defined as A is compact, $v(\cdot)$ is continuous, and E_C is a continuous and compact-valued correspondence.[4] Denoting the convex hull of stage-game payoffs by $\mathcal{V}$, we define the set of *strictly coalitionally rational payoffs* as $\mathcal{V}_{CR} := \{v \in \mathcal{V} : v_i > \underline{v}_i^\circ \text{ for every } i = 1, \ldots, n\}$.[5]

We distinguish this minmax from each player's individual minmax:

$$\underline{v}_i := \min_{a\in A} \max_{a'\in E_{\{i\}}(a)} v_i(a'), \qquad \text{(Player } i\text{'s individual minmax)}$$

Generally, $\underline{v}_i^\circ$ is higher than $\underline{v}_i$. The two coincide if player i's preferences do not align with those of any other player, i.e., $C(i) = \{i\}$. More strongly, $\mathcal{V}_{CR}$ coincides with the set of strictly individually rational payoffs if no two players have aligned preferences. The Roommates Problem, discussed in the introduction, lies in this class as do non-cooperative games that satisfy non-equivalent utilities (Abreu, Dutta, and Smith 1994) or full-dimensionality (Fudenberg and Maskin 1986).

[3]This criterion coincides with how Abreu, Dutta, and Smith (1994) assess non-equivalent utilities.

[4]Assuming that E_C is monotone in C simplifies the expression; absent monotonicity, the coalitional minmax would be $\underline{v}_i^\circ := \min_{a\in A} \max_{C\subseteq C(i)} \max_{a'\in E_C(a)} v_i(a')$. This expression coincides with that above if $E_C(a) \subseteq E_{C(i)}(a)$ for every player i, coalition $C \subseteq C(i)$, and alternative a.

[5]Although our setup does not have public randomization, the convex hull is relevant because these payoffs may be reached through intertemporal averaging (Sorin 1986; Fudenberg and Maskin 1991).

3.2 The Power of Scapegoat Schemes

With these preliminaries in place, we state our first result.

Theorem 1. *For every $\delta \geq 0$, every PCE gives each player i a payoff of at least $\underline{v}_i^\circ$. Moreover, for every $v \in \mathcal{V}_{CR}$, there is a $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a PCE with discounted payoff equal to v.*

The first part of the result identifies $\underline{v}_i^\circ$ as the appropriate minmax when coalitions can block: no PCE can push a player's payoff below her coalitional minmax. The second part shows that every strictly coalitionally rational payoff vector can be supported if players are sufficiently patient.

A natural comparison for Theorem 1 is to the folk theorem for SPE in repeated games with perfect monitoring. For this comparison, suppose every pair of players have misaligned preferences. Then Theorem 1's implication coincides with that of Fudenberg and Maskin (1986) and Abreu, Dutta, and Smith (1994). We highlight two differences. First, the result applies for both repeated "cooperative" and "non-cooperative" games, including settings such as repeated matching. Second, and more crucially, PCE are robust to both coalitional and individual deviations. Against this backdrop, our result clarifies that if players' preferences are misaligned and players are patient, deterring coalitional deviations is not harder than deterring individual deviations.

Why? In the proof, we design punishments that crack coalitions. Because blocking requires all coalition members to agree, we can deter coalitions by singling out and punishing just one member of each coalition—a "scapegoat"—as though she were the sole deviator, while granting amnesty to the rest. This approach assures that if coalition members' preferences are not aligned, a PCE can push each player's payoff arbitrarily close to her individual minmax.

This divide-and-conquer scheme fails if preferences within a coalition are aligned. A higher minmax then applies. To see why, consider such a coalition C and suppose towards a contradiction that some PCE σ could push the payoffs of players in C below their coalitional minmax. Observe that members of coalition C could guarantee their coalitional minmax if they could somehow commit to a long-run plan in which they collectively best respond to the recommended alternative after every history. Although coalition members cannot commit to this long-run plan, an argument similar to the one-shot deviation principle then implies that there must be some history at which coalition C profitably blocks, precluding σ from being a PCE.

An implication of Theorem 1 is that every PCE of common-interest games implements *only* the efficient action profile.[6] However, in general games, PCE may support inefficient payoffs because coalitional moves that increase surplus today can lead to continuation play that benefits some members at the expense of others.

3.3 Structural Properties

3.3.1 When Do Stationary PCEs Suffice?

Given its recursive form, PCE payoffs may be obtained using self-generation approaches from Abreu, Pearce, and Stacchetti (1990). Herein, we highlight a property that distinguishes PCE from SPE: in a rich class of games, all PCE payoffs may be achieved using PCE that are stationary.

We call a plan σ *stationary* if following every history h, the plan σ recommends the same alternative in each period so long as the plan is not blocked, i.e., $\sigma(h, (\sigma(h), \emptyset)) = \sigma(h)$. Blocking induces a transition where by a stationary plan would then recommend a potentially different alternative but would do so again in every subsequent period.

Stationarity generally restricts the set of supportable payoffs. However, it turns out to have little bite in *convex* games that have *default-independent power*. As the latter notion is more novel, we turn to it first. For each coalition C and $a \in A$, let $v_C(a) := \{(v_i(a))_{i \in C}\}$ denote the projection of $v(a)$ to C's payoff space.

Definition 3. *A stage game exhibits* ***default-independent power*** *if for every coalition C and alternatives a and a', $v_C(E_C(a)\backslash\{a\}) = v_C(E_C(a')\backslash\{a'\})$.*

Definition 3 asserts that what a coalition can achieve by blocking an alternative does not depend on that alternative. While this notion might appear stringent, several well-studied coalitional games (including two of our applications) exhibit this property.

For instance, consider any characteristic function game studied in the classical cooperative game theory literature (Example 2). In such a game, if a coalition blocks a recommended partition π, the set of utilities that it achieves does not depend on the partition. As an application here, one might consider a matching model without externalities; therein, the set of utilities that a group of players can obtain from blocking an assignment does not hinge on the assignment.

[6] In these settings, the coalitional minmax therefore departs from Wen (1994)'s effective minmax.

A second class of games comes from models of political economy in which any "winning" coalition can block and choose any policy—$E_C(a) = A$—and every non-winning coalition is completely powerless (i.e., $E_C(a) = \{a\}$). Such settings include the class of majoritarian voting rules studied in Bernheim and Slavov (2009) as well as voting rules in which players have unequal power.

Our result identifies how stationary PCEs suffice in default-independent games if the stage game is *convex*, i.e., $\{\widetilde{v} \in \mathbb{R}^n : \exists a \in A \text{ such that } \widetilde{v} = v(a)\}$ is a convex set.[7]

Theorem 2. *If the stage game is convex and exhibits default-independent power, then for every $\delta \geq 0$, the set of discounted payoffs supported by PCEs is identical to that supported by stationary PCEs.*

Theorem 2 offers a conclusion that would be unexpected of subgame perfect equilibria of repeated games; optimal penal codes often involve non-stationary play (Abreu 1988). The proof invokes both convexity and default-independent power: the former enables us to replace a non-stationary path of play with a stationary path and the latter assures that the replacement does not affect any coalition's incentives. Our result generalizes Bernheim and Slavov (2009) who obtain this conclusion for Dynamic Condorcet Winners. By clarifying that default-independent power is the key underlying property, Theorem 2 establishes that this conclusion holds much more broadly.

3.3.2 An Anti-Folk Theorem for Strongly Symmetric PCE

Green and Porter (1984) and Fudenberg, Levine, and Maskin (1994) elucidate how with monitoring imperfections, strongly symmetric equilibria are inefficient but asymmetric play can support near-efficient payoffs. The theory here offers a complementary rationale for asymmetric play that applies even with perfect monitoring: a strongly symmetric PCE cannot credibly punish players because it aligns their interests.

To make this point, we study a normal-form stage game (Example 1) that is *symmetric*: $A_i = A_j$ for all players i and j, and for each permutation μ of $\{1, \ldots, n\}$, $v_i(a_{\mu(1)}, \ldots, a_{\mu(n)}) = v_{\mu(i)}(a_1, \ldots, a_n)$ for every action profile a and player i. The set of symmetric action profiles is $A^S := \{a \in A : a_i = a_j \text{ for all } i, j \in N\}$ and

[7] We view convexity to be suitable for applications in which players can transfer utility or face a pure distribution problem. Alternatively, the set may be convex if players can access public correlation devices and make a choice to block before the realization of those lotteries. We note that this payoff set is compact because A is compact and v is continuous.

$V^S := \{v(a) : a \in A^S\}$ denotes their associated payoffs. Given that V^S is compact and totally ordered, a maximal element exists denoted $\widehat{v}$.

A plan σ is *strongly symmetric* if it recommends a symmetric action profile, $\sigma(h)$ in A^S, after every history h. Theorem 3 characterizes strongly symmetric PCE.

Theorem 3. *A strongly symmetric PCE exists if and only if the stage game has a symmetric core alternative $\widehat{a}$ such that $v(\widehat{a}) = \widehat{v}$; moreover, $\widehat{v}$ is then the unique payoff supported by a strongly symmetric PCE.*

This result reflects a collapse of intertemporal incentives in that a strongly symmetric PCE exists if and only if the highest symmetric payoff lies in the core and could therefore be supported without carrots and sticks altogether. The condition is highly restrictive, ruling out games like the repeated prisoner's dilemma or collusion in oligopolistic markets. For instance, consider the use of grim-trigger strategies to support mutual cooperation in these settings. Although such strategy profiles constitute subgame perfect equilibria, Theorem 3 implies that they do not qualify as PCEs. The challenge is that players would find it profitable to deviate at any history during the punishment phase, rendering the punishments non-credible. Theorem 1 nevertheless asserts that high cooperation payoffs can be supported by PCEs if players are patient. The construction must resort to asymmetric punishments; for example, a PCE in the prisoner's dilemma would punish a player by having her cooperate while the other defects for a specified number of periods before returning to mutual cooperation.

4 Do Transfers Align Incentives?

4.1 Transferable Utility Framework

We turn to the question of whether transfers align incentives. Because we vary the observability of transfers, we model them separately from alternatives; Section 4.2 considers publicly observed transfers and Section 4.3 models secret side-payments.

We represent transfers by $T := [T_{ij}]_{i,j\in N}$ where $T_{ij} \in [0, \infty)$ is the utility that player i transfers to player j. A player's *experienced payoff* is the sum of the payoff from the chosen alternative and net transfers: $u_i(a, T) := v_i(a) + \sum_{j\in N} T_{ji} - \sum_{j\in N} T_{ij}$. Let $\mathcal{T}$ be the set of all $n \times n$ transfer matrices in which entries along the main diagonal equal 0 (so that a player cannot transfer utility to herself). We denote transfers paid by members of coalition C by $T_C := [T_{ij}]_{i\in C,j\in N}$; $\mathcal{T}_C$ is the set of $|C| \times n$ transfer matrices.

An outcome of the stage game now includes the chosen alternative, the identity of a blocking coalition (if any), and the chosen transfers. The set of stage-game outcomes is $\overline{\mathcal{O}} := A \times 2^N \times \mathcal{T}$. Histories and paths are defined analogous to the NTU case with the addition of transfers. We denote the set of histories with transfers by $\overline{\mathcal{H}}$. A plan $\sigma : \overline{\mathcal{H}} \to A \times \mathcal{T}$ specifies an alternative and configuration of transfers, based on history. We use $a(h|\sigma)$ and $T(h|\sigma)$ to denote the recommended alternative and transfers in $\sigma(h)$. We modify the definition of $U_i(h|\sigma)$ to reflect the influence of transfers.

By blocking a recommendation (a, T), a coalition C can choose a different alternative $a' \in E_C(a)$ and change their transfers to any T'_C. A question we have to tackle is: *if a coalition blocks, what transfers do players outside the coalition make?* Two distinct answers strike us as reasonable. The first hews to a "simultaneous noncooperative" formulation in which the blocking by coalition C surprises players outside the coalition, who therefore make transfers T_{-C} as was recommended. The second models a "cooperative" approach in which if a coalition blocks, its members can transfer utility among themselves but players outside that coalition do not transfer any utility to them. To accommodate both answers, we formulate the transfers of others abstractly.

Assumption 1. For each coalition C, if C blocks a recommendation (a, T), the transfers made by players outside of C is $\chi^C(T)$ where $\chi^C : \mathcal{T} \to \mathcal{T}_{N\backslash C}$ satisfies:

1. For each bounded set $S \subseteq \mathcal{T}$, the image $\chi^C(S) \subseteq \mathcal{T}_{N\backslash C}$ is also bounded.
2. If T satisfies $T_{ij} = 0$ for all $i \notin C, j \in C$, then $\chi^C_{ij}(T) = 0$ for all $i \notin C, j \in C$.

Assumption 1 encompasses the two specifications described above: the former corresponds to $\chi^C(T) = T_{-C}$ whereas the latter corresponds to $\chi^C_{ij}(T) = T_{ij}$ for all $i, j \in N\backslash C$, and $\chi^C_{ij}(T) = 0$ for all $i \in N\backslash C$ and $j \in C$.

Thus, if coalition C blocks, chooses actions a' and changes transfers to T'_C, the realized outcome is then $(a', C, T'_C, \chi^C(T))$. We now define the versions of profitable blocking and PCE appropriate for this setting.

Definition 4. *Coalition C **profitably blocks** plan σ at history h if there exists an alternative $a' \in E_C(\, a(h|\sigma)\,)$ and transfers $T'_C = [T'_{ij}]_{i\in C, j\in N}$ such that for all $i \in C$,*

$$(1-\delta)u_i\big(a', T'_C, \chi^C(T(h|\sigma))\big) + \delta U_i\big(h, a', C', T'_C, \chi^C(T(h|\sigma)) \mid \sigma\big) > U_i(h|\sigma).$$

Definition 5. *A plan σ is a **perfect coalitional equilibrium** if it cannot be profitably blocked by any coalition at any history.*

To rule out Ponzi schemes, we make the following technical assumption.

Assumption 2. We consider plans σ such that continuation values are bounded across histories: $\{U(h|\sigma) : h \in \overline{\mathcal{H}}\}$ is a bounded subset of $\mathbb{R}^n$.

4.2 Publicly Observed Transfers

Transfers allow blocking coalitions to distribute gains among their members. One might intuit that transfers would then align coalition members' incentives. However, we find that public transfers have the opposite effect, undermining coalitions. Even those coalitions whose payoffs would be aligned absent transfers can now be splintered. Our result below establishes that all payoffs that are feasible and strictly *individually* rational can be supported.

To state this result, we re-define the set of feasible payoffs to account for transfers:

$$\mathcal{U} := \mathrm{co}\Big(\Big\{u \in \mathbb{R}^n : \exists a \in A \text{ such that } \sum_{i\in N} u_i = \sum_{i \in N} v_i(a)\Big\}\Big).$$

The set of feasible and strictly individually rational payoffs is

$$\mathcal{U}_{IR} := \{u \in \mathcal{U} : u_i > \underline{v}_i \text{ for every } i = 1, \ldots, n\}.$$

Theorem 4. *For every $\delta \geq 0$, every PCE gives each player i a payoff of at least $\underline{v}_i$. Moreover, for every $u \in \mathcal{U}_{IR}$, there is a $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a PCE with discounted payoff equal to u.*

The presence of transfers implies that if a member of a blocking coalition anticipates punishment, she can be bribed by others to still go along with it. But comparing Theorem 4 to Theorem 1 reveals that rather than aligning coalition members' incentives, public transfers actually undermine any existing preference alignment that was present before the transfers. The key idea is that public transfers make the distribution of utilities within any blocking coalition transparent to all. Therefore, a PCE can tailor the selection of a scapegoat in a blocking coalition to these transfers so as to punish the coalition member who benefited the least. Conceptually, players i and j have misaligned interests (or non-equivalent utilities) when one has to pay transfers to the other. This misalignment allows one to construct player-specific punishments.[8]

[8] Safronov and Strulovici (2018) also highlight how transfers can undermine groups in their study

4.3 Secret Transfers

In light of the analysis above, we ask: *what if some coalitions can make secret side-payments when they block? Could their incentives then be aligned?* We view this question to be of conceptual and practical import given that, in many contexts, transfers within coalitions are not public. For instance, a firm when poaching another firm's employees might offer a contract whose terms are observed by the worker and firm alone. In fact, these contracts are often confidential, a point to which we return in Section 5.1 in our discussion of wage transparency. More broadly, groups of players often seek and find ways to transfer money under the table when defecting from a social arrangement. Our analysis here identifies the benefits that coalitions accrue from making secret transfers even if their blocking decision is observable.

We consider a setting in which some but not all coalitions can make secret transfers; $\mathcal{S} \subseteq \mathcal{C}$ denote the set of coalitions that can. In our leading application, we consider firms that can offer contracts to workers with private wage terms. A secret side-payment is observed within a coalition but not outside it. Aligned with this idea, we define the outcomes that are publicly observed by all parties.

Definition 6. *Given the set $\mathcal{S} \subseteq \mathcal{C}$ and a stage-game outcome $o = (a, C, T) \in \overline{\mathcal{O}}$, the public component of o is*

$$o_p := \begin{cases} (a, C, T_{-C}) & \text{if } C \in \mathcal{S}, \\ (a, C, T) & \text{otherwise.} \end{cases}$$

For any history $h = (o^\tau)_{\tau=0}^t$, the public component of h is $h_p = (o_p^\tau)_{\tau=0}^t$.

If coalition C can make secret transfers, then whenever it blocks, any transfers made within it are not recorded in the public history. In this setting with imperfect public monitoring, we consider the analogue of a *perfect public equilibrium* (Abreu, Pearce, and Stacchetti 1990; Fudenberg, Levine, and Maskin 1994).

Definition 7. *A plan σ is **public** if $\sigma(h) = \sigma(h')$ for all $h, h' \in \overline{\mathcal{H}}$ satisfying $h_p = h'_p$. A **public PCE** is a public plan σ that constitutes a PCE of the repeated game.*

We argue below that secret transfers empower a coalition to act as if it were a single party. To this end, imagine that some coalition $C \in \mathcal{S}$ were a unitary actor that

of renegotiation; they show that the ability to punish players for proposals and transfers can cause inefficient norms to persist.

maximizes the total utility $\sum_{i\in C} v_i(\cdot)$ with an effectivity function $E_C(\cdot)$. We would then define its minmax as

$$\underline{u}_C := \min_{a\in A} \max_{a'\in E_C(a)} \sum_{i\in C} v_i(a') \qquad \text{(Coalition } C\text{'s minmax)}$$

Treating each coalition C in $\mathcal{S}$ in this way would lead to the set of feasible and strictly $\mathcal{S}$-coalitionally rational payoffs

$$\mathcal{U}_{CR}(\mathcal{S}) := \left\{ u \in \mathcal{U} : \begin{array}{r} u_i > \underline{v}_i \text{ for every } i \in N, \\ \sum_{i\in C} u_i > \underline{u}_C \text{ for every } C \in \mathcal{S} \end{array} \right\}.$$

The above set derives the set of feasible and "individually" rational payoffs in a fictitious game in which the set of players is $N \cup \mathcal{S}$. Our result below shows that this set characterizes the limits of public PCE.

Theorem 5. *For every $\delta \geq 0$, every public PCE gives each coalition $C \in \mathcal{S}$ a total payoff of at least $\underline{u}_C$ and every player i a payoff of at least $\underline{v}_i$. Moreover, for every $u \in \mathcal{U}_{CR}(\mathcal{S})$, there is a $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a public PCE with a discounted payoff equal to u.*

Theorem 5 identifies the significant gains that coalitions accrue from finding a channel to transfer utility secretly; all those in such a coalition can collectively enjoy a higher minmax while those outside a secret coalition can be pushed towards their individually rational payoffs.[9] One might view these high coalitional minmaxes as conveying an "anti-folk" flavor. Indeed, in our applications in Section 5, we show that secret transfers can reduce the supportable payoff set to the core of the stage game.

To see why Theorem 5 holds, we first explain why each coalition $C \in \mathcal{S}$ is assured its minmax. Consider a plan σ and suppose towards a contradiction that coalition C failed to achieve $\underline{u}_C$. Were coalition C a unitary actor, it could guarantee $\underline{u}_C$ by best-responding to the recommended alternative in each period. Coalition C would then profit from a long-term deviation. An argument similar to the one-shot deviation principle then establishes that at some history h, the total utility of members of coalition C must increase by profitably blocking. By apportioning that gain across the members of C through secret side-payments, it can then be assured that each member profits from the block at that history without affecting continuation play.

[9]Note that Theorem 4 corresponds to the special case of Theorem 5 in which $\mathcal{S}$ is empty.

We turn to why every payoff in $\mathcal{U}_{CR}(\mathcal{S})$ can be attained for patient players. Consider the fictitious game in which the set of players is $N \cup \mathcal{S}$. In this game, we directly construct "player-specific" punishments for each player; one can see that such punishments exist because the payoffs across the players in this fictitious game satisfy the NEU condition. Using these punishments, the payoff of each player in $N \cup \mathcal{S}$ can be pushed arbitrarily close to its minmax.

We contrast this result with Theorem 4. Therein, we could crack coalitions by fine-tuning the selection of the scapegoat to the details of who pays whom. Such an approach fails here because the continuation play cannot condition the punishment on these fine details for the coalitions in $\mathcal{S}$. Not only do these scapegoat schemes unravel but so do any other that pushes the payoff of one of these coalitions below its minmax.

5 Applications

5.1 Labor Market Matching and Wage Transparency

Many practices in labor markets, such as collective wage bargaining and firms' collusive wage-setting, are fundamentally driven by long-run incentives. We incorporate these considerations into the canonical model of Kelso and Crawford (1982) [henceforth KC82]. In the process, we obtain a new perspective on when and how wage transparency benefits workers.

In this stage game, the set of players is $N := \mathcal{F} \cup \mathcal{W}$, where $\mathcal{F}$ is the set of firms and $\mathcal{W}$ is the set of workers. We use f to denote a generic firm, w to denote a generic worker; i and j denote generic players who could be workers or firms.

Each firm can hire multiple workers. An assignment ϕ is a mapping on $\mathcal{F} \cup \mathcal{W}$ such that (i) every worker w is assigned to a firm or herself, $\phi(w) \in \mathcal{F} \cup \{w\}$; (ii) every firm f is assigned to a (potentially empty) set of workers, $\phi(f) \subseteq \mathcal{W}$; and (iii) $w \in \phi(f)$ if and only if $\phi(w) = f$. The set of alternatives A comprises all assignments between firms and workers. A matching is an assignment of workers to firms and a specification of transfers made between players. Following KC82, we allow non-zero transfers to occur only between employers and their employees. Therefore, the set of matchings is

$$\mathcal{M} := \big\{(\phi, T) \in A \times \mathcal{T} : T_{ij} \neq 0 \text{ only if } i = \phi(j) \text{ or } i \in \phi(j)\big\}.$$

Each firm f has a revenue function $v_f : 2^{\mathcal{W}} \to \mathbb{R}$, with $v_f(\emptyset)$ normalized to 0;

similarly, worker w has a premuneration utility function $v_w : \mathcal{F} \cup \{w\} \to \mathbb{R}$, where the payoff of being unemployed, $v_w(\{w\})$, is normalized to 0. Abusing notation, we use v_i to also denote the utility player i receives from an assignment, so $v_i(\phi) = v_i(\phi(i))$. Given a matching (ϕ, T), player i's experienced payoff is $u_i(\phi, T) := v_i(\phi) + \sum_{j \neq i} T_{ji} - \sum_{j \neq i} T_{ij}$.

Following KC82, we focus on three kinds of blocking: a worker can reject her match, a firm can fire all its workers, or a firm and a set of workers choose to match even if that departs from the original assignment. While none of our results change if other coalitions of players could also block, we make this assumption to match KC82 and because it embodies the realistic setting in which all contracting is between a firm and a set of workers. Formally, let $\mathcal{E} := \Big\{\{f\} \cup W : f \in \mathcal{F}, W \subseteq \mathcal{W}\Big\}$ denote all *essential* coalitions, i.e., those comprising a single firm and a set of workers. If C is a singleton or essential coalition, then for all $\phi \in A$, $E_C(\phi) = \{\phi, \phi'\}$,[10] where

1. A worker w can reject her match: if $C = \{w\}$, then $\phi'(w) = w$, and $\phi'(w') = \phi(w')$ for all $w' \in \mathcal{W} \backslash \{w\}$,
2. A firm f can fires its workers: If $C = \{f\}$, then $\phi'(f) = \emptyset$, and $\phi'(f') = \phi(f')$ for all $f' \in \mathcal{F} \backslash \{f\}$, and
3. A firm f and set of workers W can choose to match: If $C = \{f\} \cup W$, then $\phi'(f) = W$, and $\phi'(f') = \phi(f') \backslash W$ for all $f' \in \mathcal{F} \backslash \{f\}$.

This formulation specifies that if an assignment ϕ is blocked by coalition C, the resulting assignment coincides with ϕ apart from the departure made by coalition C. In other words, all untouched workers remain matched with their assigned partners.

Given that transfers happen only between matched players, those outside of C make no transfers to those in C if C blocks. This specification adheres to the "budget-balance" case described in Section 4.1; the mapping $\{\chi^C\}_{C \in \mathcal{C}}$ denotes the transfers made across players outside of coalition C.

We now state the definitions of profitable blocking and core used in KC82.

Definition 8. *A matching (ϕ, T) is **profitably blocked by coalition** C if there exists an alternative assignment $\phi' \in E_C(\phi)$ and transfers $T'_C = [T'_{ij}]_{i \in C, j \in N}$ such that all in C are better off from the matching $(\phi', T'_C, \chi^C(T))$:*

$$u_i(\phi', T'_C, \chi^C(T)) > u_i(\phi, T) \text{ for all } i \in C.$$

[10]For all other coalitions, $E_C(\phi) = \{\phi\}$.

A matching (ϕ, T) *is a* ***core allocation*** *if it cannot be profitably blocked by any coalition. The* ***stage-game core****, denoted by* $\mathcal{K}$*, are the payoffs of core allocations.*

KC82 show that if firms' revenue functions satisfy *gross substitutes*, the core is nonempty; we assume the same condition and define it formally in this footnote.[11]

Having described the stage game, we now consider the implications of repetition, using the framework and analyses of Section 4. The concept of PCE defined in Definition 5 naturally extends to this setting, where a plan recommends a stage-game matching at every history, and Definition 8 is modified to account for continuation play. The set of feasible payoffs in this repeated game is $\mathcal{U}^{\mathcal{M}} := \text{co}\Big(\Big\{u \in \mathbb{R}^n : \exists(\phi, T) \in \mathcal{M} \text{ such that } u = u(\phi, T)\Big\}\Big)$. Player i's individual minmax payoff is $\underline{v}_i = 0$, which is achieved by recommending a matching that ostracizes her. Thus, the set of feasible and individually rational payoffs is $\mathcal{U}^{\mathcal{M}}_{IR} := \Big\{u \in \mathcal{U}^{\mathcal{M}} : u_i > 0 \text{ for all } i \in N\Big\}$.

Public vs. Private Wages. The set of matchings that can be supported in the repeated game hinges on whether past wage terms are publicly or privately observed. In the former, we find that many outcomes may be supported; in the latter, we see a collapse of intertemporal incentives leading to only payoffs in the core being tenable.

Suppose all transfers are public. Then, an argument identical to Theorem 4's yields that all feasible and individually rational payoffs can be supported for patient players.

Proposition 1. *For every* $\delta \geq 0$*, every PCE gives each player* i *a payoff of at least* 0*. Moreover, for every* $u \in \mathcal{U}^{\mathcal{M}}_{IR}$*, there is a* $\underline{\delta} < 1$ *such that for every* $\delta \in (\underline{\delta}, 1)$*, there exists a PCE with discounted payoff equal to* u*.*

Now suppose each firm can hire and offer private wage terms to a group of workers. Formally, the set of secret coalitions, $\mathcal{S}$, includes all essential coalitions, $\mathcal{E}$. In this setting, we obtain a conclusion sharper than Theorem 5: all payoff vectors outside the core are untenable regardless of the players' patience.

Proposition 2. *Suppose the set of essential coalitions* $\mathcal{E}$ *can make secret transfers. For every* $\delta \geq 0$*, a public PCE supports a discounted payoff vector if and only if that payoff vector is in* $\mathcal{K}$*.*

[11] For a vector of wages from firm f, $T_f = (T_{fw})_{w \in \mathcal{W}}$, define $Ch_f(T_f) := \arg\max_{W \subseteq \mathcal{W}}(v_f(W) - \sum_{w \in W} T_{fw})$. For every set of workers W and pair of wage vectors T_f and T'_f such that $T'_{fw} \geq T_{fw}$ for all $w \in \mathcal{W}$, define $E(W, T_f, T'_f) := \{w \in W : T'_{fw} = T_{fw}\}$. Firm f's revenue function satisfies *gross substitutes* if $\widehat{W} \in Ch_f(T_f)$ implies that there exists $\widehat{W}' \in Ch_f(T'_f)$ such that $E(\widehat{W}, T_f, T'_f) \subseteq \widehat{W}'$.

Proposition 2 is an anti-folk theorem that asserts that empowering essential coalitions to make secret transfers cripples a PCE's ability to go beyond the stage-game core. The "if" direction is immediate as the infinite repetition of a core allocation constitutes a public PCE. For the "only if" direction, observe that by Theorem 5, every essential coalition is assured its minmax payoff in a public PCE. In other words, every firm f and group of workers W must achieve a total utility of at least what they would get from matching together, namely, $v_f(W) + \sum_{i \in W} v_w(f)$, which is this coalition's value. In our proof, we show that all payoff vectors that assure that each coalition obtains at least its value lie in the stage-game core.[12]

Who Benefits from Wage Transparency? To answer this question, we specialize to a setting in which workers are homogeneous (which KC82 also consider). Suppose that all workers have the same payoff function, $v_w(f) = \lambda(f)$ for each firm f and worker w. Additionally, each firm's revenue depends only on the number of workers it hires: $v_f(W) = \widetilde{v}_f(|W|)$. Let $\rho(f, l) := \lambda(f) + \widetilde{v}_f(l) - \widetilde{v}_f(l-1)$ be the surplus generated from assigning the l^{th} worker to firm f. We continue to assume that firm revenues satisfy gross substitutes, which KC82 show translates into a condition on diminishing marginal returns: $\rho(f, l)$ is then weakly decreasing in l for each f.

In this setting, the assignment ϕ^* that maximizes total social surplus is found by greedily assigning workers to firm slots in order of their contribution to total surplus; henceforth, we refer to ϕ^* as the efficient assignment. Formally, let $L := |\mathcal{W}|$ be the total labor supply and $\eta(\ell)$ be the ℓ^{th} highest value of $\{\rho(f, l) : f \in \mathcal{F}, l \geq 1\}$ for ℓ in $\{1, \ldots, L\}$, which represents the marginal value of assigning the ℓ^{th} worker optimally. To simplify our exposition, we assume that the set $\{\rho(f, l) : f \in \mathcal{F}, l \geq 1\}$ has no ties and excludes 0; assignment ϕ^* then fills "slots" $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ leaving all others vacant. Finally, we assume that it would be inefficient for a single firm to hire all workers so that each firm faces some competition.

The set of utilities compatible with assignment ϕ^* in which each player obtains more than her minmax is $\mathcal{U}^+ := \{\widetilde{u} = u(\phi^*, T) : (\phi^*, T) \in \mathcal{M}, \widetilde{u}_i > 0 \text{ for all } i \in \mathcal{F} \cup \mathcal{W}\}$. Among these, we consider a specific surplus division in which all workers obtain an

[12] We comment here on a subtle detail of our analysis. In the setup of Section 4.3, the public history does not record wage offers *only* when a coalition blocks. One may be interested in the setting in which wages are never recorded in the public history, both on- and off-path. Proposition 2 would remain true in that setting: the "if" direction holds as a PCE can support core allocations without observing any past wage offers and the "only if" direction holds because making all wage offers private imposes a further restriction on the public PCE.

identical payoff. Let $\eta(L+1)$ be the $(L+1)^{\text{th}}$ highest value of $\rho(f,l)$ assuming that there were an additional worker in the economy. Then we define:

$$\mathcal{U}^* := \{\widetilde{u} \in \mathcal{U}^+ : \max\{0, \eta(L+1)\} \leq \widetilde{u}_w = \widetilde{u}_{w'} \leq \max\{0, \eta(L)\} \text{ for all } w, w' \in \mathcal{W}\}.$$

In these surplus divisions, each worker obtains a net utility of approximately the "marginal product" of the last employee in the economy while firms are residual claimants. We show that $\mathcal{K} = \mathcal{U}^*$, which yields the following conclusion.

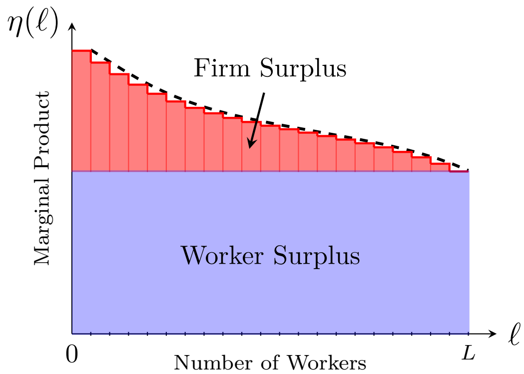
Proposition 3. *If wages are public, for each $u \in \mathcal{U}^+$, there exists $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a PCE with discounted payoff equal to u. By contrast, if wages are private, for every $\delta \geq 0$, the set of payoffs that can be supported by public PCEs is $\mathcal{U}^*$.*

Proposition 3 asserts that any surplus division from the efficient assignment ϕ^* in which individual rationality conditions hold can be supported if wages are public. Firms could collude to extract nearly all surplus from workers; alternatively, workers can collectively bargain to retain almost the entire surplus. By contrast, if wages are private, workers accrue the value of the marginal product of the least productive employee, and firms capture the remaining surplus.

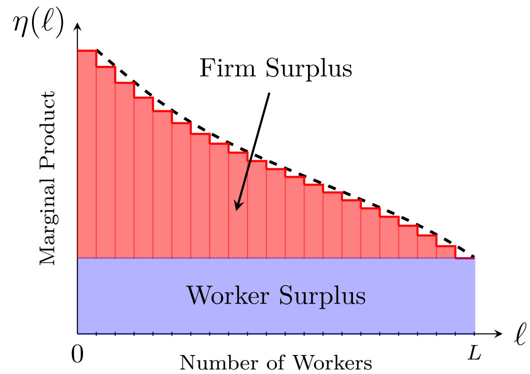
Given Proposition 3, workers favor wage transparency if they are plentiful—i.e., $\eta(L) < 0$—or their marginal product falls quickly. Without transparency, workers compete intensely for slots and thereby drive their earnings to near 0. By contrast, wage transparency enables them to use collective bargaining to obtain higher wages for them all. In such a scheme, were a firm to try to poach workers in a way that is mutually profitable, a PCE would deter workers from accepting those offers by reverting to the stage-game core from the next period onwards. Thus, workers recognize that the future promise of high wages—and the continued success of their collective bargaining efforts—requires them to reject offers that are tempting today.

By contrast, if workers are scarce or the marginal product of workers falls slowly—i.e., $\eta(1) \approx \eta(L)$—it is firms who favor wage transparency. All PCE under private wages result in high wages, as firms compete heavily for workers. Wage transparency allows firms to collusively suppress wages, with all of them setting low wages and agreeing not to poach each other's workers. Such an agreement is viable given continuation play in which poaching today triggers a "salary war" tomorrow.

We depict this prediction in Figure 2: (A) shows the distribution of worker and firm



(A) Marginal product falls slowly



(B) Marginal product falls quickly

Figure 2. (A) and (B) show the distribution of surplus under private wages when the marginal productivity falls slowly or quickly. In the latter case, workers have more to gain from wage transparency.

surplus under private wages when the marginal product of labor falls slowly and (B) shows the same when the marginal product falls quickly. As the figure shows, workers are worse off absolutely and relatively in the latter case. Were wages transparent, workers or firms could obtain better terms. In (A), workers have little to gain but much to lose from wage transparency as firms could then suppress wages; by contrast, in (B), it is workers that can use wage transparency to secure a larger share of the pie.

Formalizing this comparative statics prediction, consider two markets M_1 and M_2 that are identical in all respects but one: they differ in the productivity of labor as captured in firms' revenues. In market i, firm f's revenue function is $\widetilde{v}_{f,i}$, and the marginal value of assigning worker ℓ optimally is then $\eta_i(\ell)$. We assume that labor is valuable in each market, in that $\eta_i(L+1) > 0$ for each i, and that the two markets accrue the same gain from hiring the first worker, $\eta_1(1) = \eta_2(1)$.

Definition 9. *Market M_2 exhibits **more steeply decreasing returns to labor** than market M_1 if for every ℓ in $\{1, \ldots, L\}$,*

$$\eta_2(\ell) - \eta_2(\ell+1) \geq \eta_1(\ell) - \eta_1(\ell+1).$$

*If the inequality is strict for some ℓ, then M_2 exhibits **strictly** more steeply decreasing returns to labor.*

In each market, given gross substitutes, the marginal product of labor falls with each incremental worker; Definition 9 asserts that this fall is always more pronounced in M_2. Modulo integer issues, this definition translates into a standard condition on the second derivative of the total product being more negative in M_2.[13]

We turn to the implications for how surplus is divided between workers and firms. Let $\Pi_i := \sum_{\ell=1}^{L} \eta_i(\ell)$ denote the total surplus from the efficient assignment in market M_i. Let $\Pi_i^{\mathcal{W}} := [L\eta_i(L+1), L\eta_i(L)]$ denote the set of potential workers' total surplus under private wages; recall from Proposition 3 that each worker is paid the same, which is around the marginal product of the least productive worker. Firms capture the gap between total surplus and that taken by workers; $\Pi_i^{\mathcal{F}}$ denotes the set of potential firms' total surplus. We compare these surplus divisions between the two markets; when comparing sets, we use the strong set order denoted $\succcurlyeq_{SSO}$.

Proposition 4. *Suppose M_2 exhibits more steeply decreasing returns to labor than M_1. Then the following hold about the distribution of surplus under private wages:*

(a) *The total surplus in market M_1 is higher: $\Pi_1 \geq \Pi_2$.*

(b) *Worker surplus in market M_1 must be higher: $\Pi_1^{\mathcal{W}} \succcurlyeq_{SSO} \Pi_2^{\mathcal{W}}$.*

(c) *Firm surplus in market M_1 must be lower: $\Pi_1^{\mathcal{F}} \preccurlyeq_{SSO} \Pi_2^{\mathcal{F}}$.*

Furthermore, if M_2 exhibits strictly more steeply decreasing returns to labor, then all the orders above are strict.

Proposition 4 identifies an interesting property: while a more steeply decreasing returns reduces both total and worker surplus under private wages, it has a more pronounced effect on the latter. Hence, as seen in (c), the residual surplus captured by firms is actually higher in M_2. If wage transparency enables workers to capture firms' profits, then workers have more to gain (and less to potentially lose) in M_2 than M_1.

5.2 Distributive Politics

Herein, we study a repeated distribution problem, in which the players repeatedly choose how to divide a dollar. Such division problems feature prominently in the political economy literature (e.g. Baron and Ferejohn 1989) and relate to the *simple games* (Von Neumann and Morgenstern 1945) studied in cooperative game theory. The set of alternatives A are divisions of the dollar, $\{a \in \mathbb{R}_+^N : \sum_{i \in N} a_i = 1\}$, where player

[13] The "total product" in each market with ℓ units of labor would be $\widehat{\Pi}_i(\ell) := \sum_{l=1}^{\ell} \eta_i(l)$. As the first worker in markets M_1 and M_2 generates the same gain, Definition 9 implies that $\widehat{\Pi}_2(\cdot)$ must be a concave transformation of $\widehat{\Pi}_1(\cdot)$, which is tantamount to the standard Arrow-Pratt comparison.

i's payoff from alternative a is a_i. Divisions are chosen by a "winning" coalition: $\mathcal{W}$ is a set of coalitions such that for every coalition C in $\mathcal{W}$, $E_C(a) = A$ for every division a, and for every coalition C not in $\mathcal{W}$, $E_C(a) = \{a\}$. As standard, $\mathcal{W}$ is monotone and proper.[14] A simple-majority rule protocol corresponds to $\mathcal{W}$ comprising all coalitions that have at least $(n+1)/2$ players. This formulation also allows for *veto power*: if a player belongs to every winning coalition ($\cap_{C\in\mathcal{W}} C$), then effectively no block can happen without her approval. We denote the set of veto players by $D := \cap_{C\in\mathcal{W}} C$.

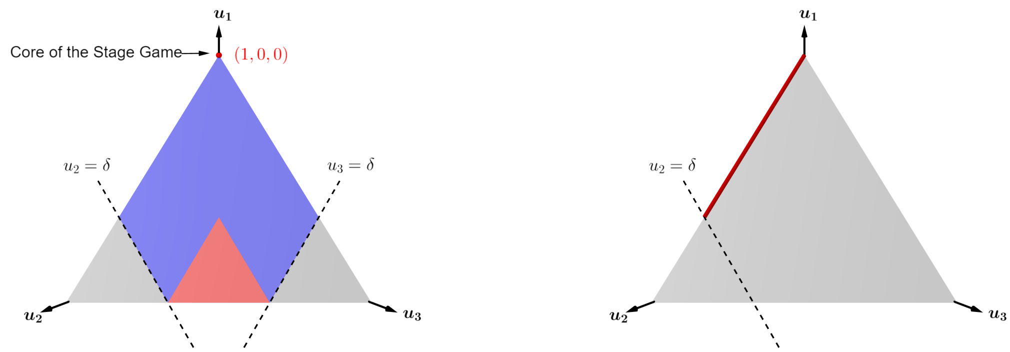
Bernheim and Slavov (2009) approach this setting with simple majority-rule in mind, emphasizing how Dynamic Condorcet Winners exist although the stage game lacks a Condorcet Winner. We focus instead on settings with at least one veto and one non-veto player, and in which veto players are not dictators ($D \notin \mathcal{W}$). Absent history dependence, these settings are prone to highly unequal splits: the veto players steal the entire dollar. Formally, the set of core alternatives of the stage game is $\mathcal{K} := \{a \in \mathbb{R}^N_+ : \sum_{i\in D} a_i = 1\}$. The logic is that any division that gives a positive share to a non-veto player would be profitably blocked by a winning coalition who would extract that share and divide it among themselves.

Against this backdrop, we evaluate how history dependence can counter this tendency towards unequal splits. Consider a three-player example in which player 1 alone has veto power; however, she needs the support of at least one other player to block. In the core of this stage game, player 1 captures the entire dollar. Nevertheless, relatively simple schemes in the repeated game can promote equal splits. Consider a core reversion plan that prescribes $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ every period if that has been the division up to now and switches to the stage-game core otherwise. On the equilibrium path, even if player 1 offers the entire dollar to either player 2 or 3, neither finds it profitable to block with her if $(1-\delta)(1) + \delta(0) \leq \frac{1}{3}$. Going further, core-reversion can support any division in the triangle formed by the vertices $\{(2\delta - 1, 1-\delta, 1-\delta), (0, \delta, 1-\delta), (0, 1-\delta, \delta)\}$, which converges to the unit simplex as $\delta \to 1$.

One could go beyond core-reversion to characterize all PCE payoffs. Because the game is convex and exhibits default-independent power, Theorem 2 implies that all PCE payoffs can be supported by stationary PCE. Using this result, we find that if players are sufficiently patient, then every payoff in which each non-veto player obtains up to δ can be supported in a PCE. We depict these outcomes in Figure 3(A).

These schemes collapse if the veto player can make and receive secret side-payments.

[14] In other words, if C is in $\mathcal{W}$, then $\mathcal{W}$ contains every superset of C but not its complement.



(A) Perfect monitoring for $\delta > 1/2$

(B) Secret transfers between 1 and 2

Figure 3. (A) depicts the set of supportable outcome. The red region depicts payoffs supported by core-reversion, and the blue region illustrates those from other PCE. (B) shows the set of supportable payoffs once coalition $\{1, 2\}$ can make secret transfers; player 3 then obtains 0.

Suppose players 1 and 2 can transfer utility under the table. All PCE payoffs then involve player 3 obtaining 0, as illustrated in Figure 3(B). Even worse, player 1 can take the entire dollar in every period if she can make secret side-payments with each player.

These intuitions generalize to n-player games in which there are at least one veto and one non-veto player, and veto players are not dictators. We call a coalition C a *minimal winning coalition* if C is a winning coalition and every proper subset is not.

Proposition 5. *The following hold:*

(a) *Absent secret transfers, there exists $\underline{\delta} \geq 0$ such that if $\delta \geq \underline{\delta}$, the set of supportable payoffs are those that give at least $(1 - \delta)$ to each winning coalition.*

(b) *If some winning coalition C can make secret transfers, then that coalition must capture the entire dollar.*

(c) *If every minimal winning coalition can make secret transfers, then the veto players must capture the entire dollar.*

Proposition 5(a) highlights how generally, egalitarian schemes can be supported by history dependence. For this fixed discount factor characterization, we exploit the feature that this game is convex and exhibits default-independent power; it turns out that if there are two or more veto players, then the $\underline{\delta}$ referred to above equals 0 so the characterization above is complete. Proposition 5(b) and (c) elucidate how secret side-payments destabilize egalitarian schemes: any coalition that can make secret transfers obtain the entire dollar, and if all minimal winning coalitions can make secret transfers, the veto players regain de facto dictatorial power.

6 Conclusion

This paper develops a portable and tractable framework for coalitional repeated games. The framework enables us to study how carrots and sticks can discipline coalitions in a wide class of games. Our analysis uncovers the importance of alignment: history dependence curbs coalitional deviations if and only if players' interests are misaligned. Simple scapegoat schemes then deter coalitional deviations. However, if players in a coalition have aligned interests, they can effectively act as a unitary agent, securing a higher minmax payoff.

This perspective delivers additional insights. Strongly symmetric schemes are generally ineffective against coalitional deviations, pushing towards the use of asymmetric punishments. Being able to transfer utility alone does not align interests; to the contrary, publicly observed transfers create a wedge between coalition partners and thereby undermine coalitions. However, the ability to make transfers under the table forges strong ties: a coalition that can do so is assured a high net payoff across PCE. In our applications, these secret side-payments can cripple intertemporal incentives, reducing the set of supportable outcomes to the stage-game core. We use these results to identify conditions under which workers favor wage transparency in repeated labor-market matching; when workers are plentiful, observing wage terms enables them to collectively bargain for a larger share of the surplus. We also apply these results to repeated multilateral negotiations to investigate when history dependence can counter the power of veto players.

While our framework answers some questions, it leaves others unanswered. First, we study purely repeated interaction in which choices today have no direct bearing on future payoffs; a natural direction is to study settings like dynamic public good provision, natural resource depletion, or experimentation in which actions in one period directly impact options (or beliefs) in the next.[15] Second, the economic forces underlying our study of labor market matching appear to apply more broadly, potentially in a setting that would allow for more general (spot-)contracts.

[15] In this vein, Bardhi, Guo, and Strulovici (2024) apply our solution concept to a dynamic matching labor market in which firms learn about worker types.

References

Abreu, Dilip. 1988. "On the Theory of Infinitely Repeated Games with Discounting." *Econometrica* 56 (2):383–396.

Abreu, Dilip, Prajit K. Dutta, and Lones Smith. 1994. "The Folk Theorem for Repeated Games: A NEU Condition." *Econometrica* 62 (4):939–948.

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. "Toward A Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica* :1041–1063.

Aumann, Robert J. 1959. "Acceptable Points in General Cooperative n-Person Games." In *Contributions to the Theory of Games IV*, vol. 4, edited by H. W. Kuhn and R. D. Luce. Princeton, NJ: Princeton University Press, 287.

Bardhi, Arjada, Yingni Guo, and Bruno Strulovici. 2024. "Early-Career Discrimination: Spiraling or Self-Correcting?" Working Paper.

Baron, David P. and John A. Ferejohn. 1989. "Bargaining in Legislatures." *American Political Science Review* 83 (4):1181–1206.

Barron, Daniel and Yingni Guo. 2021. "The use and misuse of coordinated punishments." *The Quarterly Journal of Economics* 136 (1):471–504.

Bernheim, B. Douglas and Debraj Ray. 1989. "Collective Dynamic Consistency in Repeated Games." *Games and Economic Behavior* 1 (4):295–326.

Bernheim, B. Douglas and Sita N. Slavov. 2009. "A Solution Concept for Majority Rule in Dynamic Settings." *Review of Economic Studies* 76 (1):33–62.

Chwe, Michael. 1994. "Farsighted Coalitional Stability." *Journal of Economic Theory* 63 (2):299–325.

Corbae, Dean, Ted Temzelides, and Randall Wright. 2003. "Directed Matching and Monetary Exchange." *Econometrica* 71 (3):731–756.

Cullen, Zoë. 2024. "Is Pay Transparency Good?" *Journal of Economic Perspectives* 38 (1):153–80.

Cullen, Zoë B and Bobak Pakzad-Hurson. 2023. "Equilibrium effects of pay transparency." *Econometrica* 91 (3):765–802.

DeMarzo, Peter M. 1992. "Coalitions, Leadership, and Social Norms: The Power of Suggestion in Games." *Games and Economic Behavior* 4 (1):72–100.

Doval, Laura. 2022. "Dynamically Stable Matching." *Theoretical Economics* 17 (2):687–724.

Farrell, Joseph and Eric Maskin. 1989. "Renegotiation in Repeated Games." *Games and Economic Behavior* 1 (4):327–360.

Fudenberg, Drew, David Levine, and Eric Maskin. 1994. "The Folk Theorem with Imperfect Public Monitoring." *Econometrica* 62 (5):997–1039.

Fudenberg, Drew and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* :533–554.

———. 1991. "On the Dispensability of Public Randomization in Discounted Repeated Games." *Journal of Economic Theory* 53 (2):428—438.

Gomes, Armando and Philippe Jehiel. 2005. "Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies." *Journal of Political Economy* 113 (3):626–667.

Green, Edward J and Robert H Porter. 1984. "Noncooperative collusion under imperfect price information." *Econometrica* :87–100.

Hatfield, John William, Scott Duke Kominers, and Richard Lowery. 2020. "Collusion in brokered markets." Working Paper.

Hatfield, John William, Scott Duke Kominers, Richard Lowery, and Jordan M. Barry. 2020. "Collusion in Markets with Syndication." *Journal of Political Economy* 128 (10):3779–3819.

Hatfield, John William, Scott Duke Kominers, Alexandru Nichifor, Michael Ostrovsky, and Alexander Westkamp. 2013. "Stability and competitive equilibrium in trading networks." *Journal of Political Economy* 121 (5):966–1005.

Hatfield, John William and Paul R Milgrom. 2005. "Matching with Contracts." *American Economic Review* 95 (4):913–935.

Kadam, Sangram V. and Maciej H. Kotowski. 2018a. "Multiperiod Matching." *International Economic Review* 59 (4):1927–1947.

———. 2018b. "Time Horizons, Lattice Structures, and Welfare in Multi-Period Matching Markets." *Games and Economic Behavior* 112:1–20.

Kelso, Alexander S and Vincent P Crawford. 1982. "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica* :1483–1504.

Konishi, Hideo and Debraj Ray. 2003. "Coalition Formation as A Dynamic Process." *Journal of Economic Theory* 110 (1):1–41.

Kotowski, Maciej H. 2024. "A perfectly robust approach to multiperiod matching problems." *Journal of Economic Theory* 222:105919.

Liu, Ce. 2023. "Stability in Repeated Matching Markets." *Theoretical Economics* 18 (4):1711–1757.

Liu, Ce, Ziwei Wang, and Hanzhe Zhang. 2024. "Self-Enforced Job Matching." Working Paper.

Mailath, George J., Volker Nocke, and Lucy White. 2017. "When and How The Punishment Must Fit The Crime." *International Economic Review* 58 (2):315–330.

Miller, David A and Joel Watson. 2013. "A theory of disagreement in repeated games with bargaining." *Econometrica* 81 (6):2303–2350.

Rosenthal, Robert W. 1972. "Cooperative Games in Effectiveness Form." *Journal of Economic Theory* 5 (1):88–101.

Rostek, Marzena J and Nathan Yoder. 2024. "Matching with Strategic Consistency." Working Paper.

Rubinstein, Ariel. 1980. "Strong Perfect Equilibrium in Supergames." *International Journal of Game Theory* 9 (1):1–12.

Safronov, Mikhail and Bruno Strulovici. 2018. "Contestable norms." Working Paper.

Sorin, Sylvain. 1986. "On Repeated Games with Complete Information." *Mathematics of Operations Research* 11 (1):147–160.

Von Neumann, John and Oskar Morgenstern. 1945. *Theory of Games and Economic Behavior*. Princeton University Press Princeton, NJ.

Wen, Quan. 1994. "The "Folk Theorem" for Repeated Games with Complete Information." *Econometrica* :949–954.

A Appendix

The main appendix contains proofs for Theorems 1, 2, 3, and 5. All other proofs are in the Supplementary Appendix. Throughout our analysis, we use sequences of play to convexify payoffs, following standard arguments from Sorin (1986) and Fudenberg and

Maskin (1991). Below, we reproduce the statement that we invoke in our arguments.

Lemma 1. ***(Lemma 2 of Fudenberg and Maskin 1991)*** *Let X be a convex polytope in $\mathbb{R}^n$ with vertices $x^1, \ldots, x^K$. For all $\epsilon > 0$, there exists a $\underline{\delta} < 1$ such that for all $\underline{\delta} < \delta < 1$, and any $x \in X$, there exits a sequence $\{x_\tau\}_{\tau=0}^\infty$ drawn from $\{x^1, \ldots, x^K\}$, such that $(1-\delta)\sum_{\tau=0}^\infty \delta^\tau x_\tau = x$ and at any t, $||x - (1-\delta)\sum_{\tau=t}^\infty \delta^{\tau-t} x_\tau|| < \epsilon$.*

A.1 Proof of Theorem 1 on p. 10

A Preliminary Result. A *blocking plan* by coalition C from a plan σ is a function $\alpha : \mathcal{H} \to A$ such that $\alpha(h) \in E_C(\sigma(h))$ for every history $h \in \mathcal{H}$. After each history h, the blocking plan α generates a path $(\alpha(h), \alpha(h, \alpha(h), C), \ldots)$ that is distinct from the one generated by σ. We use $U_i(h|\alpha)$ to denote player i's normalized discounted payoff from that path. The blocking plan α is profitable if there exists a history h such that $U_i(h|\alpha) > U_i(h|\sigma)$ for all $i \in C$. Below, we say that a coalition is in the alignment partition if it corresponds to a coalition $C(i)$ for some player i.

Lemma 2. *If σ is a PCE, then no coalition in the alignment partition has a profitable blocking plan.*

Proof. Consider a plan σ from which coalition $C(i^*)$ has a profitable blocking plan α for some $i^* \in N$. In particular, there exists a history $h \in \mathcal{H}$ such that $U_i(h|\alpha) > U_i(h|\sigma)$ for every $i \in C(i^*)$. We show that coalition $C(i^*)$ must then have a profitable block from the plan σ at some history, so σ is not a PCE.

Since the set of alternatives A is compact and $v : A \to \mathbb{R}^n$ is continuous, the plan σ has bounded continuation values for all players. Given discounting, the standard one-shot deviation principle applies. Therefore, there exists a history $\widehat{h} \in \mathcal{H}$ such that

$$(1-\delta)u_{i^*}\big(\alpha(\widehat{h})\big) + \delta U_{i^*}\big(\widehat{h}, \alpha(\widehat{h}), C(i^*)\big|\sigma\big) > U_{i^*}(\widehat{h}|\sigma).$$

Since $C(i^*)$ has aligned payoffs, for each $j \in C(i^*)$ there exists $\lambda_{ji^*} > 0$ and $\mu_{ji^*} \in \mathbb{R}$ such that $u_j(a) = \lambda_{ji^*} u_{i^*}(a) + \mu_{ji^*}$ and all alternatives $a \in A$; in addition, for every $j \in C(i^*)$, the discounted payoffs satisfy $U_j(h|\sigma) = \lambda_{ji^*} U_{i^*}(h|\sigma) + \mu_{ji^*}$ at every history $h \in \mathcal{H}$. Substituting into the inequality above, we have

$$(1-\delta)u_j\big(\alpha(\widehat{h})\big) + \delta U_j\big(\widehat{h}, \alpha(\widehat{h}), C(i^*)\big|\sigma\big) > U_j(\widehat{h}|\sigma) \text{ for all } j \in C(i^*).$$

Therefore, coalition $C(i^*)$ has a profitable block at history $\widehat{h}$. $\square$

Proof of Theorem 1.

<u>Part 1</u>: *For every $\delta \geq 0$, every PCE gives each player i a payoff of at least $\underline{v}_i^\circ$.*

We establish this claim by proving its contrapositive: let σ be a plan, and suppose there exists a player i^* that satisfies $U_{i^*}(\emptyset|\sigma) < \underline{v}_{i^*}^\circ$. We show that σ cannot be a PCE. Given Lemma 2, it suffices to show that coalition $C(i^*)$ has a profitable blocking plan.

Consider the following blocking plan α for coalition $C(i^*)$: at every history h, coalition $C(i^*)$ chooses its myopic best response to the recommended alternative, $\alpha(h) \in \arg\max_{a' \in E_{C(i^*)}(\sigma(h))} v_{i^*}(a')$. By the definition of $\underline{v}_{i^*}^\circ$, $v_{i^*}(\alpha(h)) \geq \underline{v}_{i^*}^\circ$ for every history h, so player i^*'s continuation value from period 0 must be higher: $U_{i^*}(\emptyset|\alpha) > U_{i^*}(\emptyset|\sigma)$. Given that all players $j \in C(i^*)$ have aligned payoffs, $U_j(\emptyset|\alpha) > U_j(\emptyset|\sigma)$ for all $j \in C(i^*)$, so α is a profitable blocking plan for coalition $C(i^*)$.

<u>Part 2</u>: *For every $v \in \mathcal{V}_{CR}$, there is a $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a PCE with discounted payoff equal to v.*

Fixing $v^* \in \mathcal{V}_{CR}$, we proceed with the following the steps.

First, observe that for any pair of players i, j such that $j \notin C(i)$, their payoffs satisfy the NEU condition. By Lemma 1 and Lemma 2 of Abreu, Dutta, and Smith (1994), we can find *coalition-specific punishments* for v^*: there exist payoff vectors $\{v^{C(i)}\}_{i=1}^n \subseteq \mathcal{V}_{CR}$ such that $v_i^{C(i)} < v_i^*$ for all $i \in N$, and $v_i^{C(i)} > v_j^{C(i)}$ for all $j \notin C(i)$.

Second, let us define *coalitional minmaxing alternatives*: for each coalition $C(i)$, let $\underline{a}_{C(i)}^\circ \in \arg\min_{a \in A} \max_{a' \in E_{C(i)}(a)} v_j(a')$ for some $j \in C(i)$—note that the specific choice of $j \in C(i)$ in the definition does not matter given the aligned payoffs within $C(i)$—as the alternative that will be used to minmax coalition $C(i)$. Since A is compact, v is continuous, and $E_{C(i)}(\cdot)$ is continuous and compact-valued, by Berge's maximum theorem, $\underline{a}_{C(i)}^\circ$ is well-defined for each $i \in N$. By construction, $v_i(a') \leq \underline{v}_i^\circ$ for all $i \in N$ and $a' \in E_{C(i)}(\underline{a}_{C(i)}^\circ)$ and in particular, $v_i(\underline{a}_{C(i)}^\circ) \leq \underline{v}_i^\circ$.

Given these payoffs and punishments, let $\kappa \in (0,1)$ be such that for every $\widetilde{\kappa} \in [\kappa, 1]$, the following is true for every i:

$$(1-\widetilde{\kappa})v_i(\underline{a}_{C(i)}^\circ) + \widetilde{\kappa} v_i^{C(i)} > \underline{v}_i^\circ \tag{1}$$

For every $i \in N$ and $j \notin C(i)$:
$$(1-\widetilde{\kappa})v_j(\underline{a}_{C(i)}^\circ) + \widetilde{\kappa} v_j^{C(i)} > (1-\widetilde{\kappa})\underline{v}_j^\circ + \widetilde{\kappa} v_j^{C(j)} \tag{2}$$

Inequality (1) implies that every player $j \in C(i)$ is willing to bear the cost of

$v_j(\underline{a}^{\circ}_{C(i)})$ with the promise of transitioning into their coalition-specific punishment rather than staying at their coalitional minmax, where the promise is discounted at $\widetilde{\kappa}$. Similarly, inequality (2) implies that player j is willing to bear the cost of minmaxing any coalition with whom j does not share equivalent utilities, given the promise of transitioning into coalition $C(i)$'s specific punishment rather than her own, when the post-minmaxing phase payoffs are discounted at $\widetilde{\kappa}$. Each inequality holds at $\widetilde{\kappa} = 1$ for all i and $j \notin C(i)$. Since the set of players is finite, there exists a value of $\kappa \in (0,1)$ such that the inequality holds for all $\widetilde{\kappa} \in [\kappa, 1]$, $i \in N$ and $j \notin C(i)$.

Let $L(\delta) := \left\lceil \frac{\log \kappa}{\log \delta} \right\rceil$ where $\lceil \cdot \rceil$ is the ceiling function. Observe that $\delta^{L(\delta)} \in [\delta^{\frac{\log \kappa}{\log \delta}+1}, \delta^{\frac{\log \kappa}{\log \delta}}] = [\delta\kappa, \kappa]$. Therefore, $\lim_{\delta \to 1} \delta^{L(\delta)} = \kappa$.

Since $\{v^{C(i)}\}_{i=1}^n \cup \{v^*\} \subseteq \text{co}\{v(a) : a \in A\} \subseteq \mathbb{R}^n$, by Carathéodory's theorem, there exist $\{\widehat{a}^1, \ldots, \widehat{a}^K\} \subseteq A$ for some integer K, such that $\{v^{C(i)}\}_{i=1}^n \cup \{v^*\} \subseteq \text{co}\{v(\widehat{a}^k) : k = 1, \ldots, K\}$. Define $\mathcal{I} := \{C(i)\}_{i=1}^n$, and $\widehat{\mathcal{I}} := \{C(i)\}_{i=1}^n \cup \{*\}$. Lemma 1 then guarantees that for any $\epsilon > 0$, there exists $\underline{\delta} \in (0,1)$ such that for all $\delta \in (\underline{\delta}, 1)$, there exist sequences $\big\{\{a^{S,\tau}\}_{\tau=0}^{\infty} : S \in \widehat{\mathcal{I}}\big\}$ such that for each $S \in \widehat{\mathcal{I}}$ and t, $(1-\delta)\sum_{\tau=0}^{\infty} \delta^{\tau} v(a^{S,\tau}) = v^S$ and $\big|\big|v^S - (1-\delta)\sum_{\tau=t}^{\infty} \delta^{\tau} v(a^{S,\tau})\big|\big| < \epsilon$. We fix an

$$\epsilon < (1-\kappa) \min\Big\{ \min_{S \in \widehat{\mathcal{I}}, i \in N \setminus S} (v_i^S - v_i^{C(i)}),\ \min_{i \in N} v_i^{C(i)} - \underline{v}_i^{\circ} \Big\},$$

and given that ϵ, consider δ exceeding the appropriate $\underline{\delta}$.

We now describe the plan that supports v^*. Consider the automaton $(W, w(*,0), f, \gamma)$:

- $W := \big\{w(d,\tau) | d \in \widehat{\mathcal{I}}, \tau \geq 0\big\} \cup \{\underline{w}(S,\tau) | S \in \mathcal{I}, 0 \leq \tau < L(\delta)\}$ is the set of possible states and $w(*,0)$ is the initial state;
- $f : W \to \mathcal{O}$ is the output function, where $f(w(d,\tau)) = (a^{d,\tau}, \emptyset)$ and $f(\underline{w}(S,\tau)) = (\underline{a}^{\circ}_S, \emptyset)$.
- $\gamma : W \times \mathcal{O} \to W$ is the transition function. For states of the form $w(d,\tau)$, the transition is

$$\gamma\big(w(d,\tau), (a, C)\big) = \begin{cases} \underline{w}(C(j^*), 0) & \text{if } C \neq \emptyset \text{ , where } j^* = \min_{j \in C} j \\ w(d, \tau+1) & \text{otherwise} \end{cases}$$

For states of the form $\{\underline{w}(S,\tau)|0 \leq \tau < L(\delta)-1\}$,

$$\gamma\big(\underline{w}(S,\tau),(a,C)\big) = \begin{cases} \underline{w}(C(j^*),0) & \text{if } C \nsubseteq S \text{ , where } j^* = \min_{j\in C\setminus S} j \\ \underline{w}(S,0) & \text{if } C \subseteq S \text{ and } C \neq \emptyset \\ \underline{w}(S,\tau+1) & \text{otherwise} \end{cases}$$

For states of the form $\underline{w}(S, L(\delta)-1)$, the transition is

$$\gamma\big(\underline{w}(S,L(\delta)-1),(a,C)\big) = \begin{cases} \underline{w}(C(j^*),0) & \text{if } C \nsubseteq S \text{ , where } j^* = \min_{j\in C\setminus S} j \\ \underline{w}(S,0) & \text{if } C \subseteq S \text{ and } C \neq \emptyset \\ w(S,0) & \text{otherwise} \end{cases}$$

The plan represented by the above automaton yields payoff profile v^*. By construction, for $\tau = 0, 1, L(\delta)$, $\big|\big|v^d - V(w(d,\tau))\big|\big| < \epsilon$ and

$$V(\underline{w}(S,\tau)) = (1-\delta^{L(\delta)-\tau})v(\underline{a}^\circ_S) + \delta^{L(\delta)-\tau}V(w(S,0)).$$

Below, we show that this plan is a PCE by showing that no coalition can profitably block in any state of this automaton.

States of the form $w(d,\tau)$: Set $B > \max_{a\in A, i\in N} v_i(a)$. Consider coalition C blocking and implementing the alternative a. Let $j^* = \min\{j \in C\}$. For all τ, without the blocking j^* obtains a payoff greater than $v^d_{j^*} - \epsilon$. By participating in the blocking, j^* obtains a payoff less than

$$(1-\delta)B + \delta V_{j^*}(\underline{w}(C(j^*),0)) = (1-\delta)B + \delta\left[(1-\delta^{L(\delta)})v_{j^*}(\underline{a}^\circ_{C(j^*)}) + \delta^{L(\delta)}v^{C(j^*)}_{j^*}\right]$$

For the blocking to be profitable, everyone in C, including player j^*, must be better off. So this is not a profitable block for C if the above term is no more than $v^d_{j^*} - \epsilon$. We prove that this is the true in two separate cases.

First consider the case where $d \in \widehat{\mathcal{I}}\setminus\{C(j^*)\}$. Observe that

$$\begin{aligned} &\lim_{\delta\to 1}(1-\delta)B + \delta\left[(1-\delta^{L(\delta)})v_{j^*}(\underline{a}^\circ_{C(j^*)}) + \delta^{L(\delta)}v^{C(j^*)}_{j^*}\right] \\ =&\lim_{\delta\to 1}\left[(1-\delta^{L(\delta)})v_{j^*}(\underline{a}^\circ_{C(j^*)}) + \delta^{L(\delta)}v^{C(j^*)}_{j^*}\right] < v^{C(j^*)}_{j^*}, \end{aligned}$$

where the inequality follows from $v_{j^*}(\underline{a}^\circ_{C(j^*)}) \leq \underline{v}^\circ_j < v^{C(j^*)}_{j^*}$. Because ϵ by construction is strictly less than $v^d_{j^*} - v^{C(j^*)}_{j^*}$, it follows that payoff from blocking is less than $v^d_{j^*} - \epsilon$ when δ is sufficiently large.

Now suppose that $d = C(j^*)$. The blocking payoff being less than $v^{C(j^*)}_{j^*} - \epsilon$ can be re-written as

$$(1-\delta)(B - v^{C(j^*)}_{j^*}) + \epsilon \leq \delta(1-\delta^{L(\delta)})(v^{C(j^*)}_{j^*} - v_{j^*}(\underline{a}^\circ_{C(j^*)}))$$

As $\delta \to 1$, the LHS converges to ϵ. Because $\lim_{\delta\to 1} \delta^{L(\delta)} = \kappa$, the RHS converges to $(1-\kappa)(v^{C(j^*)}_{j^*} - v_{j^*}(\underline{a}^\circ_{C(j^*)}))$. By definition of ϵ, the above inequality holds, and therefore, no coalition can profitably block if δ is sufficiently high.

States of the form $\underline{w}(S,\tau)$: We first consider the case where $C \subseteq S$ and $C \neq \emptyset$. Choose an arbitrary $i \in C$ and we will show that the blocking is not profitable for i. By the definition $\underline{a}^\circ_S$ and due to the monotonicity of effectivity correspondences, coalition C cannot generate a payoff of more than $\underline{v}^\circ_i$ for player i, so i finds the blocking to be unprofitable if

$$(1-\delta^{L(\delta)-\tau})v_i(\underline{a}^\circ_S) + \delta^{L(\delta)-\tau}v^S_i \geq (1-\delta)\underline{v}^\circ_i + \delta(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + \delta^{L(\delta)+1}v^S_i. \tag{3}$$

Because $v^S_i > \underline{v}^\circ_i \geq v_i(\underline{a}^\circ_S)$, it suffices to show that

$$(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + \delta^{L(\delta)}v^S_i \geq (1-\delta)\underline{v}^\circ_i + \delta(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + \delta^{L(\delta)+1}v^S_i.$$

Re-arranging terms yields that $(1-\delta)(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + (1-\delta)\delta^{L(\delta)}v^S_i \geq (1-\delta)\underline{v}^\circ_i$, and then dividing by $(1-\delta)$ yields:

$$(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + \delta^{L(\delta)}v^S_i \geq \underline{v}^\circ_i.$$

Let us verify that this inequality holds for sufficiently high δ. Taking $\delta \to 1$ yields Inequality (1), which is true. Hence Inequality (3) holds for sufficiently high δ.

Next we consider the case where $C \nsubseteq S$. By construction, $j^* \notin S$. Player j^* finds blocking to be unprofitable if

$$(1-\delta^{L(\delta)-\tau})v_{j^*}(\underline{a}^\circ_S) + \delta^{L(\delta)-\tau}v^S_{j^*} \geq (1-\delta)B + \delta(1-\delta^{L(\delta)})v_{j^*}(\underline{a}^\circ_{C(j^*)}) + \delta^{L(\delta)+1}v^{C(j^*)}_{j^*}. \tag{4}$$

We prove that this inequality is satisfied if δ is sufficiently high. Examining the LHS, observe that for all τ such that $0 \le \tau \le L(\delta)-1$,

$$\begin{aligned}\lim_{\delta\to 1}\left[(1-\delta^{L(\delta)-\tau})v_{j^*}(\underline{a}^\circ_S)+\delta^{L(\delta)-\tau}v^S_{j^*}\right] &= \lim_{\delta\to 1}\left[\left(1-\frac{\kappa}{\delta^\tau}\right)v_{j^*}(\underline{a}^\circ_S)+\frac{\kappa}{\delta^\tau}v^S_{j^*}\right]\\ &=(1-\widetilde{\kappa})v_{j^*}(\underline{a}^\circ_S)+\widetilde{\kappa}v^S_{j^*}\end{aligned}$$

for some $\widetilde{\kappa}\in[\kappa,1]$. Examining the RHS of (4), observe that

$$\begin{aligned}&\lim_{\delta\to 1}\left[(1-\delta)B+\delta(1-\delta^{L(\delta)})v_{j^*}(\underline{a}^\circ_{C(j^*)})+\delta^{L(\delta)+1}v^{C(j^*)}_{j^*}\right]\\ &=\lim_{\delta\to 1}\left[(1-\delta^{L(\delta)})v_{j^*}(\underline{a}^\circ_{C(j^*)})+\delta^{L(\delta)}v^{C(j^*)}_{j^*}\right]\\ &=(1-\kappa)v_{j^*}(\underline{a}^\circ_{C(j^*)})+\kappa v^{C(j^*)}_{j^*}\ \le\ (1-\kappa)\underline{v}^\circ_{j^*}+\kappa v^{C(j^*)}_{j^*}\ \le\ (1-\widetilde{\kappa})\underline{v}^\circ_{j^*}+\widetilde{\kappa}v^{C(j^*)}_{j^*},\end{aligned}$$

where the first equality follows from taking limits, the second from $\lim_{\delta\to 1}\delta^{L(\delta)}=\kappa$, the first weak inequality follows from $v_{j^*}(\underline{a}^\circ_{C(j^*)})\le\underline{v}^\circ_{j^*}$, the second weak inequality follows from $\widetilde{\kappa}\ge\kappa$ and $\underline{v}^\circ_{j^*}<v^{C(j^*)}_{j^*}$. Since $\widetilde{\kappa}\in[\kappa,1]$ and $j^*\notin S$, (2) delivers that $(1-\widetilde{\kappa})v_{j^*}(\underline{a}^\circ_S)+\widetilde{\kappa}v^S_{j^*}$ is strictly higher than $(1-\widetilde{\kappa})\underline{v}^\circ_{j^*}+\widetilde{\kappa}v^{C(j^*)}_{j^*}$. This term guarantees that (4) holds for sufficiently high δ. $\square$

A.2 Proof of Theorem 2 on p. 12

Given that the stage game exhibits default-independent power, the set $v_C(E_C(a)\backslash\{a\})$ is independent of the recommended alternative a for each coalition C; we can therefore define $D(C):=v_C(E_C(a)\backslash\{a\})$ for some $a\in A$. To study stationary PCEs, we define the analogue of the self-generation map (Abreu, Pearce, and Stacchetti 1990). In a stationary PCE, at any history the prescribed current-period payoff and the prescribed continuation value are the same. Accordingly, we define the stationary self-generation map as follows. For any set $Y\subseteq\mathcal{V}\subseteq\mathbb{R}^n$, define

$$\Phi_\delta(Y):=\left\{y\in Y:\forall C\in\mathcal{C}\text{ and }z_C\in D(C),\ \exists y'\in Y\text{ and }i\in C\text{ s.t. }y_i\ge(1-\delta)z_i+\delta y'_i\right\}.$$

In words, the self-generation map identifies the set of discounted payoffs that are supportable when continuation payoffs must lie in the set Y. A discounted payoff y can be supported if (i) y can be decomposed into some current payoff and continuation payoff (which must both be equal to y due to stationarity), and (ii) for any coalition

C contemplating blocking the current-period prescription and implementing payoff $z_C \in D(C)$ for its members, there is an alternative continuation payoff y' that deters C from doing so.

Lemma 3. *If $Y \subseteq \mathcal{V}$ and $Y \subseteq \Phi_\delta(Y)$, then every $y \in Y$ can be supported by a stationary PCE.*

Proof. Consider any $y \in Y \subseteq \Phi_\delta(Y)$. We will construct a stationary PCE $\sigma : \bigcup_{\tau=0}^{\infty} \mathcal{H}^\tau \to A$ such that $U(\emptyset|\sigma) = y$.

Since $y \in \mathcal{V}$, there exists $\widetilde{a} \in A$ such that $v(\widetilde{a}) = y$. Define $\sigma(\emptyset) = \widetilde{a}$. We will extend σ's domain to $\bigcup_{\tau=0}^{\infty} \mathcal{H}^\tau$ while making sure that for all $\tau \geq 0, h^\tau \in H^\tau$, σ satisfies (i) stationarity: $\sigma(h^\tau, \sigma(h^\tau), \emptyset) = \sigma(h^\tau)$ and $v(\sigma(h^\tau)) \in Y$, and (ii) no profitable block: for all $C \in \mathcal{C}$ and $a' \in E_C(\sigma(h^\tau))$, there exists $i \in C$ such that

$$v_i(\sigma(h^\tau)) \geq (1-\delta)v_i(a') + \delta v_i(\sigma(h^\tau, a', C)).$$

Since $y \in \Phi_\delta(Y)$, we know that for all $C \in \mathcal{C}$, and $a \in E_C(\widetilde{a})$, there exists $y'(a, C) \in Y$ such that $y_i \geq (1-\delta)v_i(a) + \delta y'_i(a, C)$ for some $i \in C$. Furthermore since $y'(a, C) \in Y \subseteq \mathcal{V}$, this implies the existence of $a'(a, C) \in A$ such that $y'(a, C) = v(a'(a, C))$. We extend σ to the domain $\{\emptyset\} \cup \mathcal{H}^1$ as follows: $\sigma(\widetilde{a}, \emptyset) = \widetilde{a}$, and $\sigma(a, C) = a'(a, C)$ for all $C \in \mathcal{C}$ and $a \in E_C(\widetilde{a})$. Clearly, this satisfies properties (i) and (ii) for $\tau = 0$.

Now we complete the definition of the plan σ through induction on t. Fix $t > 1$, and assume we've defined the function $\sigma : \bigcup_{\tau=0}^{t-1} \mathcal{H}^\tau \to A$ satisfying properties (i) and (ii) for $\tau = 0, \ldots, t-1$ and all $h^\tau \in \mathcal{H}^\tau$. Consider any $h^{t-1} \in \mathcal{H}^{t-1}$. Since $v(\sigma(h^{t-1})) \in Y \subseteq \Phi_\delta(Y)$, we know that for all $C \in \mathcal{C}$, and $a \in E_C(\sigma(h^{t-1}))$, there exists $y^{h^{t-1}}(a, C) \in Y$ such that $v_i(\sigma(h^{t-1})) \geq (1-\delta)v_i(a) + \delta y_i^{h^{t-1}}(a, C)$ for some $i \in C$. In addition, since $y^{h^{t-1}}(a, C) \in Y \subseteq \mathcal{V}$, this implies the existence of $a^{h^{t-1}}(a, C)$ such that $y^{h^{t-1}}(a, C) = v(a^{h^{t-1}}(a, C))$. Extend σ to the domain $\mathcal{H}^t$ by defining $\sigma(h^{t-1}, a, C)$ as follows: $\sigma(h^{t-1}, \sigma(h^{t-1}), \emptyset) = \sigma(h^{t-1})$, and $\sigma(h^{t-1}, a, C) = a^{h^{t-1}}(a, C)$ for all $h^{t-1} \in \mathcal{H}^{t-1}$, $C \in \mathcal{C}$ and $a \in E_C(\sigma(h^{t-1}))$. Note that, by construction, the function satisfies properties (i) and (ii) for $\tau = 0, \ldots, t$ and $h^\tau \in \mathcal{H}^\tau$. This completes the induction step.

By property (i), σ is stationary, which means that at any history h^t it delivers the discounted payoff $U(h^t|\sigma) = v(\sigma(h^t))$. In particular, $U(\emptyset|\sigma) = y$, as required. Property (ii) then implies that σ is a PCE. □

Proof of Theorem 2. We show that any payoff that can be supported by a PCE can be supported by a stationary PCE (given that the converse holds by definition).

Take a PCE σ, and let $\mathcal{U}(\sigma) := \{U(h|\sigma) : h \in \mathcal{H}\}$ denote the set of continuation values associated with σ. Since $\mathcal{V}$ is convex, it follows that $\mathcal{U}(\sigma) \subseteq \mathcal{V}$. Given Lemma 3, it suffices to show that $\mathcal{U}(\sigma) \subseteq \Phi_\delta(\mathcal{U}(\sigma))$.

Consider any $y \in \mathcal{U}(\sigma)$ and let h^t be some t-history such that $U(h^t|\sigma) = y$. Since σ is a PCE, we know that for any $C \in \mathcal{C}$ and any $a' \in E_C(\sigma(h^t))$, there exists $y' \in \mathcal{U}(\sigma)$ and $i \in C$ such that $(1-\delta)v_i(a) + \delta y'_i \leq y_i$. Since the stage game exhibits default-independent power, for all $C \in \mathcal{C}$ and $z_C \in D(C)$, there exists $y' \in Y$ and $i \in C$ such that $y_i \geq (1-\delta)z_i + \delta y'_i$, which implies $\mathcal{U}(\sigma) \subseteq \Phi_\delta(\mathcal{U}(\sigma))$. □

A.3 Proof of Theorem 3 on p. 13

A Preliminary Result. Let $\mathcal{U}^S := \{u \in \mathbb{R}^n : u = U(\emptyset|\sigma),\ \sigma \text{ is a strongly symmetric PCE}\}$ denote the set of discounted payoff profiles from strongly symmetric PCEs. We first establish a lemma that will be useful in proving both directions of Theorem 3.

Lemma 4. *If $\mathcal{U}^S$ is nonempty, then $\mathcal{U}^S$ is the singleton set $\{\widehat{v}\}$.*

Proof. Suppose $\mathcal{U}^S$ is nonempty. Since players accrue identical payoffs from symmetric action profiles, we have $u_i = u_j$ for all players i, j and $u \in \mathcal{U}^S$. Let $\widehat{x} := \max_{a\in A^S} v_1(a)$ be the highest symmetric stage-game payoff, so $\widehat{v} = (\widehat{x}, \ldots, \widehat{x})$. Define $\underline{x} := \inf\{x : (x, \ldots, x) \in \mathcal{U}^S\}$.

Consider a sequence $\{(x^k, \ldots, x^k)\}_{k=1}^\infty \subseteq \mathcal{U}^S$ that converges to $(\underline{x}, \ldots, \underline{x})$ and let σ^k be the PCE that supports the discounted payoff profile $(x^k, \ldots, x^k)$. As a PCE, σ^k cannot be profitably blocked by the grand coalition N choosing $\widehat{a} \in \arg\max_{a\in A^S} v_1(a)$, which would generate the stage-game payoff profile $(\widehat{x}, \ldots, \widehat{x})$. So for each k we have $x^k \geq (1-\delta)\widehat{x} + \delta \underline{x}$. Since $x^k \to \underline{x}$, it follows that

$$\underline{x} \geq \widehat{x}. \tag{5}$$

Take a payoff profile $(\widetilde{x}, \ldots, \widetilde{x}) \in \mathcal{U}^S$. The definitions of $\widehat{x}$ and $\underline{x}$ imply that $\underline{x} \leq \widetilde{x} \leq \widehat{x}$. Combining these inequalities with (5) shows that $\widetilde{x} = \widehat{x}$, so $\mathcal{U}^S = \{\widehat{v}\}$. □

Proof of Theorem 3. For the "only if" direction: Suppose there exists a strongly symmetric PCE σ. By Lemma 4, σ's continuation values satisfy $U(h|\sigma) = \widehat{v}$ for all $h \in \mathcal{H}$. Since $\widehat{v}$ is the maximal feasible payoff from symmetric action profiles, and σ must prescribe symmetric action profiles after every history, it follows that σ's recommended action profile $\sigma(h)$ must satisfy $u(\sigma(h)) = \widehat{v}$ for all $h \in \mathcal{H}$.

Take an arbitrary $\widehat{h} \in \mathcal{H}$ and let $\widehat{a} := \sigma(\widehat{h})$ be the recommended action profile. As argued above $\widehat{a}$ is symmetric and $u(\widehat{a}) = \widehat{v}$. It remains to show that $\widehat{a}$ is a core alternative. Suppose $\widehat{a}$ is not a core alternative, then there exists coalition C and $a' \in E_C(\widehat{a})$ such that $v_i(a') > v_i(\widehat{a})$ for all $i \in C$. However this implies

$$\begin{aligned} U_i(\widehat{h}|\sigma) &= (1-\delta)v_i(\widehat{a}) + \delta U_i(\widehat{h}, \widehat{a}, \emptyset) = (1-\delta)v_i(\widehat{a}) + \delta \widehat{v}_i \\ &< (1-\delta)v_i(a') + \delta \widehat{v}_i = (1-\delta)v_i(a') + \delta U_i(\widehat{h}, a', C|\sigma) \end{aligned}$$

for all $i \in C$, which contradicts σ being a PCE.

For the remainder of the theorem, suppose $\widehat{v} = v(\widehat{a})$ for some symmetric core alternative $\widehat{a}$. A plan that recommends $\widehat{a}$ at all histories is a strongly symmetric PCE that supports the discounted payoff $\widehat{v}$. Uniqueness follows from Lemma 4. $\square$

A.4 Proof of Theorem 5 on p. 17

Preliminary Results. To prove our claim, we first introduce the transferable-utility analogue of the concept of a blocking plan, as defined in Appendix A.1.

A *(transferable-utility) blocking plan* by coalition C from a plan σ is a pair (α, β), where $\alpha : \overline{\mathcal{H}} \to A$ and $\beta : \overline{\mathcal{H}} \to \mathcal{T}$ satisfy $\alpha(h) \in E_C(a(h|\sigma))$ and $\beta_{-C}(h) = \chi^C(T(h|\sigma))$ for every history $h \in \overline{\mathcal{H}}$. After each history, the blocking plan (α, β) generates a path

$$\Big(\alpha(h), C, \beta(h),\ \alpha\big(h, \alpha(h), C, \beta(h)\big), C, \beta\big(h, \alpha(h), C, \beta(h)\big), \dots\Big)$$

that is distinct from the one generated by σ. We will use $U_i(h|\alpha, \beta)$ to denote player i's discounted payoff from that path. The blocking plan (α, β) is profitable if there exists a history h such that $U_i(h|\alpha, \beta) > U_i(h|\sigma)$ for all $i \in C$.

Lemma 5. *If a plan σ is a public PCE, then no coalition $C \in \mathcal{S} \cup N$ has a profitable blocking plan.*

Proof. Consider a public plan σ from which coalition $C \in \mathcal{S} \cup N$ has a profitable blocking plan (α, β). In particular, there exists a history $h \in \mathcal{H}$ such that $U_i(h|\alpha, \beta) > U_i(h|\sigma)$ for every $i \in C$. We will show that this implies that coalition C has a profitable block from the plan σ, so σ is not a PCE.

By Assumption 2, the plan σ has bounded continuation values. Moreover, as proven in Lemma 7 in the Supplementary Appendix, it is without loss to assume that the blocking plan (α, β) also has bounded continuation values. Treating coalition C a fictitious player whose payoff is the sum of those of its members, we can see that C faces a decision tree with bounded values. Applying the standard one-shot deviation principle to the fictitious player C yields the existence of $\widehat{h} \in \overline{\mathcal{H}}$ such that such that

$$(1-\delta)\sum_{i\in C} u_i\Big(\alpha(\widehat{h}),\ \beta(\widehat{h})\Big) + \delta \sum_{i\in C} U_i\Big(\widehat{h},\ \alpha(\widehat{h}),\ C,\ \beta(\widehat{h})\Big|\sigma\Big) > \sum_{i\in C} U_i(\widehat{h}|\sigma).$$

To show that C can profitably block σ at $\widehat{h}$ amounts to showing that this total payoff can be divided so that every individual member can be made better off.

Let T^* be the transfers matrix such that for all $(j,k) \notin C \times C$, $T^*_{jk} = \beta_{jk}(\widehat{h})$; but for $(j,k) \in C \times C$, T^*_{jk} satisfies for every $i \in C$,

$$(1-\delta)u_i\Big(\alpha(\widehat{h}),\ T^*\Big) + \delta U_i\Big(\widehat{h},\ \alpha(\widehat{h}),\ C,\ \beta(\widehat{h})\Big|\sigma\Big) > U_i(\widehat{h}|\sigma). \tag{6}$$

Consider the two histories

$$h^1 := \Big(\widehat{h},\ \alpha(\widehat{h}),\ C,\ \beta(\widehat{h})\Big) \text{ and } h^2 := \Big(\widehat{h},\ a(\widehat{h}|\sigma'),\ C,\ T^*\Big).$$

By the construction of T^* and the fact that $C \in \mathcal{S} \cup N$, h^1 and h^2 share the same public component $h^1_p = h^2_p$. Since the plan σ is public, it follows that for all $i \in N$,

$$U_i\Big(\widehat{h},\ \alpha(\widehat{h}),\ C,\ \beta(\widehat{h})\Big|\sigma\Big) = U_i\Big(\widehat{h},\ \alpha(\widehat{h}),\ C,\ T^*\Big|\sigma\Big).$$

Inequality (6) can therefore be re-written as, for every $i \in C$,

$$(1-\delta)u_i\Big(\alpha(\widehat{h}),\ T^*\Big) + \delta U_i\Big(\widehat{h},\ \alpha(\widehat{h}),\ C,\ T^*\Big|\sigma\Big) > U_i(\widehat{h}|\sigma). \tag{7}$$

Inequality (7) implies that σ is not a PCE. □

The next result shows that for any payoff profile in $\mathcal{U}_{CR}(\mathcal{S})$, we can construct "$(\mathcal{S} \cup N)$-specific punishments" for all coalitions in $\mathcal{S} \cup N$.

Lemma 6. *For any $u^* \in \mathcal{U}_{CR}(\mathcal{S})$, there exist $(\mathcal{S} \cup N)$-specific punishments $\{u^C : C \in \mathcal{S} \cup N\} \subseteq \mathcal{U}_{CR}(\mathcal{S})$ such that*

$$\sum_{i\in C} u^C_i < \sum_{i\in C} u^*_i \tag{8}$$

for all $C \in \mathcal{S} \cup N$, and

$$\sum_{i \in C} u_i^C < \sum_{i \in C} u_i^{C'} \tag{9}$$

for all $C, C' \in \mathcal{S} \cup N, C' \neq C$.

Proof. For any coalition $C \in \mathcal{S} \cup N$, consider the vector u^C defined by

$$u_i^C = \begin{cases} u_i^* - \frac{\epsilon}{|C|} & i \in C \\ u_i^* + \frac{\epsilon}{|N \backslash C|} & i \notin C \end{cases}$$

Compared to the payoff vector u^*, in u^C every player in C is charged equally, with a total summing up to ϵ; by contrast, players outside of C are paid equally, with a total of amount also summing up to ϵ. The ϵ may be set sufficiently small to ensure all u^C's are in $\mathcal{U}_{CR}(\mathcal{S})$.

We show that these vectors satisfy inequalities (8) and (9). By construction, $\sum_{i \in C} u_i^C = \sum_{i \in C} u_i^* - \epsilon < \sum_{i \in C} u_i^*$, so (8) is satisfied. To verify (9), consider two coalitions $C, C' \in \mathcal{S} \cup N$ with $C \neq C'$. Coalition C can be partitioned as the union of two components $C = (C \backslash C') \cup (C \cap C')$. So

$$\begin{aligned} \sum_{i \in C} u_i^{C'} &= \sum_{i \in C \backslash C'} u_i^{C'} + \sum_{i \in C \cap C'} u_i^{C'} \\ &= \left[\sum_{i \in C \backslash C'} u_i^* + \frac{|C \backslash C'|}{|N \backslash C'|} \epsilon \right] + \left[\sum_{i \in C \cap C'} u_i^* - \frac{|C \cap C'|}{|C'|} \epsilon \right] \qquad (10) \\ &= \sum_{i \in C} u_i^* - \left[\frac{|C \cap C'|}{|C'|} - \frac{|C \backslash C'|}{|N \backslash C'|} \right] \epsilon \\ &> \sum_{i \in C} u_i^* - \epsilon \qquad (11) \\ &= \sum_{i \in C} u_i^C \end{aligned}$$

Equality (10) follows since compared to u^*, $u^{C'}$ gives every player outside of C' an extra payoff of $\frac{\epsilon}{|N \backslash C|}$, while lowering the payoff of every player inside C' by $\frac{\epsilon}{|C'|}$. Since $C \neq C'$, either $C \backslash C' \neq \emptyset$ or $C \cap C' \neq C'$ must be true; in other words, either $\frac{|C \backslash C'|}{|N \backslash C'|} > 0$ or $\frac{|C \cap C'|}{|C'|} < 1$. In either cases, (11) follows, which verifies (9). $\square$

Proof of Theorem 5.

We first show that for every $\delta \geq 0$, every public PCE gives each coalition $C \in \mathcal{S} \cup N$ a total payoff of at least $\underline{u}_C$. In fact, we prove a stronger statement: every public PCE σ guarantees that for every coalition $C \in \mathcal{S} \cup N$ and every history $h \in \overline{\mathcal{H}}$,

$$\sum_{i \in C} U_i(h|\sigma) \geq \underline{u}_C. \tag{12}$$

Consider a public plan σ such that there exists a coalition C and history $\widehat{h}$ such that $\sum_{i \in C} U_i(\widehat{h}|\sigma) < \underline{u}_C$. We prove that σ must not be a PCE.

To this end, we construct a profitable blocking plan from σ for coalition C. The plan σ recommends an alternative $a(h|\sigma)$ at every history $h \in \overline{\mathcal{H}}$. At every history $h \in \overline{\mathcal{H}}$, let $\alpha(h) \in \arg\max_{a \in E_C(a(h|\sigma))} \sum_{i \in C} v_i(a)$ be an alternative in coalition C's "best response" to the recommended alternative. By the definition of $\underline{u}_C$, it follows that $\sum_{i \in C} v_i(\alpha(h)) \geq \underline{u}_C > \sum_{i \in C} U_i(\widehat{h}|\sigma)$. Since coalition C's total generated payoff from $\alpha(h)$, $\sum_{i \in C} v_i(\alpha(h))$, is higher than $\sum_{i \in C} U_i(\widehat{h}|\sigma)$, we can find transfers among players in C such that the payoff of each individual player $i \in C$ is higher than $U_i(\widehat{h}|\sigma)$. Formally, at every history $h \in \overline{\mathcal{H}}$, there exist transfers $\widetilde{T}_C(h) := [\widetilde{T}_{ij}(h)]_{i \in C, j \in N}$ such that $\widetilde{T}_{ij}(h) = 0$ for all $j \in N \backslash C$, and

$$v_i(\alpha(h)) + \sum_{j \in C} \widetilde{T}_{ji}(h) - \sum_{j \in C} \widetilde{T}_{ij}(h) > U_i(\widehat{h}|\sigma)$$

for all $i \in C$. As a result, for each player $i \in C$, the experienced payoff from the stage-game outcome satisfies

$$\begin{aligned}
&u_i\Big(\alpha(h),\ C,\ \big[\widetilde{T}_C(h), \chi^C\big(\,T(h|\sigma)\,\big)\big]\Big) \\
=&v_i(\alpha(h)) + \sum_{j \in C} \widetilde{T}_{ji}(h) + \sum_{j \in N \backslash C} \chi^C_{ji}\big(\,T(h|\sigma)\,\big) - \sum_{j \in N} \widetilde{T}_{ij}(h) \\
\geq& \, v_i(\alpha(h)) + \sum_{j \in C} \widetilde{T}_{ji}(h) - \sum_{j \in C} \widetilde{T}_{ij}(h) \\
>& \, U_i(\widehat{h}|\sigma),
\end{aligned}$$

where the weak inequality follows because $\chi^C_{ji}(\,T(h|\sigma)\,) \geq 0$ for all $j \in N$, and $\widetilde{T}_{ij}(h) = 0$ for all $j \in N \backslash C$. Observe that the LHS holds at every history, including $\widehat{h}$ and those that follow. These steps prove that the blocking plan (α, β) by coalition C, where $\beta(h) := \big[\widetilde{T}_C(h), \chi^C(\,T(h|\sigma))\big]$ for every history $h \in \overline{\mathcal{H}}$, is profitable: $U_i(\widehat{h}|\sigma') > U_i(\widehat{h}|\sigma)$

for every $i \in C$. Lemma 5 then implies that σ is not a PCE.

We now show that for every payoff profile $u \in \mathcal{U}_{CR}(\mathcal{S})$, there is a $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a public PCE with a discounted payoff equal to u. For every $C \in \mathcal{S} \cup N$, let $\underline{a}_C \in \arg\min_{a \in A} \max_{a' \in E_C(a)} \sum_{i \in C} v_i(a')$ be an alternative that can be used to minmax C. Note that by construction, $\sum_{i \in C} v_i(\underline{a}_C) \leq \underline{u}_C$.

Fix any payoff vector $u^* \in \mathcal{U}_{CR}(\mathcal{S})$, and let $\{u^C : C \in \mathcal{S} \cup N\}$ be the $(\mathcal{S} \cup N)$-specific punishments from Lemma 6. Given these $(\mathcal{S} \cup N)$-specific punishments, let $\kappa \in (0, 1)$ be such that for every $\widetilde{\kappa} \in [\kappa, 1]$, the following is true for every $C \in \mathcal{S} \cup N$:

$$(1 - \widetilde{\kappa}) \sum_{i \in C} v_i(\underline{a}_C) + \widetilde{\kappa} \sum_{i \in C} u_i^C > \underline{u}_C \tag{13}$$

$$\forall C' \in \mathcal{S} \cup N \backslash \{C\}\colon\ (1 - \widetilde{\kappa}) \sum_{i \in C'} v_i(\underline{a}_C) + \widetilde{\kappa} \sum_{i \in C'} u_i^C > (1 - \widetilde{\kappa}) \sum_{i \in C'} v_i(\underline{a}_{C'}) + \widetilde{\kappa} \sum_{i \in C'} u_i^{C'}. \tag{14}$$

Inequality (13) implies that in terms of total value, coalition C is willing to bear the cost of $\sum_{i \in C} v_i(\underline{a}_C)$ with the promise of transitioning into its coalition-specific punishment rather than staying at its minmax. Inequality (14) implies that every coalition in $\mathcal{S} \cup N$ prefers punishing other coalitions in $\mathcal{S} \cup N$ than being punished itself. By an argument identical to what we saw in the proof of Theorem 1, there exists a value of $\kappa \in (0, 1)$ such that all the inequalities above hold for all $\widetilde{\kappa} \in [\kappa, 1]$, $C \in \mathcal{S} \cup N$ and $C' \in \mathcal{S} \cup N \backslash \{C\}$. Let $L(\delta) := \left\lceil \frac{\log \kappa}{\log \delta} \right\rceil$ where $\lceil \cdot \rceil$ is the ceiling function. As before, we use the property that $\lim_{\delta \to 1} \delta^{L(\delta)} = \kappa$.

For each alternative $a \in A$ let $\mathcal{U}(a) := \{u \in \mathbb{R}^n : \sum_i u_i = \sum_i v_i(a)\}$ denote the set of payoff profiles that can be generated by playing alternative a and redistributing through transfers. Let $\overline{a} \in \arg\max_{a \in A} \sum_{i \in N} v_i(a)$ and $\underline{a} \in \arg\min_{a \in A} \sum_{i \in N} v_i(a)$ be two alternatives that maximize and minimize players' total generated payoffs, respectively. Since $\mathcal{U}_{CR}(\mathcal{S}) \subseteq \mathcal{U}_{IR}$, by Lemma 8 in the Supplementary Appendix, there exist payoff vectors $\{\widetilde{u}^1, \ldots, \widetilde{u}^M\} \subseteq \mathcal{U}(\overline{a}) \cup \mathcal{U}(\underline{a})$ such that $\mathcal{U}_{CR}(\mathcal{S}) \subseteq \text{co}(\widetilde{u}^1, \ldots, \widetilde{u}^M)$, where $\widetilde{u}^m = u(\widetilde{a}^m, \widetilde{T}^m)$ for some alternative $\widetilde{a}^m \in \{\overline{a}, \underline{a}\}$ and transfers matrix $\widetilde{T}^m$ for each $m = 1, \ldots, M$. Lemma 1 then guarantees that for any $\epsilon > 0$, there exists $\underline{\delta} \in (0, 1)$ such that for all $\delta \in (\underline{\delta}, 1)$, there exist sequences $\big\{\{a^{d,\tau}, T^{d,\tau}\}_{\tau=0}^{\infty} : d \in \mathcal{S} \cup N \cup \{*\}\big\}$ such that for each d and t, $(1 - \delta) \sum_{\tau=0}^{\infty} \delta^{\tau} u(a^{d,\tau}, T^{d,\tau}) = u^d$ and

$\left|\left|u^d - (1-\delta)\sum_{\tau=t}^{\infty}\delta^\tau u(a^{d,\tau}, T^{d,\tau})\right|\right| < \epsilon$. We fix an ϵ such that

$$\epsilon < (1-\kappa)\min\left\{\min_{d\in\mathcal{S}\cup N, d'\in\mathcal{S}\cup N\cup\{*\}, d'\neq d}\Big(\sum_{i\in d} u_i^{d'} - \sum_{i\in d} u_i^d\Big), \min_{d\in\mathcal{S}\cup N}\sum_{i\in d} u_d^d - \underline{v}_d\right\},$$

and given that ϵ, consider δ exceeding the appropriate $\underline{\delta}$.

Let $\widetilde{\mathcal{T}} := \{\widetilde{T}^m\}_{m=1}^M$ be the set of transfer matrices defined above. By Assumption 1, when the plan recommends a transfers matrix $T \in \widetilde{\mathcal{T}}$ and a coalition $C \in \mathcal{C}$ blocks, there exists $\widetilde{B} > 0$ such that

$$\sum_{i\notin C, j\in C}\chi_{ij}^C(T) \leq \widetilde{B} \text{ for all } T \in \widetilde{\mathcal{T}} \text{ and } C \in \mathcal{C}. \tag{15}$$

In addition, since A is compact and $v(.)$ is continuous, there exists $\widehat{B}$ such that

$$\max_{C\in\mathcal{C}}\max_{a\in A}\sum_{i\in C} v_i(a) \leq \widehat{B}. \tag{16}$$

We now describe the public plan that we use to support u^*. Let $\mathbf{0}$ denote the transfer matrix where all players make no transfers. Consider the plan represented by the automaton $(W, w(0,0), f, \gamma)$, where

- $W := \big\{w(d,\tau) | d \in \mathcal{S}\cup N\cup\{*\}, \tau \geq 0\big\} \cup \{\underline{w}(S,\tau) \,|\, S \in \mathcal{S}\cup N, 0 \leq \tau < L(\delta)\}$ is the set of possible states and $w(*,0)$ is the initial state;
- $f : W \to \overline{\mathcal{O}}$ is the output function, where $f(w(d,\tau)) = (a^{d,\tau}, \emptyset, T^{d,\tau})$ and $f(\underline{w}(S,\tau)) = (\underline{a}_S, \emptyset, \mathbf{0})$;
- $\gamma : W \times \overline{\mathcal{O}} \to W$ is the transition function. For states of the form $w(d,\tau)$, the transition is

$$\gamma\big(w(d,\tau),(a,C,T)\big) = \begin{cases} \underline{w}(C,0) & \text{if } C \in \mathcal{S}\cup N \\ \underline{w}(j^*,0) & \text{if } C \in \mathcal{C}\backslash(\mathcal{S}\cup N),\ j^* \in \arg\min_{j\in C} u_j(a,T) \\ w(d,\tau+1) & \text{if } C = \emptyset \end{cases}$$

For states of the form $\{\underline{w}(i,\tau)|0 \leq \tau < L(\delta)-1\}$ where $i \in N$, the transition is

$$\gamma(\underline{w}(i,\tau),(a,C,T)) = \begin{cases} \underline{w}(C,0) & \text{if } C \in \mathcal{S} \cup N \\ \underline{w}(j^*,0) & \text{if } \{C \in \mathcal{C}\backslash(\mathcal{S}\cup N)\} \cap \big(\{u_i(a,T) > \underline{v}_i\} \cup \{i \notin C\}\big) \\ & j^* \in \arg\min_{C\backslash\{i\}} u_j(a,T) \\ \underline{w}(i,0) & \text{if } \{C \in \mathcal{C}\backslash(\mathcal{S}\cup N)\} \cap \{u_i(a,T) \leq \underline{v}_i\} \cap \{i \in C\} \\ \underline{w}(C,\tau+1) & \text{if } C = \emptyset \end{cases}$$

For states of the form $\underline{w}(i, L(\delta)-1)$ where $i \in N$, the transition is

$$\gamma(\underline{w}(i,L(\delta){-}1),(a,C,T)) = \begin{cases} \underline{w}(C,0) & \text{if } C \in \mathcal{S} \cup N \\ \underline{w}(j^*,0) & \text{if } \{C \in \mathcal{C}\backslash(\mathcal{S}\cup N)\} \cap \big(\{u_i(a,T) > \underline{v}_i\} \cup \{i \notin C\}\big) \\ & j^* \in \arg\min_{C\backslash\{i\}} u_j(a,T) \\ \underline{w}(i,0) & \text{if } \{C \in \mathcal{C}\backslash(\mathcal{S}\cup N)\} \cap \{u_i(a,T) \leq \underline{v}_i\} \cap \{i \in C\} \\ w(i,0) & \text{if } C = \emptyset \end{cases}$$

For states of the form $\{\underline{w}(S,\tau)|0 \leq \tau < L(\delta)-1\}$ where $S \in \mathcal{S}$, the transition is

$$\gamma(\underline{w}(S,\tau),(a,C,T)) = \begin{cases} \underline{w}(C,0) & \text{if } C \in \mathcal{S} \cup N \\ \underline{w}(j^*,0) & \text{if } C \in \mathcal{C}\backslash(\mathcal{S}\cup N),\ j^* \in \arg\min_{j\in C} u_j(a,T) \\ \underline{w}(S,\tau+1) & \text{if } C = \emptyset \end{cases}$$

For states of the form $\underline{w}(S, L(\delta)-1)$ where $S \in \mathcal{S}$, the transition is

$$\gamma(\underline{w}(S,L(\delta){-}1),(a,C,T)) = \begin{cases} \underline{w}(C,0) & \text{if } C \in \mathcal{S} \cup N \\ \underline{w}(j^*,0) & \text{if } C \in \mathcal{C}\backslash(\mathcal{S}\cup N),\ j^* \in \arg\min_{j\in C} u_j(a,T) \\ w(S,0) & \text{if } C = \emptyset \end{cases}$$

The plan represented by this automaton yields payoff profile u^* and is public. By construction, for all τ in $\{0,\ldots,L(\delta-1)\}$, $\big|\big|u^d - V(w(d,\tau))\big|\big| < \epsilon$ and $V(\underline{w}(S,\tau)) = (1-\delta^{L(\delta)-\tau})v(\underline{a}_S) + \delta^{L(\delta)-\tau}V(w(S,0))$ and $S \in \mathcal{S} \cup N$. As the arguments from here on are standard, we verify in the Supplementary Appendix that no coalition can profitably block in any state of this automaton. $\square$

B Supplementary Appendix

The Supplementary Appendix contains the proof of Theorem 4, completes the proof of Theorem 5, the proofs of all propositions, and considers the extension in which multiple coalitions may simultaneously block.

B.1 Preliminary Result

Below, we list a preliminary result used in our proofs of Theorems 4 and 5

Lemma 7. *Let σ be a PCE. Suppose coalition C has a blocking plan (α, β) such that $\sum_{i\in C} U_i(\overline{h}|\alpha, \beta) > \sum_{i\in C} U_i(\overline{h}|\sigma)$ for some $\overline{h} \in \overline{\mathcal{H}}$, then C has a blocking plan (α', β') such that $\sum_{i\in C} U_i(\overline{h}|\alpha', \beta') > \sum_{i\in C} U_i(\overline{h}|\sigma)$, and the set $\{\sum_{i\in C} U_i(h|\alpha', \beta') : h \in \overline{\mathcal{H}}\}$ is bounded.*

Proof. We break this argument into two parts.

Part 1: We show that the set $\{\sum_{i\in C} U_i(h|\alpha, \beta) : h \in \overline{\mathcal{H}}\}$ is bounded from above. To this end, it suffices to show that the set of stage-game payoffs from the blocking plan, $\{\sum_{i\in C} u_i(\alpha(h), \beta(h)) : h \in \overline{\mathcal{H}}\}$ is bounded from above.

Consider an arbitrary coalition $C \in \mathcal{C}$ and an arbitrary history $h \in \overline{\mathcal{H}}$. Let $\widetilde{a} = a(h|\sigma)$ denote the alternative recommended by σ and $\widetilde{T} = T(h|\sigma)$ denote the recommended transfers. By the definition of a blocking plan, $\alpha(h) \in E_C(\widetilde{a})$ and $\beta(h) = (T'_C, \chi^C(\widetilde{T}))$ for some $T'_C \in \mathcal{T}_C$. Since the transfers T'_C may involve nonzero transfers to players outside of C, we have

$$\begin{aligned}\sum_{i\in C} u_i(\alpha(h), \beta(h)) =& \sum_{i\in C} v_i(\widetilde{a}) + \sum_{i\in C, j\notin C} \chi^C_{ji}(\widetilde{T}) - \sum_{i\in C, j\notin C} T'_{ij} \\ \leq& \sum_{i\in C} v_i(\widetilde{a}) + \sum_{i\in C, j\notin C} \chi^C_{ji}(\widetilde{T}) \qquad (17)\end{aligned}$$

Now suppose the coalition C blocks at history h and chooses alternative $\alpha(h) \in E_C(\widetilde{a})$; however, instead of $\beta(h) = (T'_C, \chi^C(\widetilde{T}))$, C chooses transfers $(T''_C, \chi^C(\widetilde{T}))$, where the transfers T''_C are such that members of C make zero payment to players outside of C while splitting the total payoff within C evenly. If C carries out this block, each member $i \in C$ obtains a discounted utility of at least

$$(1-\delta)\frac{1}{|C|}\Big[\sum_{i\in C} v_i(\alpha(h)) + \sum_{i\in C, j\notin C} \chi^C_{ji}(\widetilde{T})\Big] + \delta \inf_{h\in\overline{\mathcal{H}}, i\in N} U_i(h|\sigma),$$

whereas adhering to σ at h yields each member at most $\sup_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma)$. Since σ is a PCE, $\big(\alpha(h), T''_C, \chi^C(\widetilde{T})\big)$ cannot be a profitable block for C, so it must be true that

$$(1-\delta)\frac{1}{|C|}\Big[\sum_{i\in C} v_i(\widetilde{a}) + \sum_{i\in C, j\notin C} \chi^C_{ji}(\widetilde{T})\Big] + \delta \inf_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma) \le \sup_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma).$$

Combining the inequality above with (17) yields

$$(1-\delta)\frac{1}{|C|}\Big[\sum_{i\in C} u_i(\alpha(h),\beta(h))\Big] + \delta \inf_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma) \le \sup_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma).$$

Rearranging terms, we have

$$\begin{aligned}\sum_{i\in C} u_i(\alpha(h),\beta(h)) &\le \frac{|C|}{1-\delta}\left[\sup_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma) - \delta \inf_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma)\right] \\ &\le \frac{|C|}{1-\delta}\left|\sup_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma)\right| + \frac{|C|\delta}{1-\delta}\left|\inf_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma)\right|.\end{aligned}$$

Since $\{U(h|\sigma) : h\in\overline{\mathcal{H}}\}$ is bounded by Assumption 2, there exists $L>0$ such that

$$\left|\sup_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma)\right| \le L \quad \text{and} \quad \left|\inf_{h\in\overline{\mathcal{H}},i\in N} U_i(h|\sigma)\right| \le L.$$

Therefore,

$$\sum_{i\in C} u_i(\alpha(h),\beta(h)) \le \frac{1+\delta}{1-\delta}\,|C|\,L.$$

Note that the inequality above holds for all $h\in\overline{\mathcal{H}}$ while the right hand side does not depend on h, so our claim follows.

Part 2: It is without loss to assume that $\{\sum_{i\in C} U_i(h|\alpha,\beta) : h\in\overline{\mathcal{H}}\}$ is bounded from below. If not, we can construct another blocking plan (α',β') such that $\sum_{i\in C} U_i(\overline{h}|\alpha',\beta') > \sum_{i\in C} U_i(\overline{h}|\sigma)$ while ensuring $\{\sum_{i\in C} U_i(h|\alpha',\beta') : h\in\overline{\mathcal{H}}\}$ is bounded from below: if $\sum_{i\in C} U_i(\widehat{h}|\alpha,\beta)$ falls below $\min_{a\in A}\sum_{i\in C} v_i(a)$ for some history $\widehat{h}\in\overline{\mathcal{H}}$, we will ask C to refuse all outgoing transfers at all histories following $\widehat{h}$.

Formally, for a history $\widehat{h}\in\overline{\mathcal{H}}$, let $F(\widehat{h}) := \{h\widehat{h} : h\in\overline{\mathcal{H}}\}$ denote the set of histories that can follow from $\widehat{h}$. Let $\underline{H}_C := \{h\in\overline{\mathcal{H}} : \sum_{i\in C} U_i(h|\alpha,\beta) < \min_{a\in A}\sum_{i\in C} v_i(a)\}$. Let $\mathbf{0}_C$ denote the vector of zero-valued transfers made from players in C. Set $\alpha'=\alpha$, and define

$$\beta'(h) = \begin{cases} \left(\mathbf{0}_C, \chi^C(T(h|\sigma))\right) & \forall h \in F(\widehat{h}) \text{ for some } \widehat{h} \in \underline{H}_C, \\ \beta(h) & \text{otherwise.} \end{cases}$$

By construction, the blocking plan (α', β') has continuation values bounded below by $\min_{a \in A} \sum_{i \in C} v_i(a)$. In addition, compared to (α, β), the blocking plan (α', β') gives coalition C weakly higher total continuation value following any history, so $\sum_{i \in C} U_i(\overline{h}|\alpha', \beta') > \sum_{i \in C} U_i(\overline{h}|\sigma)$. □

Next we argue that there exists a finite set of payoff vectors whose convex hull contains the set $\mathcal{U}_{IR}$.

Lemma 8. *For each alternative $a \in A$ let $\mathcal{U}(a) := \{u \in \mathbb{R}^n : \sum_i u_i = \sum_i v_i(a)\}$ denote the set of payoff profiles that can be generated by playing alternative a and redistributing through transfers. Let $\overline{a} \in \arg\max_{a \in A} \sum_{i \in N} v_i(a)$ and $\underline{a} \in \arg\min_{a \in A} \sum_{i \in N} v_i(a)$ be two alternatives that maximize and minimize players' total generated payoffs, respectively. There exist payoff vectors $\{\widetilde{u}^1, \ldots, \widetilde{u}^M\} \subseteq \mathcal{U}(\overline{a}) \cup \mathcal{U}(\underline{a})$, such that $\mathcal{U}_{IR} \subseteq co(\widetilde{u}^1, \ldots, \widetilde{u}^M)$.*

Proof. By definition,

$$\mathcal{U}_{IR} \subseteq \overline{\mathcal{U}}_{IR} := \left\{ u \in \mathbb{R}^n : \sum_{i \in N} v_i(\underline{a}) \leq \sum_{i \in N} u_i \leq \sum_{i \in N} v_i(\overline{a}) \text{ and } u_i \geq \underline{v}_i \forall i \in N \right\}.$$

Since $\overline{\mathcal{U}}_{IR}$ is a bounded polyhedron, it is also a polytope. Let $x^1, \ldots, x^K$ be its vertices. Any point inside $\mathcal{U}_{IR}$ can then be expressed as convex combinations of these vertices. Since $x^k \in \text{co}(\mathcal{U}(\overline{a}) \cup \mathcal{U}(\underline{a}))$ for all $1 \leq k \leq K$, for each k, there exist $\{\widetilde{u}^{k,1}, \ldots, \widetilde{u}^{k,m_k}\} \subseteq \mathcal{U}(\overline{a}) \cup \mathcal{U}(\underline{a})$ such that $x^k \subseteq \text{co}(\widetilde{u}^{k,1}, \ldots, \widetilde{u}^{k,m_k})$. As a result $\mathcal{U}_{IR} \subseteq \text{co}(\cup_{1 \leq k \leq K} \{\widetilde{u}^{k,1}, \ldots, \widetilde{u}^{k,m_k}\})$. □

B.2 Proof of Theorem 4 on p. 15

<u>Part 1</u>: *For every $\delta \geq 0$, every PCE gives each player i a payoff of at least $\underline{v}_i$.*

Suppose σ is a PCE but $U_i(\emptyset|\sigma) < \underline{v}_i$ for some player i. To demonstrate that this creates a contradiction, we first show that player i has a profitable blocking plan from σ. Consider the following blocking plan: in every period, player i blocks and best-responds to the recommended alternative from the plan, and refuses to make any outgoing transfers. Formally, this is a blocking plan (α, β) defined by

$$\alpha(h) \in \underset{a' \in E_{\{i\}}(a(h|\sigma))}{\arg\max} \; v_i(a') \quad \text{and} \quad \beta(h) = \big(\mathbf{0}_i, \chi^{\{i\}}(T(h|\sigma))\big) \quad \text{for all } h \in \overline{\mathcal{H}}.$$

By the definition of $\underline{v}_i$, this blocking plan gives i at least $\underline{v}_i$ after every history, so $U_i(\emptyset|\alpha, \beta) > U_i(\emptyset|\sigma)$ and therefore (α, β) is a profitable blocking plan.

By Assumption 2, the plan σ has bounded continuation values. Moreover, as proven in Lemma 7 in the Supplementary Appendix, it is without loss to assume that the blocking plan (α, β) also has bounded continuation values. As a result, player i faces a decision tree with bounded values and we can apply the standard one-shot deviation principle to prove that there exists a profitable one-shot deviation for i. This implies that $\{i\}$ can profitably block σ at some history, contradicting σ being a PCE.

Part 2: *For every $u \in \mathcal{U}_{IR}$, there is a $\underline{\delta} < 1$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a PCE with a discounted payoff equal to u.*

The proof mirrors that of the same part in Theorem 1, and so we elaborate on the necessary changes to the arguments below. Fix any $u^* \in \mathcal{U}_{IR}$. First, using transfers, we can find player-specific punishments for u^*: consider the vectors $\{u^i : i \in N\}$ defined by

$$u^i_j = \begin{cases} u^*_j - \zeta & \text{if } j = i, \\ u^*_j + \frac{\zeta}{n-1} & \text{if } j \neq i. \end{cases}$$

Observe that $\{u^i\}_{i=1}^n \subseteq \mathcal{U}_{IR}$ when ζ is sufficiently small, and that for all i, $u^i_i < u^*_i$ and for all $j \neq i$, $u^j_i > u^i_i$. Therefore, this is a vector of player-specific punishments.

Second, let us define *individual minmaxing alternatives*: for each player $i \in N$, let $\underline{a}_i \in \arg\min_{a \in A} \max_{a' \in E_{\{i\}}(a)} v_j(a')$ denote the alternative that will be used to min-max player i. Since A is compact, v is continuous, and $E_{\{i\}}(\cdot)$ is continuous and compact-valued, by Berge's maximum theorem, $\underline{a}_i$ is well-defined for each $i \in N$. By construction, $v_i(a') \leq \underline{v}^\circ_i$ for all $i \in N$ and $a' \in E_{\{i\}}(\underline{a}_i)$ and in particular, $v_i(\underline{a}_i) \leq \underline{v}_i$.

Given these payoffs and alternatives, let $\kappa \in (0, 1)$ be such that for every $\widetilde{\kappa} \in [\kappa, 1]$, the following is true for every i:

$$(1-\widetilde{\kappa})v_i(\underline{a}_i) + \widetilde{\kappa} u_i^i > \underline{v}_i \tag{18}$$

$$\text{For every } j \neq i: \quad (1-\widetilde{\kappa})v_j(\underline{a}_i) + \widetilde{\kappa} u_j^i > (1-\widetilde{\kappa})\underline{v}_j + \widetilde{\kappa} u_j^j \tag{19}$$

By an argument identical to that which we saw in Theorem 1, there exists a value of $\kappa \in (0,1)$ such that the inequality holds for all $\widetilde{\kappa} \in [\kappa, 1]$, $i \in N$ and $j \in N \backslash \{i\}$. Let $L(\delta) := \left\lceil \frac{\log \kappa}{\log \delta} \right\rceil$ where $\lceil \cdot \rceil$ is the ceiling function. As before, we use the property that $\lim_{\delta \to 1} \delta^{L(\delta)} = \kappa$.

Lemma 8 implies that there exist payoff vectors $\{\widetilde{u}^1, \ldots, \widetilde{u}^M\} \subseteq \mathcal{U}(\overline{a}) \cup \mathcal{U}(\underline{a})$ such that $\mathcal{U}_{IR} \subseteq \operatorname{co}(\widetilde{u}^1, \ldots, \widetilde{u}^M)$, where $\widetilde{u}^m = u(\widetilde{a}^m, \widetilde{T}^m)$ for some alternative $\widetilde{a}^m \in \{\overline{a}, \underline{a}\}$ and transfers matrix $\widetilde{T}^m$ for each $m = 1, \ldots, M$. Lemma 1 then guarantees that for any $\epsilon > 0$, there exists $\underline{\delta} \in (0,1)$ such that for all $\delta \in (\underline{\delta}, 1)$, there exist sequences $\left\{\{a^{d,\tau}, T^{d,\tau}\}_{\tau=0}^{\infty} : d \in N \cup \{*\}\right\}$ such that for each d and t, $(1-\delta)\sum_{\tau=0}^{\infty} \delta^\tau u(a^{d,\tau}, T^{d,\tau}) = u^d$ and $\left|\left|u^d - (1-\delta)\sum_{\tau=t}^{\infty} \delta^\tau u(a^{d,\tau}, T^{d,\tau})\right|\right| < \epsilon$. We fix an

$$\epsilon < (1-\kappa) \min\left\{ \min_{d \in N \cup \{*\}, j \in N \backslash \{d\}} (u_j^d - u_j^j),\ \min_{i \in N} (u_i^i - \underline{v}_i) \right\}$$

and given that ϵ, consider δ exceeding the appropriate $\underline{\delta}$.

Let $\widetilde{\mathcal{T}} := \{\widetilde{T}^m\}_{m=1}^M$ be the set of transfer matrices defined above. By Assumption 1, when the plan recommends a transfers matrix $T \in \widetilde{\mathcal{T}}$ and a coalition $C \in \mathcal{C}$ blocks, there exists $\widetilde{B} > 0$ such that

$$\sum_{i \notin C, j \in C} \chi_{ij}^C(T) \leq \widetilde{B} \text{ for all } T \in \widetilde{\mathcal{T}} \text{ and } C \in \mathcal{C}. \tag{20}$$

Now we describe the plan that we use to support u^*. Let $\mathbf{0}$ denote the transfer matrix where all players make no transfers. Consider the plan represented by the automaton $(W, w(*, 0), f, \gamma)$, where

- $W := \left\{w(d,\tau) | d \in N \cup \{*\}, \tau \geq 0\right\} \cup \{\underline{w}(i,\tau) | i \in N, 0 \leq \tau < L(\delta)\}$ is the set of possible states and $w(*,0)$ is the initial state;
- $f : W \to \overline{\mathcal{O}}$ is the output function, where $f(w(d,\tau)) = (a^{d,\tau}, \emptyset, T^{d,\tau})$ and $f(\underline{w}(i,\tau)) = (\underline{a}_i, \emptyset, \mathbf{0})$;
- $\gamma : W \times \overline{\mathcal{O}} \to W$ is the transition function. For states of the form $w(d,\tau)$, the

transition is

$$\gamma\big(w(d,\tau),(a,C,T)\big) = \begin{cases} \underline{w}(j^*,0) & \text{if } C \neq \emptyset \\ & \text{where } j^* \in \arg\min_{j\in C} u_j(a,T), \\ w(d,\tau+1) & \text{otherwise.} \end{cases}$$

For states in $\{\underline{w}(i,\tau)|0 \leq \tau < L(\delta)-1\}$,

$$\gamma\big(\underline{w}(i,\tau),(a,C,T)\big) = \begin{cases} \underline{w}(j^*,0) & \text{if } \{C \neq \emptyset\} \cap \big(\{u_i(a,T) > \underline{v}_i\} \cup \{i \notin C\}\big) \\ & \text{where } j^* \in \arg\min_{C\setminus\{i\}} u_j(a,T), \\ \underline{w}(i,0) & \text{if } \{C \neq \emptyset\} \cap \{u_i(a,T) \leq \underline{v}_i\} \cap \{i \in C\}, \\ \underline{w}(i,\tau+1) & \text{otherwise.} \end{cases}$$

For states of the form $\underline{w}(i,L(\delta)-1)$, the transition isa

$$\gamma\big(\underline{w}(i,L(\delta)-1),(a,C,T)\big) = \begin{cases} \underline{w}(j^*,0) & \text{if } \{C \neq \emptyset\} \cap \big(\{u_i(a,T) > \underline{v}_i\} \cup \{i \notin C\}\big) \\ & \text{where } j^* \in \arg\min_{C\setminus\{i\}} u_j(a,T), \\ \underline{w}(i,0) & \text{if } \{C \neq \emptyset\} \cap \{u_i(a,T) \leq \underline{v}_i\} \cap \{i \in C\}, \\ w(i,0) & \text{otherwise.} \end{cases}$$

The plan represented by the above automaton yields payoff profile u^*. By construction, the continuation values in different states, $V(\cdot)$, satisfy:

$$\big|\big|u^d - V(w(d,\tau))\big|\big| < \epsilon, \quad \tau = 0,1,\ldots$$

$$V(\underline{w}(i,\tau)) = (1-\delta^{L(\delta)-\tau})v(\underline{a}_i) + \delta^{L(\delta)-\tau}V(w(i,0)), \quad 0 \leq \tau \leq L(\delta)-1$$

In the NTU environment, since the feasible payoff set $\mathcal{V}$ is bounded, whenever a coalition deviates, we can find number $B > 0$ that bounds every player's stage-game payoff. With transferable utility, the *total* stage-game payoff of the deviating coalition is still bounded. The definition of j^* in the automaton above ensures that the "scapegoat" selected by the plan can be effectively deterred as $\delta \to 1$. It remains to show that this plan is a PCE. This is the next step.

States of the form $w(d,\tau)$: If a coalition $C \neq \emptyset$ blocks in automaton state $w(d,\tau)$

and the outcome $(\widehat{a}, C, \widehat{T})$ is realized, the plan punishes $j^* = \arg\min_{j \in C} u_j(\widehat{a}, \widehat{T})$. It follows that

$$u_{j^*}(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|} \sum_{j \in C} u_j(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|} \left[\max_{a \in A} \sum_{j \in C} v_j(a) + \sum_{j \in C} \sum_{k \notin C} \chi^C_{kj}(T^{d,\tau}) \right]$$
$$\leq \frac{1}{|C|} \left[\max_{a \in A} \sum_{j \in C} v_j(a) + \widetilde{B} \right] \leq \max_{C \in \mathcal{C}} \frac{1}{|C|} \left[\max_{a \in A} \sum_{j \in C} v_j(a) + \widetilde{B} \right].$$

where the first inequality follows from the minimum among a set of numbers being less than their average; the second inequality follows since C's total payoff comes from the generated payoffs plus the net transfers paid by players outside of C; lastly, the third inequality follows from (20) and the fact that all $T^{d,\tau}$ are drawn from $\{\widetilde{T}^m\}_{m=1}^M$. Therefore, we can find a uniform bound B_1 such that $u_{j^*}(\widehat{a}, \widehat{T}) < B_1$ for every δ and (d, τ).

Given this bound, we can use the analogue of the argument used in Theorem 1. For all τ, j^* obtains a payoff greater than $u^d_{j^*} - \epsilon$. By being participating in the blocking coalition C, j^* obtains a payoff less than

$$(1-\delta)B_1 + \delta V_{j^*}(\underline{w}(j^*, 0)) = (1-\delta)B_1 + \delta \left[(1-\delta^{L(\delta)}) v_j(\underline{a}_j) + \delta^{L(\delta)} u^{j^*}_{j^*} \right]$$

By the exact same argument as in Theorem 1, this not profitable for j^* and hence, coalition C cannot profitably block the plan if δ is sufficiently high.

States of the form $\underline{w}(i, \tau)$: Suppose coalition C blocks leading to the outcome $(\widehat{a}, C, \widehat{T})$, we prove that at least one player in C does not find it profitable to join the blocking coalition. There are two cases to consider.

Case I: *$i \in C$ and $u_i(\widehat{a}, \widehat{T}) \leq \underline{v}_i$.* In this case, the plan selects player i to be the scapegoat. She finds it unprofitable to block if

$$(1 - \delta^{L(\delta)-\tau}) v_i(\underline{a}_i) + \delta^{L(\delta)-\tau} u^i_i \geq (1-\delta)\underline{v}_i + \delta(1 - \delta^{L(\delta)}) v_i(\underline{a}_i) + \delta^{L(\delta)+1} u^i_i.$$

which follows from (18) for sufficiently high δ (using steps identical to the analogous argument in Theorem 1).

Case II: *Either $i \notin C$ or $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$.* First observe that in this case it must be that $C \neq \{i\}$. To see why, note that the plan recommends zero transfers in state $\underline{w}(i, \tau)$, so if $C = \{i\}$, it would receive no transfers from other players in $\widehat{T}$ by Assumption 1, and therefore $u_i(\widehat{a}, \widehat{T}) \leq \underline{v}_i$ by the definition of $\underline{a}_i$, which contradicts both $i \notin C$ and $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$. Given that $C \neq \{i\}$, the scapegoat $j^* \in \arg\min_{j \in C \setminus \{i\}} u_j(\widehat{a}, \widehat{T})$ is well defined.

Next we show that j^*'s stage-game payoff is bounded. If $i \notin C$, then

$$j^* = \arg\min_{j \in C \setminus \{i\}} u_j(\widehat{a}, \widehat{T}) = \arg\min_{j \in C} u_j(\widehat{a}, \widehat{T}),$$

and

$$u_{j^*}(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|} \sum_{j \in C} u_j(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|} \max_{a \in A} \sum_{j \in C} v_j(a). \tag{21}$$

Alternatively, if $i \in C$ and $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$, then

$$\begin{aligned} u_{j^*}(\widehat{a}, \widehat{T}) &\leq \frac{1}{|C|-1} \sum_{j \in C \setminus \{i\}} u_j(\widehat{a}, \widehat{T}) = \frac{1}{|C|-1} \Big[\sum_{j \in C} u_j(\widehat{a}, \widehat{T}) - u_i(\widehat{a}, \widehat{T}) \Big] \\ &\leq \frac{1}{|C|-1} \Big[\sum_{j \in C} u_j(\widehat{a}, \widehat{T}) - \underline{v}_i \Big] \leq \frac{1}{|C|-1} \Big[\max_{a \in A} \sum_{j \in C} v_j(a) - \underline{v}_i \Big]. \end{aligned} \tag{22}$$

In the second line above, the first inequality follows because we are considering the case $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$, the second inequality follows from Assumption 1 and the fact that the plan is recommending zero transfers. Inspecting the RHS of both (21) and (22), we note that given the continuity of $v_j(.)$ and the compactness of A, we can find $B_2 > 0$ such that $u_{j^*}(\widehat{a}, \widehat{T}) < B_2$.

Finally, we show that the scapegoat j^* does not find it profitable to join blocking coalition C. Player j^* does not benefit from blocking if

$$(1 - \delta^{L(\delta)-\tau}) v_{j^*}(\underline{a}_i) + \delta^{L(\delta)-\tau} u^i_{j^*} \geq (1-\delta) B_2 + \delta(1 - \delta^{L(\delta)}) v_{j^*}(\underline{a}_{j^*}) + \delta^{L(\delta)+1} u^{j^*}_{j^*}.$$

This inequality is satisfied for sufficiently high δ, and the argument follows the same steps as that of the analogous part of Theorem 1.

B.3 Completing the proof of Theorem 5

States of the form $w(d, \tau)$: Depending on whether or not the blocking coalition can make secret transfers, there are two cases to consider.

Case I: $C \in \mathcal{C}\backslash(\mathcal{S} \cup N)$. Suppose the outcome $(\widehat{a}, C, \widehat{T})$ is realized, then the plan punishes player $j^* = \arg\min_{j\in C} u_j(\widehat{a}, \widehat{T})$. It follows that

$$\begin{aligned} u_{j^*}(\widehat{a}, \widehat{T}) &\leq \frac{1}{|C|}\sum_{j\in C} u_j(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|}\left[\max_{a\in A}\sum_{j\in C} v_j(a) + \sum_{j\in C}\sum_{k\notin C}\chi^C_{kj}(T^{d,\tau})\right] \\ &\leq \frac{1}{|C|}\left[\max_{a\in A}\sum_{j\in C} v_j(a) + \widetilde{B}\right] \leq \widehat{B} + \widetilde{B}. \end{aligned}$$

where the first inequality follows from the minimum among a set of numbers being less than their average; the second inequality follows since C's total payoff comes from the generated payoffs plus the net transfers paid by players outside of C; the third inequality follows from (15) and the fact that all $T^{d,\tau}$ are drawn from $\{\widetilde{T}^m\}_{m=1}^M$; the last inequality follows from (16). Therefore, we can find a uniform bound B_1 such that $u_{j^*}(\widehat{a}, \widehat{T}) < B_1$ for every δ and (d, τ).

Given this bound, we can use the analogue of the argument used in Theorem 1. For all τ, j^* obtains a payoff greater than $u^d_{j^*} - \epsilon$. By being participating in the blocking coalition C, j^* obtains a payoff less than

$$(1-\delta)B_1 + \delta V_{j^*}(\underline{w}(j^*, 0)) = (1-\delta)B_1 + \delta\left[(1-\delta^{L(\delta)})v_j(\underline{a}_j) + \delta^{L(\delta)}u^{j^*}_{j^*}\right]$$

By the exact same argument as in Theorem 1, this not profitable for j^* and hence, coalition C cannot profitably block the plan if δ is sufficiently high.

Case II: $C \in \mathcal{S} \cup N$. The plan punishes coalition C. Since all $T^{d,\tau}$ are drawn from $\{\widetilde{T}^m\}_{m=1}^M$, we have

$$\begin{aligned} \sum_{i\in C'} u_i(\widehat{a}, \widehat{T}) &\leq \max_{a\in A}\sum_{j\in C} v_j(a) + \sum_{j\in C}\sum_{k\notin C}\chi^C_{kj}(T^{d,\tau}) \\ &\leq \max_{a\in A}\sum_{j\in C} v_j(a) + \widetilde{B} \leq \widehat{B} + \widetilde{B}. \end{aligned}$$

The second inequality above follows from (15), while the third inequality follows from (16). Thus, we can find a uniform bound B_2 such that the total payoff from deviation for coalition C, $\sum_{i\in C} u_i(\widehat{a}, \widehat{T})$ is less than B_2 for every C, δ and (d, τ).

Coalition C has total payoff of at least $\sum_{i\in C} u^d_i - \epsilon$ without blocking. By blocking the recommendation from the plan, C obtains a total payoff less than

$$(1-\delta)B_2 + \delta \sum_{i \in C} V_i(\underline{w}(C,0)) = (1-\delta)B_2 + \delta \left[(1-\delta^{L(\delta)}) \sum_{i \in C} v_i(\underline{a}_C) + \delta^{L(\delta)} \sum_{i \in C} u_i^C \right]$$

For the blocking to be profitable, the total value for C must be higher. So the blocking is unprofitable if the above term is no more than $\sum_{i \in C} u_i^d - \epsilon$. We prove that this is the case both for $C \neq d$ and $C = d$.

First consider $C \neq d$. Observe that

$$\begin{aligned} &\lim_{\delta \to 1} (1-\delta)B_2 + \delta \left[(1-\delta^{L(\delta)}) \sum_{i \in C} v_i(\underline{a}_C) + \delta^{L(\delta)} \sum_{i \in C} u_i^C \right] \\ &\qquad = (1-\kappa) \sum_{i \in C} v_i(\underline{a}_C) + \kappa \sum_{i \in C} u_i^C < \sum_{i \in C} u_i^C < \sum_{i \in C} u_i^d - \epsilon, \end{aligned}$$

where the last inequality above follows from the definition of ϵ. It follows that C cannot profitably block the plan.

Now suppose that $C = d$. The payoff from blocking being less than $\sum_{i \in C} u_i^C - \epsilon$ can be re-written as

$$(1-\delta)(B_2 - \sum_{i \in C} u_i^C) + \epsilon \leq \delta(1-\delta^{L(\delta)})(\sum_{i \in C} u_i^C - \sum_{i \in C} v_i(\underline{a}_C))$$

As $\delta \to 1$, the LHS converges to ϵ. Because $\lim_{\delta \to 1} \delta^{L(\delta)} = \kappa$, the RHS converges to $(1-\kappa)(\sum_{i \in C} u_i^C - \sum_{i \in C} v_i(\underline{a}_C))$. By the definition of ϵ, the above inequality holds, and therefore, there is no profitable blocking if δ is sufficiently high.

States of the form $\underline{w}(i,\tau)$ where $i \in N$: Suppose coalition C blocks leading to the outcome $(\widehat{a}, C, \widehat{T})$, there are 3 cases to consider.

<u>Case I: $C \in \mathcal{C} \backslash (\mathcal{S} \cup N)$, $i \in C$, and $u_i(\widehat{a}, \widehat{T}) \leq \underline{v}_i$</u>. In this case, the plan selects player i to be the scapegoat. She finds it unprofitable to block if

$$(1-\delta^{L(\delta)-\tau})v_i(\underline{a}_i) + \delta^{L(\delta)-\tau} u_i^i \geq (1-\delta)\underline{v}_i + \delta(1-\delta^{L(\delta)})v_i(\underline{a}_i) + \delta^{L(\delta)+1} u_i^i.$$

which follows from (13) for sufficiently high δ (using steps identical to the analogous argument in Theorem 1).

<u>Case II: $C \in \mathcal{C} \backslash (\mathcal{S} \cup N)$, but either $i \notin C$ or $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$</u>. First observe that in this

case it must be that $C \neq \{i\}$, so the scapegoat $j^* \in \arg\min_{j \in C \setminus \{i\}} u_j(\widehat{a}, \widehat{T})$ is well defined.

Next we show that j^*'s stage-game payoff is bounded. If $i \notin C$, then

$$j^* \in \underset{C \setminus \{i\}}{\arg\min}\, u_j(\widehat{a}, \widehat{T}) = \underset{C}{\arg\min}\, u_j(\widehat{a}, \widehat{T}),$$

and

$$u_{j^*}(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|} \sum_{j \in C} u_j(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|} \max_{a \in A} \sum_{j \in C} v_j(a). \tag{23}$$

Alternatively, if $i \in C$ and $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$, then

$$\begin{aligned} u_{j^*}(\widehat{a}, \widehat{T}) &\leq \frac{1}{|C|-1} \sum_{j \in C \setminus \{i\}} u_j(\widehat{a}, \widehat{T}) = \frac{1}{|C|-1} \Big[\sum_{j \in C} u_j(\widehat{a}, \widehat{T}) - u_i(\widehat{a}, \widehat{T}) \Big] \\ &\leq \frac{1}{|C|-1} \Big[\sum_{j \in C} u_j(\widehat{a}, \widehat{T}) - \underline{v}_i \Big] \leq \frac{1}{|C|-1} \Big[\max_{a \in A} \sum_{j \in C} v_j(a) - \underline{v}_i \Big]. \end{aligned} \tag{24}$$

In the second line above, the first inequality follows because we are considering the case $u_i(\widehat{a}, \widehat{T}) > \underline{v}_i$, the second inequality follows from Assumption 1 and the fact that the plan is recommending zero transfers. Inspecting the RHS of both (23) and (24), we note that by (16) we can find $B_3 > 0$ such that $u_{j^*}(\widehat{a}, \widehat{T}) < B_3$.

Finally, we show that the scapegoat j^* does not find it profitable to join blocking coalition C. Player j^* does not benefit from blocking if

$$(1 - \delta^{L(\delta)-\tau}) v_{j^*}(\underline{a}_i) + \delta^{L(\delta)-\tau} u^i_{j^*} \geq (1-\delta) B_3 + \delta(1 - \delta^{L(\delta)}) v_{j^*}(\underline{a}_{j^*}) + \delta^{L(\delta)+1} u^{j^*}_{j^*}.$$

This inequality is satisfied for sufficiently high δ, and the argument follows the same steps as that of the analogous part of Theorem 1.

<u>Case III: $C \in \mathcal{S} \cup N$</u>. If $C = \{i\}$, then the analysis is identical to Case I above. Suppose $C \neq \{i\}$. For the blocking to be profitable, the total value for C must be higher. Note that since the plan is recommending zero transfers, when coalition C blocks, it receives net zero incoming transfers by Assumption 1. Given (16), coalition C's total stage-game payoff is bounded above by $\widehat{B}$. So the claim amounts to showing

$$(1 - \delta^{L(\delta)-\tau}) \sum_{j \in C} v_j(\underline{a}_i) + \delta^{L(\delta)-\tau} \sum_{j \in C} u^i_j \geq (1-\delta)\widehat{B} + \delta(1 - \delta^{L(\delta)}) \sum_{j \in C} v_j(\underline{a}_C) + \delta^{L(\delta)+1} \sum_{j \in C} u^C_j.$$

This inequality is satisfied for sufficiently high δ, and the argument follows the same steps as that of the analogous part of Theorem 1.

States of the form $\underline{w}(S, \tau)$ where $S \in \mathcal{S}$: There are two cases to consider.
Case I: $C \in \mathcal{S} \cup N$. The argument mirrors the same steps as Case III for states in the form of $\underline{w}(i, \tau)$.
Case II: $C \in \mathcal{C} \backslash (\mathcal{S} \cup N)$. Suppose the outcome $(\widehat{a}, C, \widehat{T})$ is realized. The plan punishes $j^* = \arg\min_{j \in C} u_j(\widehat{a}, \widehat{T})$. It follows that

$$u_{j^*}(\widehat{a}, \widehat{T}) \leq \frac{1}{|C|}\Big[\sum_{j \in C} u_j(\widehat{a}, \widehat{T})\Big] \leq \frac{1}{|C|}\sum_{j \in C} v_j(\widehat{a}) \leq \widehat{B}.$$

The first inequality above follows since the minimum among any numbers is less than their average; the second inequality follows because the plan recommends zero transfers when C blocks; the last inequality follows from (16).

Player j^* does not benefit from this blocking if

$$(1 - \delta^{L(\delta)-\tau}) v_{j^*}(\underline{a}_C) + \delta^{L(\delta)-\tau} u^S_{j^*} \geq (1-\delta)\widehat{B} + \delta(1 - \delta^{L(\delta)}) v_{j^*}(\underline{a}_{j^*}) + \delta^{L(\delta)+1} u^{j^*}_{j^*}.$$

Since $S \in \mathcal{S}$ and $j^* \in N$, it must be that $S \neq j^*$. The inequality above is satisfied for sufficiently high δ, and the argument follows the same steps as that of the analogous part of Theorem 1. $\square$

B.4 Proof of Proposition 1 on p. 20

Note that for each player $i \in \mathcal{F} \cup \mathcal{W}$, the individual minmax is $\underline{v}_i = 0$. The result then follows from applying Theorem 4. $\square$

B.5 Proof of Proposition 2 on p. 20

Preliminary Results.

Lemma 9. *Let $\widehat{x} = \max_{\phi \in A} \sum_{i \in N} v_i(\phi)$. If $u \in \mathbb{R}^n$ satisfies $\sum_{i \in N} u_i \leq \widehat{x}$ and $\sum_{i \in C} u_i \geq \underline{u}_C$ for all $C \in \mathcal{E}$, then $\sum_{i \in N} u_i = \widehat{x}$.*

Proof. Take any $\widetilde{u} \in \mathbb{R}^n$ satisfying $\sum_{i \in N} \widetilde{u}_i \leq \widehat{x}$ and $\sum_{i \in C} \widetilde{u}_i \geq \underline{u}_C$ for all $C \in \mathcal{E}$. Towards a contradiction, suppose that $\widetilde{u}$ is not utilitarian efficient; that is, suppose $\sum_{i \in N} \widetilde{u}_i < \widehat{x}$. Then there exists an assignment $\phi' \in A$ such that $\sum_{i \in N} v_i(\phi') > \sum_{i \in N} \widetilde{u}_i$.

Let π' denote the partition of players into essential coalitions induced by the matching ϕ', so $\pi' \subseteq \mathcal{E}$. It follows that there exists $C' \in \pi' \subseteq \mathcal{E}$ such that $\underline{u}_{C'} = \sum_{i \in C'} v_i(\phi') > \sum_{i \in C'} \widetilde{u}_i$, which is a contradiction to the assumption that $\sum_{i \in C} \widetilde{u} \geq \underline{u}_C$ for all $C \in \mathcal{E}$. So $\widetilde{u}$ must be utilitarian efficient. $\square$

Lemma 10. *Let* $\widehat{x} = \max_{\phi \in A} \sum_{i \in N} v_i(\phi)$. *The set* $\mathcal{K}$ *is characterized by*

$$\mathcal{K} = \{u \in \mathbb{R}^n : \sum_{i \in N} u_i = \widehat{x}, \sum_{i \in C} u_i \geq \underline{u}_C \textit{ for all } C \in \mathcal{E}\}. \tag{25}$$

Proof. Take any $\widetilde{u} \in \mathcal{K}$. Suppose, for the sake of contradiction, that there exists some $C \in \mathcal{E}$ such that $\sum_{i \in C} \widetilde{u}_i < \underline{u}_C$, then $\widetilde{u}$ would be blocked by C, which contradicts the assumption that $\widetilde{u} \in \mathcal{K}$. So $\sum_{i \in C} \widetilde{u}_i \geq \underline{u}_C$ must hold for all $C \in \mathcal{E}$. Lemma 9 then implies that $\widetilde{u}$ is utilitarian efficient, so $\widetilde{u}$ satisfies the conditions in (25).

For the converse, take any $\widetilde{u}$ that satisfies the conditions in (25). We will show that $\widetilde{u} \in \mathcal{K}$, i.e., there exists a core allocation (ϕ, T) such that $\widetilde{u} = u(\phi, T)$. Since $\mathcal{K}$ is nonempty, there exists a core alternative $(\widetilde{\phi}, \widetilde{T})$, which by the arguments above must satisfy $\sum_{i \in N} v_i(\widetilde{\phi}) = \widehat{x}$. Since $\sum_{i \in N} \widetilde{u}_i = \widehat{x}$, there exists $\widetilde{T}' \in \mathcal{T}$ such that $\widetilde{u} = u(\widetilde{\phi}, \widetilde{T}')$. Note however that $\widetilde{T}'$ may involve nonzero transfers between players who are not in an employment relationship, so $(\widetilde{\phi}, \widetilde{T}')$ may not constitute a matching. Nevertheless, let $\widetilde{\pi}$ denote the partition of players induced by $\widetilde{\phi}$. For every $C \in \widetilde{\pi}$, it must hold that

$$\sum_{i \in C, j \notin C} \widetilde{T}'_{ij} - \sum_{i \in C, j \notin C} \widetilde{T}'_{ji} = 0,$$

for otherwise we would have $\sum_{i \in C'} \widetilde{u}_i < \sum_{i \in C'} v_i(\widetilde{\phi})$ for some $C' \in \widetilde{\pi}$, contradicting the fact that $\widetilde{u}$ satisfies (25). Therefore, we can construct $\widetilde{T}'' \in \mathcal{T}$ such that

$$\widetilde{u} = u(\widetilde{\phi}, \widetilde{T}''), \quad \text{and} \quad \widetilde{T}''_{ij} \neq 0 \text{ only if } i = \widetilde{\phi}(j) \text{ or } i \in \widetilde{\phi}(j),$$

so $(\widetilde{\phi}, \widetilde{T}'')$ is a matching that induces payoff profile $\widetilde{u}$. Since $\sum_{i \in C} \widetilde{u}_i \geq \underline{u}_C$ for all $C \in \mathcal{E}$, $(\widetilde{\phi}, \widetilde{T}'')$ cannot be blocked by any coalition, so $(\widetilde{\phi}, \widetilde{T}'')$ is a core allocation, and therefore $\widetilde{u} \in \mathcal{K}$. $\square$

Lemma 11. *Let*

$$\mathcal{U}^{\mathcal{M}} := co\Big(\Big\{u \in \mathbb{R}^n : \exists (\phi, T) \in \mathcal{M} \textit{ such that } u = u(\phi, T)\Big\}\Big)$$

denote the convex hull of all feasible matching payoffs. Then

$$\left\{u \in \mathcal{U}^{\mathcal{M}} : \sum_{i \in C} u_i \geq \underline{u}_C \text{ for all } C \in \mathcal{E}\right\} = \mathcal{K}.$$

Proof. The fact that $\mathcal{K} \subseteq \left\{u \in \mathcal{U}^{\mathcal{M}} : \sum_{i \in C} u_i \geq \underline{u}_C \text{ for all } C \in \mathcal{E}\right\}$ follows from the definition of $\mathcal{K}$.

To show $\left\{u \in \mathcal{U}^{\mathcal{M}} : \sum_{i \in C} u_i \geq \underline{u}_C \text{ for all } C \in \mathcal{E}\right\} \subseteq \mathcal{K}$, take any $\widetilde{u} \in \mathcal{U}^{\mathcal{M}}$, since $\widetilde{u}$ is a convex combination of feasible payoff vectors, it must be that

$$\sum_{i \in N} \widetilde{u}_i \leq \widehat{x} := \max_{\phi \in A} \sum_{i \in N} v_i(\phi).$$

Lemma 9 then implies that $\sum_{i \in N} \widetilde{u}_i = \widehat{x}$, so $\widetilde{u} \in \mathcal{K}$ by Lemma 10. $\square$

Proof of Proposition 2.

We first prove that every payoff vector in $\mathcal{K}$ can be supported by a public PCE. For any $\widetilde{u} \in \mathcal{K}$ there exists core allocation (ϕ, T) such that $\widetilde{u} = u(\phi, T)$. Consider the plan $\widetilde{\sigma}$ defined by $\widetilde{\sigma}(h) = (\phi, T)$ for all $h \in \overline{\mathcal{H}}$. The plan $\widetilde{\sigma}$ is obviously public and produces discounted payoff profile u. Given that (ϕ, T) is a core allocation, $\widetilde{\sigma}$ is also a PCE.

We now prove that for every $\delta \geq 0$, every public PCE implements a discounted payoff profile in $\mathcal{K}$. By Theorem 5, for every $\delta \geq 0$, every discounted payoff profile $\widetilde{u}$ produced by a public PCE must satisfy $\sum_{i \in C} \widetilde{u}_i \geq \underline{u}_C$ for all $C \in \mathcal{E}$, so $\widetilde{u} \in \left\{u \in \mathcal{U}^{\mathcal{M}} : \sum_{i \in C} u_i \geq \underline{u}_C \text{ for all } C \in \mathcal{E}\right\}$. By Lemma 11, $\widetilde{u}$ then must be an element of $\mathcal{K}$. $\square$

B.6 Proof of Proposition 3 on p. 22

Preliminary Results.

Lemma 12. *All static stable matchings fill slots* $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ *while leaving other slots vacant; more over, all workers receive the same payoff* r *where* $\max\{0, \eta(L+1)\} \leq r \leq \max\{0, \eta(L)\}$.

Proof. We break down the proof into two parts.
<u>Part 1</u>: All static stable matchings fill slots $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ while leaving other slots vacant.

Let m be any static stable matching. We first show that m must be utilitarian efficient. Suppose, for the sake of contradiction, that m is not utilitarian efficient.

Then there exists a reassignment of players that increases players' total payoff, which implies the existence of $f \in \mathcal{F}$ and $W \subseteq \mathcal{W}$ such that $v_f(W) + \sum_{w \in W} v_w(f) > v_f(m) + \sum_{w \in W} v_w(m)$. But this implies that m is profitably blocked by (f, W), contradicting the stability of m.

Next, we show that since m is utilitarian efficient, it fills all slots in $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$. Suppose, for the sake of contradiction, that there exists a slot $(\widetilde{f}, \widetilde{l}) \in \{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ that is not filled. Let $\widetilde{l}^* = \min\{l : (\widetilde{f}, l) \text{ is unfilled}\}$ denote the first unfilled position at firm $\widetilde{f}$. Since firms have diminishing marginal products, we have $\rho(\widetilde{f}, \widetilde{l}^*) \geq \rho(\widetilde{f}, \widetilde{l})$, so $(\widetilde{f}, \widetilde{l}^*)$ is an open slot in $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ that is immediately accessible by workers. Since there are L workers in total, if not all slots in $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ are filled, there exists $w' \in W$ who is either unemployed or filling a slot outside of $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$. In the first scenario, matching w' to the unfilled slot $(\widetilde{f}, \widetilde{l}^*)$ would strictly increase the total surplus. In the second scenario, let $(\widehat{f}, \widehat{l})$ be the slot filled by w', and let $\widehat{l}^* = \max\{l : (\widehat{f}, l) \text{ is filled}\}$ denote the last occupied slot at firm $\widehat{f}$, and $\widehat{w}^*$ denote the worker filling $(\widehat{f}, \widehat{l}^*)$. It follows from decreasing marginal product that $(\widehat{f}, \widehat{l}^*)$ is also outside of $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$, so matching $\widehat{w}^*$ to the unfilled slot $\rho(\widetilde{f}, \widetilde{l}^*)$ instead would strictly increase the total surplus, again contradicting the utilitarian efficiency of m. Thus, all stable matchings must fill the slots in $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$.

To show that all slots outside of $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ are vacant, there are two cases to consider. If $\eta(L) > 0$, we know from the arguments above that the L slots in $\{(f, l) : \rho(f, l) \geq \eta(L)\}$ are filled, so all other slots must be vacant. If $\eta(L) < 0$, then the set $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}$ becomes $\{(f, l) : \rho(f, l) \geq 0\}$, and let us suppose, for the sake of a contradiction, that some slot $(\widetilde{f}, \widetilde{l})$ with $\rho(\widetilde{f}, \widetilde{l}) < 0$ is filled. Let $\widetilde{l}^* = \max\{l : (\widetilde{f}, l) \text{ is filled}\}$ denote the last filled slot at firm $\widetilde{f}$, and let $\widetilde{w}$ denote the worker matched to this position. Due to decreasing marginal returns, we have $\rho(\widetilde{f}, \widetilde{l}^*) < 0$ as well, so simply unmatching $\widetilde{w}$ from $(\widetilde{f}, \widetilde{l}^*)$ will increase the total surplus. This contradicts the efficiency of m, which implies that m would not not stable. It follows that all stable matchings must leave slots outside $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ vacant.

<u>Part 2</u>: All workers are paid the same wage r, where $\max\{0, \eta(L+1)\} \leq r \leq \max\{0, \eta(L)\}$.

First we establish that all workers have the same wage. Take any static stable matching m. From Part 1, all positions in $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ are filled.

We prove that all workers have the same wage under two separate cases.

First, suppose $\eta(L) < 0$. It follows that $|\{(f,l) : \rho(f,l) \geq \max\{0, \eta(L)\}\}| < L$, so in the stable matching m there exists a worker $\widetilde{w}$ who is unmatched. This means $\widetilde{w}$ receives 0 payoff in the stable matching m. It then follows that any other employed worker must also receive 0 payoff, since otherwise there is a profitable block where their employer replaces them with $\widetilde{w}$.

Second, suppose $\eta(L) > 0$. Since by assumption $\rho(f, L) < \max\{0, \eta(L)\}$, there exists f_1 and f_2 such that both f_1 and f_2 employ workers in m. Since workers are identical, each worker working for f_1 must receive the same payoff as any worker at f_2 in m. This implies that workers at f_1 and f_2 all have the same payoff. The same argument applies to workers employed by any other firm, so all workers receive the same payoff.

Let r denote the payoff that workers receive, we next show that $\max\{0, \eta(L+1)\} \leq r \leq \max\{0, \eta(L)\}$. It is obvious that $r \geq 0$ by workers' individual rationality. To complete the arguments, it suffices to demonstrate the validity of three statements: A. $r \geq \eta(L+1)$ if $\eta(L+1) > 0$; B. $r \leq \eta(L)$ if $\eta(L) > 0$; and C. $r = 0$ if $\eta(L) \leq 0$.

Statement A: if $\eta(L+1) > 0$, then decreasing marginal return implies $\eta(L) > 0$, so from Part 1 we know all L workers are assigned to $\{(f,l) : \rho(f,l) \geq \max\{0, \eta(L)\}\} = \{(f,l) : \rho(f,l) \geq \eta(L)\}$. Let $(\widetilde{f}, \widetilde{l})$ denote the slot with value $\rho(\widetilde{f}, \widetilde{l}) = \eta(L+1)$. By decreasing marginal return, any slot $\{(\widetilde{f}, l) : l < \widetilde{l}\}$ at $\widetilde{f}$ is in $\{(f,l) : \rho(f,l) \geq \eta(L)\}$ and already filled. It follows that $r \geq \eta(L+1)$, since otherwise $\widetilde{f}$ can profitably block m by poaching a worker from other firms, which generates additional surplus $\eta(L+1)$, while offering wage r' satisfying $\eta(L+1) > r' > r$.

Statement B: if $\eta(L) > 0$, again from Part 1 we know that all L workers are assigned to $\{(f,l) : \rho(f,l) \geq \eta(L)\}$. Let $(\widetilde{f}, \widetilde{l})$ be the slot such that $\rho(\widetilde{f}, \widetilde{l}) = \eta(L)$. By decreasing marginal return we know that $(\widetilde{f}, \widetilde{l})$ must be the last filled slot at firm $\widetilde{f}$. It follows that workers' payoff is no more than $\eta(L)$ since otherwise $\widetilde{f}$ can profitably block by firing the worker matched to the slot $(\widetilde{f}, \widetilde{l})$.

Statement C: if $\eta(L) < 0$, then there are at most $(L-1)$ slots with a positive surplus, which by Part 1 implies that in any stable matching there exists a worker $\widetilde{w}$ who is unmatched. In this case, workers' payoff must be 0 since otherwise the matching is profitably blocked by a firm replacing one of its employees with worker $\widetilde{w}$.

Combining statements A, B, and C lets us conclude that $\max\{0, \eta(L+1)\} \leq r \leq \max\{0, \eta(L)\}$. $\square$

Proof of Proposition 3. The first half of Proposition 3 follows from Proposition 1, while the second half of Proposition 3 follows from combining Proposition 2 and Lemma 12. □

B.7 Proof of Proposition 4 on p. 24

Since by assumption both markets M_1 and M_2 satisfy $\eta_i(L+1) > 0$, Lemma 12 implies that in each market M_i, where $i = 1$ or 2, all static stable matchings fill the slots in $\{(f,l) : \rho_i(f,l) \geq \max\{0, \eta_i(L)\}\} = \{(f,l) : \rho_i(f,l) \geq \eta_i(L)\}$. Moreover, the workers' payoff r in market M_i satisfies $\eta_i(L+1) \leq r \leq \eta_i(L)$. Recall that the total surplus is $\Pi_i := \sum_{\ell=1}^{L} \eta_i(\ell)$, while the set of potential workers' total surplus is $\Pi_i^{\mathcal{W}} = [L\eta_i(L+1), L\eta_i(L)]$, and the set of potential firms' total surplus is

$$\Pi_i^{\mathcal{F}} = \Pi_i - \Pi_i^{\mathcal{W}} = \Big[\sum_{\ell=1}^{L} \eta_i(\ell) - L\eta_i(L), \sum_{\ell=1}^{L} \eta_i(\ell) - L\eta_i(L+1)\Big].$$

To simplify notation let us denote $\underline{b}_i^{\mathcal{W}} := L\eta_i(L+1)$ and $\bar{b}_i^{\mathcal{W}} := L\eta_i(L)$, so $\Pi_i^{\mathcal{W}} = [\underline{b}_i^{\mathcal{W}}, \bar{b}_i^{\mathcal{W}}]$. Similarly, let $\underline{b}_i^{\mathcal{F}} := \sum_{\ell=1}^{L} \eta_i(\ell) - L\eta_i(L)$ and $\bar{b}_i^{\mathcal{F}} = \sum_{\ell=1}^{L} \eta_i(\ell) - L\eta_i(L+1)$, so $\Pi_i^{\mathcal{F}} = [\underline{b}_i^{\mathcal{F}}, \bar{b}_i^{\mathcal{F}}]$.

Let $s := \eta_1(1) = \eta_2(1)$. For each $2 \leq \ell \leq L$, define $\Delta_\ell^i := \eta_i(\ell-1) - \eta_i(\ell)$, so $\eta_i(\ell) = s - \sum_{k=2}^{\ell} \Delta_k^i$ for all $\ell \geq 2$. It follows that

$$\sum_{\ell=1}^{L} \eta_i(\ell) = sL - \sum_{\ell=2}^{L} (L+1-\ell)\Delta_\ell^i,$$
$$L\eta_i(L) = sL - L\sum_{\ell=2}^{L} \Delta_\ell^i, \quad \text{and } L\eta_i(L+1) = sL - L\sum_{\ell=2}^{L+1} \Delta_\ell^i.$$

This allows us to express the bounds for firms' and workers' aggregate surplus in terms of s and Δ_ℓ^i's, yielding

$$\underline{b}_i^{\mathcal{W}} = sL - L\sum_{\ell=2}^{L+1} \Delta_\ell^i, \text{ and } \bar{b}_i^{\mathcal{W}} = sL - L\sum_{\ell=2}^{L} \Delta_\ell^i, \tag{26}$$
$$\underline{b}_i^{\mathcal{F}} = \sum_{\ell=2}^{L} (\ell-1)\Delta_\ell^i, \quad \text{and } \bar{b}_i^{\mathcal{F}} = \sum_{\ell=2}^{L+1} (\ell-1)\Delta_\ell^i. \tag{27}$$

Market M_2 exhibits more more steeply decreasing returns than M_1 is equivalent to $\Delta_\ell^2 \geq \Delta_\ell^1$ for all $2 \leq \ell \leq L$, which implies $\eta_2(\ell) \leq \eta_1(\ell)$ for all $1 \leq \ell \leq L$, so $\Pi_2 \leq \Pi_1$.

In (26), all the Δ_ℓ^i's enter the bounds for worker surplus with negative coefficients, so $\underline{b}_2^{\mathcal{W}} \leq \underline{b}_1^{\mathcal{W}}$ and $\bar{b}_2^{\mathcal{W}} \leq \bar{b}_1^{\mathcal{W}}$, where the inequalities are strict if M_2 has strictly more steeply decreasing returns than M_1. By contrast, in (27) the Δ_ℓ^i terms enter the bounds with positive coefficients, so $\underline{b}_2^{\mathcal{F}} \geq \underline{b}_1^{\mathcal{F}}$ and $\bar{b}_2^{\mathcal{F}} \geq \bar{b}_1^{\mathcal{F}}$, where, again, the inequalities are strict if M_2 has strictly more steeply decreasing returns than M_1. Together, the directions of change for these bounds imply $\Pi_2^{\mathcal{W}} \preccurlyeq_S \Pi_1^{\mathcal{W}}$ and $\Pi_2^{\mathcal{F}} \succcurlyeq_S \Pi_1^{\mathcal{F}}$, with strict set orders if M_2 has strictly more steeply decreasing returns than M_1. $\square$

B.8 Proof of Proposition 5 on p. 26

Preliminary Results. We will use an alternative $a \in A$ to also represent its generated payoff profile $v(a)$. We establish two preliminary results. Lemma 13 establish the existence of "punishment PCEs" $\{\sigma^i\}_{i=1}^n$ that guarantee $U_i(\emptyset|\sigma^i) = 0$ for each player i. Lemma 14 proves that any PCE can be enforced by punishments where every member of a deviating coalition simultaneously obtains 0.

Lemma 13. *Under perfect monitoring, for every player $i \in N$, there is a PCE σ^i such that $U_i(\emptyset|\sigma^i) = 0$ when $\delta > \frac{n-2}{n-1}$.*

Proof. We consider two case, $|D| = 1$ and $|D| \geq 2$. The case where $|D| = 1$ requires the discount factor to be sufficiently high. The case where there are two or more veto players ($|D| \geq 2$) applies for every discount factor.

Case 1: $|D| = 1$. Suppose without loss of generality that D consists of player 1. Let $\widehat{a} := (1, 0, \ldots, 0)$ denote the unique core alternative, and $\overline{a} := (0, \frac{1}{n-1}, \ldots, \frac{1}{n-1})$ denote the alternative that equally divides the total payoff among all non-veto players.

For $i \neq 1$, let σ^i be the plan that recommends the core alternative $\widehat{a}$ after every history, so each σ^i is a PCE that satisfies $U_i(\emptyset|\sigma^i) = 0$

For $i = 1$, let σ^1 be the plan that recommends $(\overline{a}, \emptyset)$ on path, and recommends $(\widehat{a}, \emptyset)$ at any history where blocking has occurred in the past. Note that $U_1(\emptyset|\sigma^1) = 0$. We will verify that σ^1 is a PCE. No coalition can profitably block once continuation play reverts back to the core alternative. On the path of play, consider a winning coalition $C \in \mathcal{W}$ blocking and choosing alternative a'. Since the game is non-dictatorial, if C is a winning coalition, player 1 cannot be its only member. Let $j \neq 1$ be a player in C.

Since $a'_j \leq 1$, we have

$$(1-\delta)a'_j + \delta 0 \leq 1 - \delta \leq \frac{1}{n-1}$$

so player j prefers following the plan σ^1 over blocking and reverting to the core alternative. As a result, no coalition C can profitably block the plan σ^1 at any history, so σ^1 is a PCE.

Case 2: $|D| \geq 2$. Without loss of generality, suppose $\{1, 2\} \subseteq D$. Let $a^1 := (1, 0, \ldots, 0)$ and $a^2 := (0, 1, 0, \ldots, 0)$ be two alternatives that allocate all payoff to player 1 and 2, respectively. It follows that both a^1 and a^2 are core alternatives.

Let σ^1 be the plan that recommends $(a^2, \emptyset)$ at all histories; for all $i \neq 1$, let σ^i be the plan that recommends $(a^1, \emptyset)$ at all histories. Each σ^i is a PCE, and $U_i(\emptyset|\sigma^i) = 0$ for every $i \in N$. □

Lemma 14. *Suppose $\mathcal{U}$ is the set of PCE-supportable payoff profiles. For each player $i \in N$, let $\underline{u}_i := \min_{u \in \mathcal{U}} u_i$ be player i's smallest possible payoff from PCEs. There is a stationary PCE with payoff profile a if and only if for every coalition C and alternative $a' \in E_C(a)$, there is a player $i \in C$ such that*

$$(1-\delta)a'_i + \delta \underline{u}_i \leq a_i \tag{28}$$

Proof. To see the "only if" direction, suppose (28) fails for some coalition C and $a' \in E_C(a)$. In other words, suppose there exists a coalition C and alternative a' such that

$$(1-\delta)a'_i + \delta \underline{u}_i > a_i \text{ for all } i \in C.$$

Towards a contradiction, suppose also that there exists a stationary PCE σ that supports payoff a. Since σ is a PCE, it follows that $U_i(h|\sigma) \geq \underline{u}_i$ for every $i \in C$ and all $h \in \mathcal{H}$. As a result, for every $i \in C$,

$$(1-\delta)a'_i + \delta U_i(a', C|\sigma) \geq (1-\delta)a'_i + \delta \underline{u}_i > a_i.$$

Moreover, since σ is stationary, it always plays a on path. The inequality above then implies that (a', C) is a profitable block for coalition C on path, contradicting σ being a stationary PCE.

For the "if" direction, (28) implies that for every coalition C and alternative $a' \in$

$E_C(a)$, there exits a player $i[a', C]$ and a PCE $\sigma[a', C]$ such that

$$(1-\delta)a'_{i[a',C]} + \delta U_{i[a',C]}\big(\emptyset \,|\, \sigma[a', C]\big) \leq a_{i[a',C]}. \tag{29}$$

Since the stage game exhibits default-independent power, by Theorem 2, we can without loss assume that each $\sigma[a', C]$ is a stationary PCE.

Consider a plan σ^* that recommends a on path, but switches to $\sigma[a', C)]$ if coalition C blocks to implement a'. Inequality (29) implies that on path, no coalition can find profitably block. In addition, the fact that each $\sigma[a', C]$ is a PCE ensures that after any off-path history, no coalition can profitably block. Finally, σ is also stationary since it is stationary on path, and each $\sigma[a', C]$ is also stationary. Therefore, σ is a stationary PCE that supports payoff a. □

Proof of Proposition 5.

Statement (a). Set $\underline{\delta} = \frac{n-2}{n-1}$. By Lemma 13, there exist PCEs $\{\sigma^i : i \in N\}$ satisfying $U_i(\emptyset|\sigma^i) = 0$ for all $i \in N$. It is straightforward to see that no players shared aligned payoffs in the stage game; in addition, no single player can form a winning coalition since the game is non-dictatorial. It follows that each player i's individual minmax is $\underline{v}_i = 0$. Moreover, this minmax payoff is achieved by the PCE σ^i.

By Lemma 14, in order for a payoff profile u to be supported by a stationary PCE, it is necessary and sufficient that for every winning coalition $C \in \mathcal{W}$, there exist no alternative $a' \in E_C(a)$ such that

$$(1-\delta)a'_i + \delta \cdot 0 = (1-\delta)a'_i > u_i \text{ for all } i \in C. \tag{30}$$

Note that the condition above is equivalent to $\sum_{i \in C} u_i \geq 1 - \delta$ for every $C \in \mathcal{W}$, since if $\sum_{i \in C} u_i < (1-\delta) \cdot 1$ for some coalition $C \in \mathcal{W}$, there would be a certain $a' \in E_C(u)$ representing a division of total payoff 1 among players in C, such that (30) holds for every $i \in C$. It follows that a payoff profile u is supportable by a stationary PCE if and only if $\sum_{i \in C} u_i \geq 1 - \delta$ for every $C \in \mathcal{W}$. Finally, Theorem 2 implies that this same set is also the set of PCE-supportable payoff profiles.

Statement (b). Because a winning coalition can obtain the entire dollar by blocking, its minmax value is 1. This statement then follows immediately from Theorem 5.

Statement (c). Let $\widehat{\mathcal{W}}$ denote the set of minimal winning coalitions. By definition, $\widehat{\mathcal{W}} \subseteq \mathcal{W}$ so $\cap_{C \in \mathcal{W}} C \subseteq \cap_{C \in \widehat{\mathcal{W}}} C$. Furthermore, $\cap_{C \in \widehat{\mathcal{W}}} C \subseteq \cap_{C \in \mathcal{W}} C$, since otherwise there exists $i \in \cap_{C \in \widehat{\mathcal{W}}} C$ and $\widetilde{C} \in \mathcal{W}$ such that $i \notin \widetilde{C}$, but this would lead to a contradiction since $\widetilde{C}$ must contains a winning coalition $\widehat{C}$, and $i \in \widehat{C}$. So $\cap_{C \in \widehat{\mathcal{W}}} C = \cap_{C \in \mathcal{W}} C = D$. By Theorem 5, every $C \in \widehat{\mathcal{W}}$ obtains total payoff 1. This implies the total payoff for players in D is 1. $\square$

B.9 Allowing Multiple Coalitions to Simultaneously Block

We consider two protocols that allow multiple coalitions to simultaneously block, neither of which affects our results.

Normal-Form Games: Consider the game of Example 1 on p. 6. For this setting, one could suppose that the coalitional blocking decision is observable or unobservable; to maintain consistency with the main text, we assume the former. If a coalition C blocks, then it can change its action profile to any $a_C \in \times_{i \in C} A_i$. We allow multiple disjoint coalitions to block simultaneously. Given that multiple coalitions may block, a stage-game outcome $\mathcal{O} := (a, \mathcal{D})$, records the chosen action profile and a collection of disjoint coalitions, $\mathcal{D} \subseteq \mathcal{C}$. A plan recommends an alternative after every history. Definitions 1 and 2 have natural counterparts in this setting.

Definition 10. *Coalition C* ***profitably blocks*** *plan σ at history h if given $\sigma(h) = a$, there exists $a'_C \in \times_{i \in C} A_i$ such that*

$$(1-\delta) v_i(a'_C, a_{-C}) + \delta U_i(\, h, ((a'_C, a_{-C}), C) \mid \sigma) > U_i(h \mid \sigma) \quad \textit{for all } i \in C.$$

This definition of profitably blocks specifies that coalition C has an incentive to block, assuming that no other coalition is doing so. The definition leaves open whether coalition C would choose to block when it assumes that others do block.

Definition 11. *A plan σ is a* ***perfect coalitional equilibrium*** *if, at every history, no coalition would profitably block.*

These definitions are essentially identical to those of our baseline model. That a plan can record the identity of multiple blocking coalitions creates no substantive difference: a PCE could respond to a past outcome in which multiple coalitions blocked as if no coalition had blocked without affecting any player's incentives.

Random Recognition: In this protocol, if two or more coalitions block in a given period, then one of the blocking decisions is implemented; the selection can be done with uniform probability. We view this protocol to be particularly relevant for political economy applications in which a "proposer" is randomly selected (Baron and Ferejohn 1989). We allow for the outcome in each period to record all blocking coalitions as well as the coalition whose block is recognized. Given this protocol, our formulation applies in its current form. In this setup, if a coalition blocks, its choice of alternative applies only when it is the recognized blocking coalition. A PCE would constitute a plan in which no coalition profitably blocks. Any plan that conditions its choice at a history only on the past alternatives and the identities of the recognized blocking coalitions is isomorphic to a plan of our baseline model; therefore, the PCE of our baseline model have natural counterparts in this extension.