Coalitions in Repeated Games*

S. Nageeb Ali† Ce Liu‡

June 4, 2026

Abstract

This paper proposes a framework and solution concept for repeated coalitional behavior. We model history-dependent schemes that deter coalitions from blocking using continuation promises and punishments. We evaluate the effectiveness of these schemes with and without transferable utility, and apply the framework to study the implications of wage transparency in labor markets.

*We thank anonymous referees, Dan Barron, Federico Echenique, Jon Eguia, Matt Elliott, Drew Fudenberg, Ben Golub, Yingni Guo, Navin Kartik, Scott Kominers, Maciej Kotowski, Elliot Lipnowski, George Mailath, Francesco Nava, Refine.ink, Ziwei Wang, Alex Wolitzky, Nathan Yoder, and various 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 study of repeated games models history-dependent schemes that enable players to cooperate even if each myopically favors defection. This canonical approach focuses on non-cooperative play in which actions are chosen only by individuals. But, in many contexts, analysts have found it more useful to allow groups of players to act jointly. For instance, matching and network theory model "pairwise stable arrangements" from which no pair of players can profitably deviate. Political economy models emphasize the "Condorcet Winner," a policy preferred by a majority of voters to all others. More broadly, the study of cooperative games studies the "core," an arrangement that no group of players would find it profitable to block. These notions are all modeled for static interactions, without harnessing the power of promises and punishments.

A natural question is how to marry these two approaches to group behavior. In this regard, we make two contributions. First, we develop a tractable and portable framework for studying repeated interaction in matching, voting, and other coalitional games. Second, we identify settings in which dynamic incentives deter group defection and those in which they fail to do so.

We illustrate our framework using the *Roommates Problem* (Gale and Shapley 1962). 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).

As is well known, no arrangement is pairwise stable. For instance, were Ann and Bella paired as roommates, Bella and Carol would each gain if they defected and roomed together instead. Our point of entry is to see how punishment and rewards can "solve" this problem.

Suppose that instead of a one-time decision, the trio made choices monthly. Each accrues the flow payoffs described above, and weights them by the per period discount factor δ. As in repeated games, no player can commit to her future behavior on- or off-path. What long-run stable matches can be supported through continuation play?

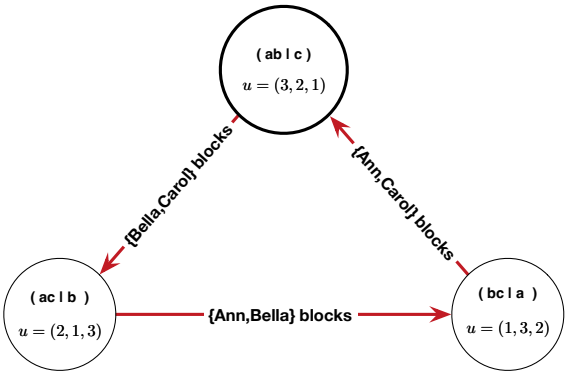


FIGURE 1. A perfect coalitional equilibrium for the Roommates Problem if $\delta \geq 1/2$.

Figure 1 depicts a stable scheme. On the path of play, Ann and Bella room together each month, leaving Carol out. While Bella and Carol might like to defect and share a room this month, the scheme assures that they refrain from doing so if δ exceeds $1/2$. For 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 rooming with Carol would 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, including those off-path, is self-enforcing.

We model schemes of this form in general games. We consider the repeated play of an abstract stage game that accommodates many settings: (i) a strategic form game in which players choose action profiles, (ii) a characteristic function game in which players obtain payoffs based on a partitional structure, (iii) matching games with and without transferable utility, and (iv) voting games in which different coalitions have varying power to push through policies. In this repeated game, we propose history-dependent schemes such that no coalition profits from blocking at any history given how it affects continuation play. We call such schemes *perfect coalitional equilibria* (PCE).

PCEs are inherently recursive. In strategic form games, PCE refines subgame perfect equilibria. More generally, PCE offers a tractable way to model how continuation promises shape coalitional behavior; PCE's recursive nature implies that its payoff set can be obtained via self-generation approaches developed by Abreu, Pearce, and Stacchetti (1990). In some settings, it can be even simpler: all PCE-supportable payoffs can be supported by stationary PCE if the stage game exhibits *default-independent power*. This property holds in the characteristic function games studied in cooperative game theory as well as in matching and voting models.

Our main results characterize when continuation promises and punishments deter coalitional blocking. 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 principle: a coalition can withstand punishments if and only if its members have highly aligned interests. Then, all members of that coalition enjoy a high minmax, considerably above their individual minmax. However, if there is any wedge in members' incentives, player-specific punishments can splinter coalitions. Then, the set of PCE-supportable payoffs virtually coincides with those that are feasible and *individually* rational.

Building on this principle, we explore features that align coalitions' interests. One such feature, we find, are *strongly symmetric* schemes in which players behave symmetrically after every history. Given their tractability, 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 constitute subgame perfect equilibria, we show that they typically are not PCE. The reason is that once players' incentives are aligned, punishments are no longer credible. By contrast, schemes that feature asymmetric play off-path can credibly deter blocking coalitions and support a larger payoff set.

We also study whether transferable utility aligns incentives within a coalition. One might have expected the answer to be yes: if a coalition achieves a net gain by blocking, transferable utility allows it to distribute those gains among its members. However, we show that if all transfers are publicly observed, a PCE can break coalitions apart by conditioning its continuation play on who pays whom. Only if a coalition can make transfers "secretly"—that is, without the transfers being publicly observed—can it entirely align its incentives. Such a coalition is then difficult to punish, which then limits what a PCE can enforce.

We consider a detailed application of our results to repeated labor-market matching, building on Kelso and Crawford (1982)'s seminal work. Here, we study the kinds of matches and wages that can be supported through repeated play. It turns out that the set of supportable outcomes hinges crucially on the transparency of past wages. Under wage transparency, a vast range of outcomes can be supported, enabling workers or firms to capture much of the surplus. By contrast, if wage terms are observed only by the employer and employee, there is a complete collapse of intertemporal incentives: the

supportable payoff set reduces to the core of the stage game. As to who then benefits from wage transparency depends on economic primitives. Workers benefit if they are plentiful or their marginal returns fall quickly. Absent transparency, competitive forces bid down their wages; by contrast, wage transparency enables them to enforce higher wages through collective-bargaining schemes. If workers are scarce or their marginal returns fall slowly, then it is firms who profit from wage transparency because that enables them to collusively suppress wages. Thus, our application highlights a new dimension to the debate on wage transparency, connecting it to the side of the market that it empowers to collude.

In the Supplementary Appendix, we consider a second application to repeated negotiations when some players have veto power. The core of this stage game entails that veto players capture the entire surplus, making them de facto dictators. Against that backdrop, we show that history dependence can promote egalitarianism but that these schemes destabilize once players can make secret side-payments.

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 labor market matching. Section 6 concludes. All proofs are in the Appendix and the Supplementary Appendix (Ali and Liu 2026b). The remainder of this introduction briefly discusses the related literature.

Related Literature: We build on results in repeated games, particularly Fudenberg and Maskin (1986), Abreu, Dutta, and Smith (1994), and Wen (1994); see Mailath and Samuelson (2006) for a textbook treatment. Our result on how secret side-payments can undermine intertemporal incentives also aligns with the findings of Barron and Guo (2021), who model how it disrupts collective punishment in relational contracting.

Several precursors model coalitional deviations in repeated strategic form games. Aumann (1959) and Rubinstein (1980) study Strong Nash and Strong Perfect Equilibria of infinitely repeated games. These concepts presume that deviating players can commit to long-term plans, even if those plans are not credible. DeMarzo (1992) adds interim incentive constraints and identifies benefits of scapegoat policies in finite horizon games. More closely related, Greenberg (1990) develops a general theory of social situations, encompassing strategic form, extensive form, and characteristic function games. In Chapter 9 of this book and Greenberg (1989), he applies this theory to repeated strategic form games and defines a set-valued concept, *Conservative Stable*

Standard of Behavior (CSSB). Intuitively, a CSSB models worst-case decisionmaking: a player contemplates deviating only if she is guaranteed to gain for every continuation play allowed by the standard, and a CSSB describes behavior that is immune to such deviations. Greenberg focuses on individual deviations and finds that the largest "nondiscriminating" CSSB coincides with the set of paths supported by subgame perfect Nash equilibria. He also proposes an extension that permits coalitional deviations but does not obtain general results for this setting. Instead, he illustrates using a specific common-interest game that every nondiscriminating "coalitional" CSSB selects the efficient action profile. This prediction matches PCE's for this game, as seen in our Theorem 1. While we formulate our solution concept differently—our approach is rooted in rational expectations rather than conservatism—we establish a connection between CSSB and PCE in Ali and Liu (2026a). Therein, we show that, for every discount factor, the largest nondiscriminating CSSB for Greenberg's coalitional repeated game coincides with the set of paths supportable by a PCE.

Outside of strategic form games, Bernheim and Slavov (2009) develop the notion of *Dynamic Condorcet Winners* (DCW) for repeated elections; our solution concepts coincide in that setting. While they characterize some properties of DCW, they do not identify its limits. Hence, apart from our more general focus, many issues central to our study do not feature in their work.

Like the work above, our baseline approach takes as given the exogenously specified blocking technology in a cooperative game and studies its implications for repeated play. A different route would instead demand that blocking be coalition-proof in the sense of Bernheim, Peleg, and Whinston (1987). They define inductively a solution concept for one-shot and finite-horizon games that is immune to "credible" coalitional deviations, where credibility requires that such deviations themselves be immune to credible sub-coalitional deviations. In the Supplementary Appendix, we develop the notion of *perfect coalition-proof equilibrium* (PCPE) that combines ideas of PCE and coalition proofness. Because coalition proofness imposes a more demanding criterion for which deviations are admissible, every PCE is necessarily a PCPE. However, we show that the two concepts result in identical coalitional minmaxes and that the sets of PCE and PCPE payoffs coincide as $\delta \to 1$.

Also related is the work on renegotiation-proofness initiated by Bernheim and Ray (1989) and Farrell and Maskin (1989); also see Miller and Watson (2013) and Safronov and Strulovici (2018). Renegotiation embodies the idea that players can collectively

replace an agreed upon continuation with another that they all prefer. A PCE, by contrast, studies the complementary question of when coalitions refrain from profitably blocking *given* the continuation plan. Like Greenberg (1989), our approach presupposes that players have a shared understanding of future play, can form temporary agreements to block, but cannot coordinate a shift in their beliefs about continuation play. This shared understanding must be internally consistent in that players can see that no coalition can profitably block at any future stage, given the continuation. An implication of Theorem 1 is that if players' interests are fully aligned—i.e., a common-interest game—then the unique PCE selects the efficient action profile, a prediction shared by Farrell and Maskin's strong renegotiation-proofness. However, the two notions diverge when interests are misaligned. A further difference is that the renegotiation literature largely emphasizes renegotiation by the *grand* coalition alone, whereas our results stem from allowing other coalitions to deviate. Two exceptions in this regard are Genicot and Ray (2003) and Cole, Krueger, Mailath, and Park (2024), who evaluate risk-sharing arrangements that are immune to subcoalitional renegotiation.

Our work connects to that on farsighted coalition formation. This literature has two interrelated strands. One evaluates farsighted coalitional negotiations over behavior within a static game; see, for example, Chwe (1994), and Ray and Vohra (2015). A more closely related strand, exemplified by Konishi and Ray (2003) and Gomes and Jehiel (2005), studies dynamic games in which payoffs accrue in real time and coalitions evaluate moves based on discounted continuation values. Our work makes two departures from this strand. One is that we study a repeated game, in which the past behavior has no direct payoff relevance for the future; by contrast, this literature studies stochastic games in which the alternative chosen in period t is the designated default for period $(t+1)$. The second is that we model the implications of history dependence whereas the prior work largely concerns Markov perfect behavior. An exception is Vartiainen (2011) who models history-dependence with limit-of-means discounting, and his focus is on existence of deterministic absorbing processes.

Dynamic considerations feature in studies of matching in which players account for future play when deciding with whom to match, e.g., Corbae, Temzelides, and Wright (2003), Damiano and Lam (2005), Kadam and Kotowski (2018a,b), Doval (2022), and Kotowski (2024). Rostek and Yoder (2024) propose a stability notion for static matching in which similar considerations emerge from players' strategic thinking.

Our application to labor market matching and wage transparency contribute to a

growing literature, surveyed in Cullen (2024). In a bargaining model with incomplete information, Cullen and Pakzad-Hurson (2023) show that wage transparency disadvantages workers. We offer a complementary perspective, studying the implications of wage transparency for the repeated game. 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) but relatively little is known about how carrots and sticks affect these matching markets.

Since our initial draft, Bardhi, Guo, and Strulovici (2026) use our approach to model stability in labor markets in which firms learn about workers' types, evaluating how early-career discrimination can result in persistent wage gaps. Liu (2023) and Liu, Wang, and Zhang (2025) also apply the framework here to study repeated matching between long-run firms and short-run workers. The former studies matching without transfers and shows that repetition overturns the Rural Hospital Theorem; the latter shows that repetition in a setting with transferable utility can lead to stable solutions despite peer effects and complementarities. Neither of these papers considers matching between long-run players nor the effect of wage transparency.

2 Model

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

The Stage Game. In each period, the players collectively choose an *alternative* a from A, a compact metrizable space. The alternative a generates 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.

Given an alternative, a coalition of players can choose to block or participate in it. Our specification allows for blocking by either a single coalition or multiple disjoint coalitions, but the former is what matters when evaluating stability. If a generic coalition C alone blocks alternative a, then it can choose any alternative in $E_C(a)$. The correspondence $E_C : A \rightrightarrows A$ is C's *effectivity correspondence* and offers a standard approach to model coalitional power.[1] For every coalition C, $E_C(\cdot)$ is continuous,

[1]The use of effectivity functions to model the cooperative game dates back to Rosenthal (1972), Moulin and Peleg (1982), Greenberg (1989, 1990), and Chwe (1994).

compact-valued, and reflexive (i.e., $a \in E_C(a)$). In our analysis, we also assume that larger coalitions can do more: for each alternative a, $E_{C'}(a) \subseteq E_C(a)$ for $C' \subseteq C$. This assumption is for notational convenience; we detail in footnotes, when necessary, how to adapt notation if this assumption fails.

We interpret alternatives as a description of behavior, for instance, an action profile in a strategic form game, the matches that form in a matching game, or the partition in a characteristic function game. Given an alternative a that players anticipate will occur, if a coalition C blocks, the set $E_C(a)$ reflects the set of alternatives it can freely choose *given the behavior of others*.

To see what this formulation captures, we revisit the Roommates Problem described in the introduction. An alternative is a rooming arrangement, and the set of alternatives, $A = \{ab|c, bc|a, ac|b, a|b|c\}$ is that of all arrangements, where $ij|k$ denotes i and j rooming together leaving k out. For the alternative that puts Ann and Bella together, Bella could block as an individual and choose an arrangement in $E_{\{\text{Bella}\}}(ab|c) = \{ab|c, a|b|c\}$. This specification is tantamount to her only choice *as an individual* being whether to accept or reject Ann as a roommate. Carol has even less power—$E_{\{\text{Carol}\}}(ab|c) = \{ab|c\}$—because she cannot room with someone else without that player's consent. But by teaming up and blocking as a pair, Bella and Carol could choose any alternative in $E_{\{\text{Bella,Carol}\}}(ab|c) = \{bc|a, ab|c, a|b|c\}$ where the first element denotes the pair rooming together.

The Roommates Problem is a specific illustration but the abstract form is considerably more general. We formalize below how to embed strategic form, voting, and characteristic function games in this setup.

Example 1. *Consider a strategic 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 comprises all 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 that are used to define Nash equilibria.*[2]

[2]We note two conceptual considerations. First, this example assumes that when coalition C blocks action profile a, the actions of players outside the coalition, a_{-C}, are held fixed. Our general formulation is agnostic about what players outside coalition C do; alternative specifications of how these players respond correspond to alternative specifications of $E_C(\cdot)$. Second, we do not require that the actions chosen by coalition C be individual best responses for each of its members. As we describe

Example 2. *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.*

Example 3. *Consider a characteristic function game (N, U), where for each coalition $C \subseteq N$, the set $U(C) \subseteq \mathbb{R}^{|C|}$ specifies the payoff vectors that coalition C can feasibly secure if it forms. An alternative is a pair $a = (\pi, u)$, where π is a partition of N and $u \in \mathbb{R}^n$ is a payoff vector satisfying $u_C \in U(C)$ for every coalition $C \in \pi$. Payoffs are projections of the alternative onto its second coordinate space, that is, $v((\pi, u)) = u$. For every coalition C and alternative a, the effectivity correspondence $E_C(a)$ specifies the set of alternatives to which coalition C may move.*[3]

Outcomes, Histories, and Plans. We develop our notation recursively. A plan specifies a default alternative, say a, at the beginning of period $t = 0$. This default is chosen if no coalition blocks it, in which case we record the stage-game outcome as $(a, \emptyset)$. If coalitions $\{C_1, \ldots, C_k\}$ block the default, and their moves result in the alternative a', we record the stage-game outcome as $(a', \{C_1, \ldots, C_k\})$. Based on the outcome at $t = 0$, a plan specifies a default at $t = 1$, and the game continues recursively.[4]

Proceeding abstractly, let $\mathcal{P}$ be the set of all partitions over players, and define $\mathcal{B} := \{B \subseteq 2^N : B \subseteq \pi \text{ for some } \pi \in \mathcal{P}\}$, so that each B in $\mathcal{B}$ is a collection of disjoint coalitions. We denote the set of stage-game outcomes by $\mathcal{O} := A \times \mathcal{B}$; an outcome specifies an alternative a and a collection of disjoint blocking coalitions. At the beginning of period t, the history $h := (a^\tau, B^\tau)_{\tau=0}^{t-1}$ records the stage-game outcomes up to the start of period t. We denote the set of all t-period histories by $\mathcal{H}^t$ for $t \geq 1$. The set of all histories is $\mathcal{H} := \bigcup_{t=0}^{\infty} \mathcal{H}^t$, where $\mathcal{H}^0 = \{\emptyset\}$. A *plan* $\sigma : \mathcal{H} \to A$ specifies a default 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

later, we develop a coalition-proof variant of our solution concept in the Supplementary Appendix that imposes stronger requirements on this margin.

[3] Special cases of this example include the Roommates Problem and two-sided matching.

[4] We assume that coalitional blocking is directly observable so as to hew closely to repeated games with perfect monitoring, which we view to be the natural starting point. In some settings, the identity of a blocking coalition is implied by the chosen alternative. But, in other settings (e.g., matching), the chosen alternative alone might not be enough to encode who initiated the block. For instance, in the Roommates Problem, if the alternative $ab|c$ is blocked and we only record that $a|b|c$ is chosen instead, it does not distinguish which of Ann and Bella blocked.

$[0, 1)$ is a common discount factor.[5] After a history h, a plan σ recursively specifies a path of default alternatives and we denote by $U_i(h|\sigma)$ player i's payoff from that path.[6]

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 $\widetilde{v} = v(a)$. For example, in the Roommates Problem, $ab|c$ fails to be a core alternative because Bella and Carol together can profitably block it.

We build on this notion in the repeated game: when players contemplate blocking the alternative $\sigma(h)$ specified by plan σ at history h, they care not only about their instantaneous payoffs but also about how their choices today affect future behavior.

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

$$(1-\delta)v_i(a') + \delta U_i(\, h, (a', \{C\}) \mid \sigma) > U_i(h \mid \sigma).$$

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 "self-enforcing" plan in that, given the continuation play, no coalition finds it profitable to block.[7] In the language of self-generation, the alternative specified at each history is *enforced* by continuation promises that themselves are credible given that the requirement is imposed at every history (including those off-path). Thus, the continuation of a PCE at every history must itself be a PCE. This recursive form implies that the set of PCE supportable payoffs is amenable to dynamic programming

[5] In this repeated game, the history affects future payoffs only through the continuation plan; the alternative chosen in period t has no direct bearing on future payoffs or the set of feasible alternatives. In this way, our setup departs from the stochastic games studied by Konishi and Ray (2003) and Gomes and Jehiel (2005) in which the alternative chosen at the end of period t becomes the default in period $(t+1)$ and thereby directly restricts the set of feasible alternatives in that period.

[6] To be explicit, after a period-t history h, the plan σ specifies a default alternative $\sigma(h)$ for period t. If no coalition blocks this default, the resulting period-$(t+1)$ history is $(h, \sigma(h), \emptyset)$, where $\emptyset$ indicates that no blocking occurred in period t. Given this extended history, the plan then specifies the next default alternative $\sigma(h, \sigma(h), \emptyset)$ for period $(t+1)$. Iterating this construction yields the path of default alternatives induced by plan σ after history h: $(\sigma(h), \sigma(h, \sigma(h), \emptyset), \ldots)$.

[7] An alternative definition of profitable blocking might stipulate that each coalition member gains weakly and at least one does so strictly. This modification would not affect our results.

à la Abreu, Pearce, and Stacchetti (1990). If $\delta = 0$, PCEs implement only core alternatives of the stage game. Hence, a PCE may not exist if the stage game has an empty core and players are sufficiently myopic. If the core is nonempty, then a plan that specifies a core alternative a^* after every history is a PCE for every $\delta \geq 0$.[8]

Before characterizing PCE, we briefly compare it to other solution concepts. Fundamentally, a PCE captures the idea that coalitions can block through temporary agreements but coalition partners cannot commit in advance to the agreements they will sign in the future. By contrast, Strong Nash Equilibrium (Aumann 1959) and Strong Perfect Equilibrium (Rubinstein 1980) allow deviating coalitions to commit to a strategy in the repeated game, including how they will behave in every future period. Relative to those concepts, PCE therefore takes a conservative stance on the agreements that blocking coalitions can form.

From a different perspective, however, PCE may endow deviating coalitions with too much power, since a blocking coalition need not worry about further deviations by a subcoalition in the same period. To address this concern, we formalize perfect coalition-proof equilibrium (PCPE) in the Supplementary Appendix, combining the ideas above with coalition proofness (Bernheim, Peleg, and Whinston 1987). Although the resulting solution concept is distinct, we find that it yields identical coalitional minmaxes and supports the same payoff set when players are patient.

3 What Payoffs Are Supported by PCE?

3.1 Coalitional Minmaxes

In the introduction, we mentioned how coalitions can withstand punishments if they share aligned interests, which in turn limits the scope of PCE-supportable payoffs. We illustrate this phenomenon using the common-interest game depicted in Table 2(A), where we adopt the specification of coalitional moves stipulated in Example 1.

This game has a unique PCE, which prescribes the efficient action profile (U, L) at every history guaranteeing each player a payoff of 1. To see why, let $\underline{w}$ denote the infimum of the normalized discounted payoffs from all PCEs, and consider an arbitrary PCE in which each player accrues $w \in [0, 1]$. Since the continuation of a PCE at any

[8] The Bondareva-Shapley Theorem (Bondareva 1963; Shapley 1967) establishes that a cooperative game has a nonempty core if it is balanced. Thus, under these conditions, a PCE is also guaranteed to exist for every $\delta \geq 0$.

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

(A)

	L	R
U	$1,1$	$-\epsilon,\epsilon$
D	$0,0$	$0,0$

(B)

TABLE 2. Payoffs in (A) are perfectly aligned while those in (B) are slightly misaligned ($\epsilon > 0$).

history must itself be a PCE, if the pair blocks the alternative in the first period 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 $\underline{w} \geq 1$.

In this common-interest game, there is 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 by SPE of the repeated game. It turns out that the complete alignment of preferences is both sufficient and necessary for this gap. We find that coalitions of perfectly "like-minded" players can guarantee themselves a high coalitional payoff across all PCEs, regardless of discount factors. But misaligning preferences ever so slightly disrupts coalitional power.

For instance, our analysis shows that in the stage game of Table 2(B), PCE can support payoffs arbitrarily close to 0 when players are patient. We discuss why the logic used for the common-interest game does not apply here. Let $\underline{w}_i$ denote the infimum of player i's PCE payoffs. In a PCE, the pair do not block and choose (U, L) if at least one of them finds it unprofitable. Hence, a PCE payoff (w_1, w_2) must satisfy at least one, but not necessarily both, of the following inequalities:

$$w_1 \geq (1-\delta)+\delta\underline{w}_1 \quad \text{or} \quad w_2 \geq (1-\delta)+\delta\underline{w}_2.$$

Because players' payoffs are not perfectly aligned, there exists a sequence of PCE payoff profiles (w_1, w_2) such that $w_1 \to \underline{w}_1$ while w_2 stays bounded away from $\underline{w}_2$, and along this sequence the first inequality fails but the second holds. In these PCE, player 1 would profit from blocking with player 2 but player 2 does not. By using continuation play that punishes one player and not the other, such PCE can push players' payoffs below 1 without inviting a coalitional deviation. In contrast, both players are simultaneously rewarded and punished in the common-interest game, which makes it impossible to push payoffs below 1.

Generalizing these ideas, the right formulation of alignment here uses Abreu, Dutta,

and Smith (1994)'s notion of *equivalent utilities*: players i and j have equivalent utilities 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 utilities are not equivalent. We partition the set of players on this basis; let $C(i)$ be the set of players with whom player i shares equivalent utilities.

This set leads to player i's *coalitional minmax*, namely the lowest payoff that she can be pushed down to when coalition $C(i)$ collectively best responds.

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

The pure minmax is relevant because plans are pure, specifying an alternative following every history.[9] Using $\mathcal{V}$ to denote the convex hull of stage-game payoffs, we define $\mathcal{V}_{CR} := \{v \in \mathcal{V} : v_i > \underline{v}_i^\circ \text{ for every } i = 1, \ldots, n\}$ as the set of *strictly coalitionally rational payoffs*.[10] We distinguish this from the 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 does not share equivalent utilities with 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 equivalent utilities. The Roommates Problem lies in this class as do non-cooperative games that satisfy the NEU condition or full-dimensionality.

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 the coalitional minmax, $\underline{v}_i^\circ$, as relevant when coalitions can block. The second part shows that every strictly coalitionally rational

[9] The term above is also well-defined as A is compact, $v(\cdot)$ is continuous, and E_C is continuous and compact-valued. Assuming that E_C is monotone in C simplifies the expression; absent monotonicity, the coalitional minmax is $\underline{v}_i^\circ := \min_{a \in A} \max_{C \subseteq C(i)} \max_{a' \in E_C(a)} v_i(a')$. 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.

[10] 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).

payoff vector can be supported if players are sufficiently patient. An ancillary implication of this result is that, if $\mathcal{V}_{CR}$ is nonempty, PCE are guaranteed to exist in the repeated game when players are patient, even for stage games with an empty core.

We compare Theorem 1 to the folk theorem for SPE in repeated games with perfect monitoring. For this comparison, suppose that no two players share equivalent payoffs. Then Theorem 1 has the same implication as the folk theorems 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, a PCE is robust to both coalitional and individual deviations. Thus, when players are patient and have misaligned preferences, deterring coalitional deviations is no harder than deterring individual deviations.

Why? The proof constructs punishments that crack coalitions. Since a coalition blocks only if all of its members agree, it suffices to make just one member unwilling to participate. We do so by singling out and punishing only one member of each coalition as though she were the sole deviator, while granting amnesty to the rest. Specifically, consider the construction in Fudenberg and Maskin (1986) in which individual deviations are deterred by minmaxing a player before shifting to her player-specific punishment; such a construction is feasible whenever players have non-equivalent utilities. We adapt that scheme by assigning to each history and each (non-singleton) coalition C a designated member—a "scapegoat"—who is punished as if she were the sole deviator. This punishment makes the scapegoat unwilling to block with coalition C. By associating each coalition with a scapegoat, a PCE can ensure that no coalition finds it profitable to block whenever players are sufficiently patient.[11]

This divide-and-conquer scheme fails if all members of a coalition C share equivalent utilities. A higher minmax then applies for those players. To see why, consider such a coalition C and suppose towards a contradiction that some PCE σ could push the payoffs of its coalition members below their coalitional minmax. Members of coalition C could guarantee their coalitional minmax if they could commit to a long-run plan in which they collectively best respond to the alternative specified by the plan after every history. Because payoffs are equivalent, all their gains and losses move in sync so we can effectively treat coalition C as a unitary actor. Using an argument like that for the One-Shot Deviation Principle, it follows that there then exists a history at which

[11] We observe that a coalition can be cracked even if all coalition members share the same *ordinal* rankings over alternatives; a PCE can nevertheless create player-specific punishments and isolate a coalition member as a scapegoat through its choice of how to sequence alternatives.

coalition C could profitably block, which precludes σ from being a PCE.

Thus, Theorem 1 highlights that a coalition withstands the force of repeated games if and only if its members share *completely* aligned preferences. On the one hand, this conclusion might appear to emphasize a "knife-edge" consideration. Yet, it applies to common-interest games, a commonly studied setting, in which we find that only efficient action profiles are chosen in a PCE.[12] More importantly, as we show in our study of strongly symmetric equilibria (Theorem 3) and secret side-payments (Theorem 5), this theme of alignment emerges even if players do not have equivalent utilities.

3.3 Structural Properties

3.3.1 When Do Stationary PCEs Suffice?

Herein, we highlight how, in a rich class of games, all PCE payoffs can be achieved using stationary PCE. A plan σ is *stationary* if following every history h, the plan σ specifies the same alternative in each period so long as the plan is not blocked, i.e., $\sigma(h, (\sigma(h), \emptyset)) = \sigma(h)$. After a block, a stationary plan may transition to a different alternative but would prescribe that alternative in every subsequent period. We prove that it suffices to study stationary PCE in *convex* games with *default-independent power*. As the latter notion is new, we turn to it first.

For a coalition C and alternative a, let $v_C(a) := \{(v_i(a))_{i \in C}\}$ be the projection of $v(a)$ onto C's payoff space. Let $V_C(a) := \{v_C(a') : a' \in E_C(a)\}$ be the set of $|C|$-dimensional payoff vectors that coalition C can obtain from blocking that alternative.

Definition 3. *A stage game exhibits* ***default-independent power*** *if for every coalition C, there exists a $|C|$-dimensional payoff set D_C such that $V_C(a) = D_C \cup \{v_C(a)\}$ for every alternative a.*

Definition 3 asserts that the payoffs a coalition can achieve through blocking an alternative does not depend on that alternative. Although this notion rules out strategic form games (Example 1), several well-studied coalitional games exhibit this property.

For instance, consider any characteristic function game studied in the classical cooperative game theory literature (Example 3). In such a game, if a coalition blocks a default partition π, the set of payoffs it can achieve is independent of the partition. An

[12] For these games, the coalitional minmax therefore departs from—and is generally higher than—Wen (1994)'s *effective minmax* for SPE, which would be $\min_{a \in A} \max_{j \in C(i)} \max_{a'_j \in A_j} v_i(a_{-j}, a'_j)$.

important application is matching without externalities; there too, the set of payoffs that a coalition obtains from blocking an assignment does not hinge on the assignment.

Our result identifies how stationary PCEs suffice in default-independent games if the stage game is *convex*, i.e., $\{v(a) : a \in A\}$ is a convex set.[13]

Theorem 2. *If the stage game is convex and exhibits default-independent power, then for every $\delta \geq 0$, the set of PCE-supportable payoffs coincides with 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 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 strategic 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 $V^S := \{v(a) : a \in A^S\}$ denotes the associated symmetric payoffs. Given that V^S is compact and totally ordered, a maximal element exists denoted $\widehat{v}$, the highest symmetric payoff.

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

[13] 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.

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

This result reflects a collapse of intertemporal incentives: a strongly symmetric PCE exists if and only if the highest symmetric payoff lies in the core and thus can be supported without any carrots or sticks. This requirement is demanding, excluding games like the prisoner's dilemma or oligopolistic collusion. In such games, a PCE must instead rely on asymmetric punishments.

	E	S
E	$2,2$	$-1,3$
S	$3,-1$	$0,0$

(A)

	E	S
E	$2,2$	$-1,1$
S	$1,-1$	$0,0$

(B)

TABLE 3. (A) is a Prisoner's Dilemma, which fails the condition of Theorem 3; (B) is a strategic form game that satisfies it.

We illustrate this finding in Table 3. In the Prisoner's Dilemma depicted in (A), the highest symmetric payoff $(2,2)$ can only be generated by the action profile EE, which is not a core alternative. Hence, the game admits no strongly symmetric PCE for any discount factor. Grim trigger, a strongly symmetric plan (and subgame perfect equilibrium if $\delta \geq 1/3$), fails to be a PCE because players would find it profitable to block during the punishment phase. By contrast, in the game depicted in (B), the payoff $(2,2)$ is in the core; Theorem 3 asserts then that the *unique* strongly symmetric PCE prescribes EE after every history.

The key step in the proof of Theorem 3 shows that strongly symmetric PCE payoffs, if they exist, must be unique. The logic is that once punishments are required to be symmetric, the game effectively becomes a common-interest game: unless players are already at the highest symmetric payoff profile $\widehat{v}$, they would profitably block and move to it.[14] Therefore, the continuation payoffs for any strongly symmetric PCE are pinned down independently of history, which implies that such equilibria can enforce only those actions that are myopically optimal.

[14]Suppose $\underline{v}$ is the lowest strongly symmetric PCE profile. Then players do not profitably block to an action that generates the payoff $\widehat{v}$ only if $\underline{v} \geq (1-\delta)\widehat{v} + \delta\underline{v}$, which implies that $\underline{v} \geq \widehat{v}$.

4 Do Transfers Align Incentives?

4.1 Transferable Utility Framework

We turn to the question of whether transferable utility aligns incentives. We model transfers separately from alternatives because we vary their observability. Section 4.2 studies publicly observed transfers and Section 4.3 models secret side-payments. Like the bonus in relational contracting (e.g., Levin 2003), transfers are discretionary and not contracted upon ahead of time. In introducing transfers, our primary goal is to evaluate the extent to which coalitions can use transfers to profitably block.

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 blocking coalitions (if any), and the chosen transfers. The set of stage-game outcomes is $\overline{\mathcal{O}} := A \times \mathcal{B} \times \mathcal{T}$. 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 default alternative and transfers in $\sigma(h)$. We modify the definition of $U_i(h|\sigma)$ to reflect the influence of transfers.

By blocking the default (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 coalition C's block surprises players outside the coalition, who therefore make transfers T_{-C} as was specified by the plan. 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 default transfers matrix T, the transfers made by players outside of C is $\chi^C(T)$ where $\chi^C : \mathcal{T} \to \mathcal{T}_{N \setminus C}$ satisfies:

(a) For each bounded set $S \subseteq \mathcal{T}$, the image $\chi^C(S) \subseteq \mathcal{T}_{N\setminus C}$ is also bounded.

(b) 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 $i, j \in N\setminus C$, and $\chi^C_{ij}(T) = 0$ for all $i \in N\setminus C$ 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: $\big\{U(h|\sigma) : h \in \overline{\mathcal{H}}\big\}$ 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:

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

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.*

We omit a formal proof of Theorem 4 because it follows as a special case of Theorem 5 in Section 4.3. Comparing Theorems 1 and 4 reveals that rather than aligning coalition members' incentives, public transfers 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 the post-transfer payoffs of coalition members so as to deter blocking alongside all transfer choices. We illustrate this logic in Example 4 below. Moreover, players i and j have misaligned interests (or non-equivalent utilities) once one can pay transfers to the other. This misalignment allows us to construct player-specific punishments.

Example 4. Consider a setting in which, in each period, a single firm F procures up to 10 units of perfectly divisible goods each period from two suppliers, S_1 and S_2, each of whom faces zero marginal cost. Each unit is worth \$1 to the firm, and prices can be tailored to each supplier. Thus, the total surplus each period is \$10.

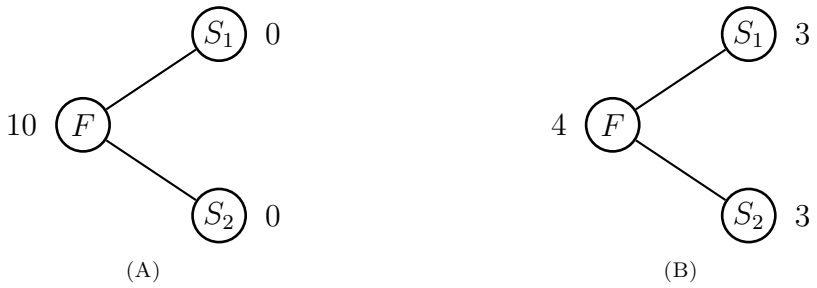


FIGURE 2. (A) shows the unique core of the stage game, where the firm keeps the entire surplus; (B) shows an allocation where the suppliers retain some surplus.

Because the firm F can meet its demand through either supplier individually, each pair $\{F, S_1\}$ and $\{F, S_2\}$ must secure a total surplus of 10 in the stage-game core. Otherwise, at least one pair could profitably block the allocation by dividing the share of the other supplier between themselves.[15] Thus, the firm captures the entire surplus in the stage-game core, as depicted in Figure 2(A).

[15] We consider the specification of transfers from Assumption 1 in which players outside a blocking coalition make no transfers to members of a blocking coalition.

A PCE captures how repeated play could distribute surplus to the suppliers. Suppose $\delta = 2/3$. Then the payoffs in Figure 2(B) can be supported by a PCE using a modified core-reversion scheme: players play the allocation in (B) indefinitely unless some pair blocks. Should the coalition $\{F, S_1\}$ block and the resulting allocation give F a stage-game payoff of $x \geq 4$, then play reverts permanently to the stage-game core in every subsequent period; otherwise, their block is ignored and play continues as in (B). A similar response applies to blocking by $\{F, S_2\}$. Observe that this scheme conditions the continuation play not only on the identity of the blocking coalition but also on the distribution of surplus within it. Finetuning punishments in this manner allows the scheme to ensure that at least one player is made worse off in each (potential) blocking coalition. For example, if $\{F, S_1\}$ contemplates a block where $x \geq 4$, then S_1 suffers from participating in this block because her payoff can be no more than $(1-\delta)(6) + \delta(0)$, which is less than 3 given that $\delta = 2/3$; by contrast, if $x < 4$, the scheme deters firm F from blocking because doing so leads to a lower payoff today without affecting its continuation payoff. ■

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. These contracts are often confidential, a point to which we return in Section 5 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; denote the set of all non-singleton coalitions that can by $\mathcal{S} \subseteq \mathcal{C}$. 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 $o = (a, B, T) \in \overline{\mathcal{O}}$, the*

public transfers, denoted T^p, exclude those made within any blocking coalition in $\mathcal{S}$:

$$T^p_{ij} = \begin{cases} \# & \text{if } \exists\, C \in \mathcal{S} \cap B \text{ such that } \{i,j\} \subseteq C, \\ T_{ij} & \text{otherwise.} \end{cases}$$

The public component of o, denoted by o_p, is $o_p := (a, B, T^p)$. For any history $h = (o^\tau)_{\tau=0}^t$, the public component of h is $h_p = (o^\tau_p)_{\tau=0}^t$.

The definition stipulates that if coalition C can make secret transfers, then transfers within it are not recorded in the public history whenever it blocks; instead, those transfers are recorded as #, indicating that they are missing. In this setting with imperfect public monitoring, we consider the analog 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, which in turn limits what a public PCE can support. In particular, a public PCE can condition only on the identity of a blocking coalition and not on how its members divide their surplus. Before proceeding to our results, we illustrate this tension by returning to Example 4.

Example 4 (Continued). The PCE supporting the allocation in Figure 2(B) relies on continuation play that conditions on transfers within the blocking coalition. Suppose transfers are instead secret: if the coalition $\{F, S_i\}$ blocks, transfers between F and S_i are unobserved by the other seller S_{-i}. We show that, in this setting, simple core reversion cannot support the allocation in (B). Consider a plan in which, on-path, the players obtain the payoffs in (B) but any block triggers permanent reversion to the stage-game core. Firm F and seller S_1 can then find it profitable to block, generate the full surplus of 10, and use an appropriate choice of transfer to make sure that each comes out ahead, given continuation play.[16] Therefore, this plan is not a PCE.

Would other plans work? The challenge is that every plan can be defeated by some choice of transfers between F and S_1. In particular, Theorem 5 below implies that every public PCE must deliver a total payoff of 10 to each of the coalitions $\{F, S_1\}$

[16] E.g., firm F could transfer all of the current-period surplus to S_1, which results in a normalized discounted payoff of $(1-\delta)(10) + \delta(0) > 3$ to S_1 and $(1-\delta)(0) + \delta(10) > 4$ to F (recall $\delta = 2/3$).

and $\{F, S_2\}$. Only the stage-game core, wherein the firm captures the entire surplus, is compatible with this division of surplus. ■

To formalize the general conclusion, suppose a coalition $C \in \mathcal{S}$ acted as a unitary actor that maximizes the total utility $\sum_{i \in C} v_i(\cdot)$ equipped 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

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

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. One might view these high coalitional minmaxes as conveying an "anti-folk" flavor. Indeed, in the example above and both applications, secret transfers reduce the supportable payoff set to the core of the stage game.

To see why Theorem 5 holds, we first explain why each coalition $C \in \mathcal{S}$ can guarantee its minmax for every $\delta \geq 0$. 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 a payoff no lower than $\underline{u}_C$ by executing a long-run plan where it best responds to the default alternative in each period, blocking when necessary. An argument similar to the one-shot deviation principle then establishes that the total utility of members of coalition C must increase by blocking once at some history h. By apportioning that

gain across the members of C through secret side-payments, it can then be assured that each member simultaneously profits from the block at that history without affecting continuation play.[17] Such a block is then profitable, which means σ is not a PCE.

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, which corresponds to the special case in which $\mathcal{S}$ is empty. Therein, we cracked 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 vary the identify of the scapegoat with the transfers within 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 An Application to Labor Market Matching

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 application, 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. We use $C \subseteq N$ to denote a generic coalition. We call a coalition *essential* if it comprises a single firm and a nonempty set of workers; let $\mathcal{E} := \left\{\{f\} \cup W : f \in \mathcal{F}, W \subseteq \mathcal{W}, W \neq \emptyset\right\}$ be the set of all essential coalitions.

Each firm can hire multiple workers. An *assignment* is a mapping $\phi : \mathcal{F} \cup \mathcal{W} \to 2^{\mathcal{F} \cup \mathcal{W}}$ such that (i) every worker w is assigned to at most one firm; (ii) every firm f is assigned to a (potentially empty) set of workers; and (iii) $w \in \phi(f)$ if and only

[17] In addition to secrecy, this step uses the one-to-one transferability of utility. If utility were costly to transfer, then the coalition C as a whole may be on a higher utility frontier from blocking but have no way to apportion those gains so that each player profits.

if $f \in \phi(w)$.[18] 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 $\mathcal{M} := \{(\phi, T) \in A \times \mathcal{T} : T_{ij} \neq 0 \text{ only if } i \in \phi(j)\}$.

Each firm f has a revenue function v_f defined over all subsets of $\mathcal{W}$, with $v_f(\emptyset)$ normalized to 0; similarly, worker w has a pre-renumeration utility function v_w defined over singleton or empty subsets of $\mathcal{F}$, where the payoff of being unemployed, $v_w(\emptyset)$, 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 $u_i(\phi, T) := v_i(\phi) + \sum_{j \neq i} T_{ji} - \sum_{j \neq i} T_{ij}$.

If a coalition C blocks, its members sever their links with players outside C and may arbitrarily rearrange the matching within C.[19] Formally, for all $\phi \in A$,

$$E_C(\phi) = \{\phi' \in A : \phi'(i) \subseteq C \text{ for all } i \in C, \text{ and } \phi'(i) = \phi(i) \backslash C \text{ for all } i \notin C\}.$$

This formulation specifies that if coalition C blocks assignment ϕ, any match between agents both outside C remains intact while those involving a member of C and a nonmember is severed. Because 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 of 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.

Definition 8. *A matching (ϕ, T) is **profitably blocked by coalition** C if there exists an 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 $(\phi', T'_C, \chi^C(T))$:*

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

*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.*

[18] Formally, the first two conditions stipulate that $\phi(w) \subseteq \mathcal{F}$ with $|\phi(w)| \leq 1$ and that $\phi(f) \subseteq \mathcal{W}$. Throughout, we write $\phi(i) = \emptyset$ if player i is unassigned.

[19] KC82 study a more restrictive notion of blocking, allowing only essential coalitions to block. Our results apply also to that setting.

KC82 show that when firms' revenue functions satisfy the *gross substitutes* condition, formally defined in the footnote,[20] the core is nonempty. We impose the same assumption and, although we permit blocking by non-essential coalitions, we prove that the resulting core remains unchanged.

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 specifies a stage-game matching at every history.[21] 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 through 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. Using Theorem 4's argument, we show that any feasible and individually rational payoff can be supported when players are patient.

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.*

While we assume gross substitutes to maintain comparability with KC82, Proposition 1 itself does not require gross substitutes. Even if the stage-game core is empty, stable schemes exist in the repeated game if players are sufficiently patient.

Now suppose each firm can hire and offer private wage terms to a group of workers. Recall that a coalition C is essential if it comprises a single firm and a nonempty set of workers, and $\mathcal{E}$ is the set of all essential coalitions. For the next result, we suppose

[20] 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 $E(\widehat{W}, T_f, T'_f) \subseteq \widehat{W}'$.

[21] Our approach models firms and workers who interact via spot contracts and studies how intertemporal incentives discipline which spot contracts are signed over time. An alternative modeling approach might allow firms and workers to sign long-term contracts. Diamantoudi, Miyagawa, and Xue (2015) and Zhang and Zheng (2016) study two-period matching models with commitment, emphasizing how the availability of long-term contracts influences equilibrium outcomes.

that 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 $\mathcal{E} \subseteq \mathcal{S}$. For every $\delta \geq 0$, a public PCE supports a discounted payoff vector if and only if that payoff vector is in the stage-game core, $\mathcal{K}$.*

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 the same logic as in 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, $v_f(W) + \sum_{w \in W} v_w(f)$, 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.

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, an assignment ϕ^* that maximizes total social surplus is found by greedily assigning workers to firm slots in order of their contribution to total surplus. 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; every efficient assignment must then fills "slots" $\{(f, l) : \rho(f, l) \geq \max\{0, \eta(L)\}\}$ leaving all others vacant. We use A^* to denote the set of efficient assignments and note that all its elements are identical up to a relabeling of workers. Finally, we assume that it would be inefficient for a single firm to hire all workers so that each firm faces some competition.

Let $\mathcal{U}^\circ := \text{co}\{u(\phi^*, T) : (\phi^*, T) \in \mathcal{M} \text{ and } \phi^* \in A^*\}$ denote the efficient frontier of the set of feasible payoff profiles. Within this set, let $\mathcal{U}^+ := \{\widetilde{u} \in \mathcal{U}^\circ : \widetilde{u}_i > 0 \text{ for all } i \in \mathcal{F} \cup \mathcal{W}\}$ denote those surplus divisions under which every player obtains more than her minmax. We also consider the surplus divisions in which each worker obtains a net utility of approximately the "marginal product" of the last employee in the economy while firms are residual claimants. 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:

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

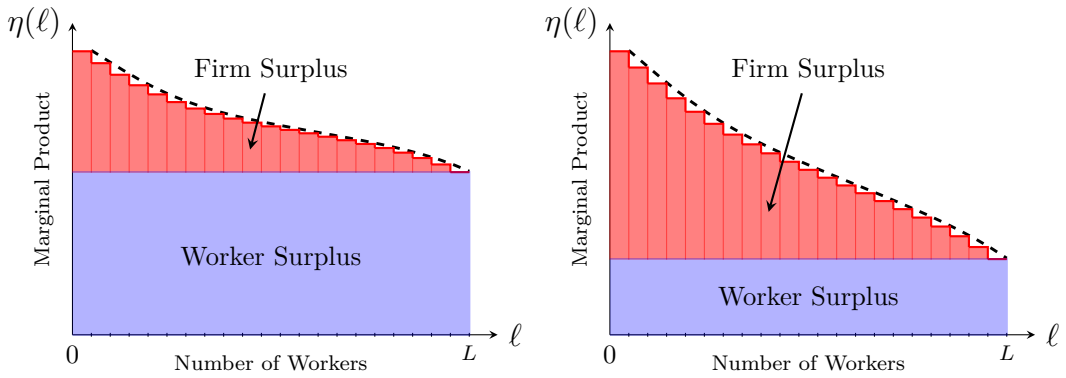
We show that $\mathcal{K} = \mathcal{U}^*$, which yields the following conclusion.

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 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 PCEs 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 the



(A) Marginal product falls slowly (B) Marginal product falls quickly

FIGURE 3. (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.

continuation play in which poaching today triggers a "salary war" tomorrow.

We depict this prediction in Figure 3: (A) shows the distribution of worker and firm 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 who 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 $\eta_2(\ell) - \eta_2(\ell+1) \geq \eta_1(\ell) - \eta_1(\ell+1)$ for every ℓ in $\{1, \ldots, L\}$.*[22]

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

[22] If the inequality is strict for some ℓ, then we add the qualifier "strictly."

the second derivative of the total product being more negative in M_2.[23]

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 and workers' surplus; let $\Pi_i^{\mathcal{F}}$ 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 $\succcurlyeq_{SSO}$.[24]

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

Although more steeply decreasing returns reduce both total surplus and worker surplus under private wages, Proposition 4 shows that this effect is more pronounced for workers. Hence, as seen in (c), the residual surplus captured by firms is actually higher in M_2 than in M_1. If wage transparency enables workers to capture firms' profits, then workers have more to gain (and less to lose) in M_2 than M_1.

6 Conclusion

This paper develops a portable framework for coalitional repeated games, which enables us to evaluate the role of dynamic incentives in coalitional behavior across a range of settings. Our analysis uncovers the importance of alignment: history dependence keeps coalitions in line if coalition members' interests are even slightly misaligned. Simple scapegoat schemes then deter coalitional deviations. However, if players in a coalition

[23] 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.

[24] For sets $A, B \subseteq \mathbb{R}$, $A \succcurlyeq_{SSO} B$ if for every $a \in A$ and $b \in B$, $\max\{a, b\} \in A$ and $\min\{a, b\} \in B$.

have completely aligned interests, they can secure a higher minmax payoff by effectively acting as a unitary agent.

This perspective delivers additional insights. Strongly symmetric schemes do not deter coalitional deviations, thereby 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 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. In the Supplementary Appendix, we also study repeated negotiations and show that history dependence can counter the tendency of veto players to become de facto dictators; we show therein that secret transfers restore their dictatorial power.

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.

Ali, S. Nageeb and Ce Liu. 2026a. "On Conservative Stable Standard of Behavior and Perfect Coalitional Equilibrium." Working Paper.

———. 2026b. "Supplement to 'Coalitions in Repeated Games'." *Econometrica Supplementary Material* .

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. 2026. "Early-Career Discrimination: Spiraling or Self-Correcting?" *Review of Economic Studies* .

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

Bernheim, B. Douglas, Bezalel Peleg, and Michael D. Whinston. 1987. "Coalition-proof nash equilibria i. concepts." *Journal of Economic Theory* 42 (1):1–12.

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.

Bondareva, Olga N. 1963. "Some Applications of Linear Programming Methods to the Theory of Cooperative Games." *Problemy Kibernetiki* 10:119–139. In Russian.

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

Cole, Harold L, Dirk Krueger, George J Mailath, and Yena Park. 2024. "Trust in Risk Sharing: A Double-Edged Sword." *Review of Economic Studies* 91 (3):1448–1497.

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.

Damiano, Ettore and Ricky Lam. 2005. "Stability in Dynamic Matching Markets." *Games and Economic Behavior* 52 (1):34–53.

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

Diamantoudi, Effrosyni, Eiichi Miyagawa, and Licun Xue. 2015. "Decentralized matching: The role of commitment." *Games and Economic Behavior* 92:1–17.

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.

Gale, David and Lloyd S. Shapley. 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly* 69 (1):9–15.

Genicot, Garance and Debraj Ray. 2003. "Group Formation in Risk-Sharing Arrangements." *Review of Economic Studies* 70 (1):87–113.

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.

Greenberg, Joseph. 1989. "An application of the theory of social situations to repeated games." *Journal of Economic Theory* 49 (2):278–293.

———. 1990. *The theory of social situations: an alternative game-theoretic approach*. Cambridge University Press.

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.

Levin, Jonathan. 2003. "Relational incentive contracts." *American Economic Review* 93 (3):835–857.

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

Liu, Ce, Ziwei Wang, and Hanzhe Zhang. 2025. "Self-Enforced Job Matching." *American Economic Journal: Microeconomics* .

Mailath, George and Larry Samuelson. 2006. *Repeated Games and Reputations*. New York, NY: Oxford University Press.

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

Moulin, Herve and Bezalel Peleg. 1982. "Cores of effectivity functions and implementation theory." *Journal of Mathematical Economics* 10 (1):115–145.

Ray, Debraj and Rajiv Vohra. 2015. "The Farsighted Stable Set." *Econometrica* 83 (3):977–1011.

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.

Shapley, Lloyd S. 1967. "On Balanced Sets and Cores." *Naval Research Logistics Quarterly* 14 (4):453–460.

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

Vartiainen, Hannu. 2011. "Dynamic Coalitional Equilibrium." *Journal of Economic Theory* 146 (2):672–698.

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

Zhang, Mu and Jie Zheng. 2016. "Multi-period Matching with Commitment." Working

Paper.

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. 13

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

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

Since players in $C(i^*)$ have equivalent payoffs, for each $j \in C(i^*)$ there exists $\lambda_{ji^*} > 0$ and $\mu_{ji^*} \in \mathbb{R}$ such that $v_j(a) = \lambda_{ji^*} v_{i^*}(a) + \mu_{ji^*}$ 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

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

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

Proof of Theorem 1.

Part 1: *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 default 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 equivalent utilities, $U_j(\emptyset|\alpha) > U_j(\emptyset|\sigma)$ for all $j \in C(i^*)$, so α is a profitable blocking plan for coalition $C(i^*)$.

Part 2: *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.*

Fix $v^* \in \mathcal{V}_{CR}$. First, observe that for any pair of players i, j such that $j \notin C(i)$, their payoffs satisfy the non-equivalent utilities (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 $v_j^{C(i)} > v_j^{C(j)}$ 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 equivalent 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, for all $i \in N$ and $a' \in E_{C(i)}(\underline{a}_{C(i)}^\circ)$, $v_i(a') \leq \underline{v}_i^\circ$. Observe that the reflexivity of effectivity correspondences implies that $\underline{a}_{C(i)}^\circ \in E_{C(i)}(\underline{a}_{C(i)}^\circ)$ and therefore, $v_i(\underline{a}_{C(i)}^\circ) \leq \underline{v}_i^\circ$.

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:

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

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

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 $\left\{(a^{S,\tau})_{\tau=0}^\infty : S \in \widehat{\mathcal{I}}\right\}$ 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 $\left|\left|v^S - (1-\delta)\sum_{\tau=t}^\infty \delta^{\tau-t} v(a^{S,\tau})\right|\right| < \epsilon$. We fix an

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

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

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

- $W := \left\{w(d,\tau) | d \in \widehat{\mathcal{I}}, \tau \geq 0\right\} \cup \{\underline{w}(S,\tau) | S \in \mathcal{I}, 0 \leq \tau < L(\delta)\}$ 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 any collection of blocking

coalitions $B \in \mathcal{B}$, let $\widehat{C}(B) = \cup_{C \in B} C$ denote their union. For states of the form $w(d, \tau)$, the transition is

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

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

$$\gamma\big(\underline{w}(S,\tau),(a,B)\big) = \begin{cases} \underline{w}(C(j^*),0) & \text{if } \widehat{C}(B) \not\subseteq S \text{ , where } j^* = \min(\widehat{C}(B) \backslash S) \\ \underline{w}(S,0) & \text{if } \widehat{C}(B) \subseteq S \text{ and } \widehat{C}(B) \neq \emptyset \\ \underline{w}(S,\tau+1) & \text{otherwise} \end{cases}$$

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

$$\gamma\big(\underline{w}(S,L(\delta)-1),(a,B)\big) = \begin{cases} \underline{w}(C(j^*),0) & \text{if } \widehat{C}(B) \not\subseteq S, \\ & \quad \text{where } j^* = \min(\widehat{C}(B) \backslash S) \\ \underline{w}(S,0) & \text{if } \widehat{C}(B) \subseteq S \text{ and } \widehat{C}(B) \neq \emptyset \\ w(S,0) & \text{otherwise} \end{cases}$$

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

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

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 $j^* = \min C$. 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 $(1-\delta)b + \delta V_{j^*}(\underline{w}(C(j^*),0)) = (1-\delta)b + \delta[(1-\delta^{L(\delta)}) v_{j^*}(\underline{a}^\circ_{C(j^*)}) + \delta^{L(\delta)} v^{C(j^*)}_{j^*}]$. The blocking is not profitable if the preceeding term is no more than $v^d_{j^*} - \epsilon$. We prove that this is true in two separate cases.

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

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

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

Now suppose that $d = C(j^*)$. The blocking payoff being less than $v_{j^*}^{C(j^*)} - \epsilon$ can be re-written as $(1-\delta)(b - v_{j^*}^{C(j^*)}) + \epsilon \leq \delta(1-\delta^{L(\delta)})(v_{j^*}^{C(j^*)} - v_{j^*}(\underline{a}^\circ_{C(j^*)}))$. As $\delta \to 1$, the LHS converges to ϵ. Because $\lim_{\delta \to 1}\delta^{L(\delta)} = \kappa$, the RHS converges to $(1-\kappa)(v_{j^*}^{C(j^*)} - v_{j^*}(\underline{a}^\circ_{C(j^*)}))$. 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 $\underline{a}^\circ_S$, 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

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

Because $v_i^S > \underline{v}^\circ_i \geq v_i(\underline{a}^\circ_S)$, it suffices to show that the inequality above holds at $\tau = 0$. Re-arranging terms yields that $(1-\delta)(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + (1-\delta)\delta^{L(\delta)}v_i^S \geq (1-\delta)\underline{v}^\circ_i$, and then dividing by $(1-\delta)$ yields $(1-\delta^{L(\delta)})v_i(\underline{a}^\circ_S) + \delta^{L(\delta)}v_i^S \geq \underline{v}^\circ_i$. Let us verify that this inequality holds for sufficiently high δ. Taking $\delta \to 1$ yields (1), which is true. Hence (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

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

We prove that this inequality is satisfied if δ is sufficiently high. Examining the LHS,

observe that for all τ such that $0 \leq \tau \leq L(\delta) - 1$,

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

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

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

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 $v_{j^*}(\underline{a}^\circ_{C(j^*)}) \leq \underline{v}^\circ_{j^*}$, the second weak inequality follows from $\widetilde{\kappa} \geq \kappa$ 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. ■

A.2 Proof of Theorem 2 on p. 16

Since the stage game exhibits default-independent power, for each $C \in \mathcal{C}$ there exists set $D(C) \subseteq \mathbb{R}^{|C|}$ such that $V_C(a) = D(C) \cup \{v_C(a)\}$ for all $a \in A$. 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 continuation value are the same. Accordingly, we define the stationary self-generation map as follows. Let $V^\circ := \{v(a) : a \in A\}$ denote the set of stage-game payoff profiles. For any set $Y \subseteq V^\circ \subseteq \mathbb{R}^n$, define

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

The self-generation map identifies the set of discounted payoffs that are supportable when continuation payoffs must lie in the set Y, so for any blocking coalition C contemplating 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 $Y \subseteq V^\circ$ 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 $y \in V^\circ$, 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 $y'[a, C] \in Y \subseteq V^\circ$, 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: $\sigma(a, B) = \widetilde{a}$ if B is either an empty set or comprises more than one coalitions; 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 $\tau = 0, \ldots, t-2$ 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 $y^{h^{t-1}}[a, C] \in Y \subseteq V^\circ$, 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 $\sigma(h^{t-1}, a, B)$ as follows: $\sigma(h^{t-1}, a, B) = \sigma(h^{t-1})$ if B is either an empty set or comprises more than one coalitions; 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 $\tau = 0, \ldots, t-1$ and $h^\tau \in \mathcal{H}^\tau$. This completes the induction step.

By property (i), σ is stationary, which implies 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 V° is convex, it follows that $\mathcal{U}(\sigma) \subseteq V^\circ$. 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, this is equivalent to the statement that for all $C \in \mathcal{C}$ and $z_C \in D(C)$, there exists $y' \in \mathcal{U}(\sigma)$ and $i \in C$ such that $y_i \geq (1-\delta)z_i + \delta y'_i$, which implies $\mathcal{U}(\sigma) \subseteq \Phi_\delta(\mathcal{U}(\sigma))$. The claim then follows from Lemma 3. ■

A.3 Proof of Theorem 3 on p. 17

A Preliminary Result. Let $\mathcal{V}^S$ denote the set of discounted payoff profiles from strongly symmetric PCEs. The following lemma is useful for proving Theorem 3.

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

Proof. Suppose $\mathcal{V}^S$ is nonempty. Since players accrue identical payoffs from symmetric action profiles, we have $v_i = v_j$ for all players i, j and $v \in \mathcal{V}^S$. Let $\widehat{x} := \max_{a \in A^S} v_1(a)$ be the highest feasible symmetric payoff, so $\widehat{v} = (\widehat{x}, \ldots, \widehat{x})$. Let $\underline{x} := \inf\{x : (x, \ldots, x) \in \mathcal{V}^S\}$ denote the lowest symmetric PCE payoff.

Consider a sequence $\left((x^k, \ldots, x^k)\right)_{k=1}^{\infty} \subseteq \mathcal{V}^S$ that converges to $(\underline{x}, \ldots, \underline{x})$ and let σ^k be the PCE that supports 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 $\underline{x} \geq \widehat{x}$. However, by definitions $\underline{x} \leq \widehat{x}$, so $\mathcal{V}^S = \{\widehat{v}\}$. ■

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 the default action profile $\sigma(h)$ must satisfy $v(\sigma(h)) = \widehat{v}$ for all $h \in \mathcal{H}$.

Take an arbitrary $\widehat{h} \in \mathcal{H}$ and let $\widehat{a} := \sigma(\widehat{h})$ be the default action profile. As argued above $\widehat{a}$ is symmetric and $v(\widehat{a}) = \widehat{v}$, so it remains to show that $\widehat{a}$ is a core alternative. Suppose otherwise, namely there exists coalition C and $a' \in E_C(\widehat{a})$ such that $v_i(a') > v_i(\widehat{a})$ for all $i \in C$. Then,

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

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

For the "if" direction: suppose $\widehat{v} = v(\widehat{a})$ for some symmetric core alternative $\widehat{a}$. A plan that specifies $\widehat{a}$ as default at all histories is a strongly symmetric PCE that supports the discounted payoff $\widehat{v}$. Uniqueness follows from Lemma 4. ■

A.4 Proof of Theorem 5 on p. 23

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

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

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 $C \in \mathcal{S} \cup \mathcal{N}$ has a profitable blocking plan.*

Proof. Consider a public plan σ from which coalition $C \in \mathcal{S} \cup \mathcal{N}$ 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

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

If $C \in \mathcal{N}$, we have already shown that σ is not a PCE, which completes the proof. If instead $C \in \mathcal{S}$, showing that C can profitably block σ at $\widehat{h}$ amounts to showing that

the total payoff can be divided so that every member of C is 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 $i \in C$,

$$(1-\delta)u_i\Big(\alpha(\widehat{h}),\ T^*\Big) + \delta U_i\Big(\widehat{h},\ \alpha(\widehat{h}),\ \{C\},\ \beta(\widehat{h})\Big|\sigma\Big) > U_i(\widehat{h}|\sigma). \tag{5}$$

Consider the two histories $h^1 := \big(\widehat{h},\ \alpha(\widehat{h}),\ \{C\},\ \beta(\widehat{h})\big)$ and $h^2 := \big(\widehat{h},\ \alpha(\widehat{h}),\ \{C\},\ T^*\big)$.

By the construction of T^* and the fact that $C \in \mathcal{S} \cup \mathcal{N}$, 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 $i \in C$, $U_i\big(\widehat{h},\ \alpha(\widehat{h}),\ \{C\},\ \beta(\widehat{h})\big|\sigma\big) = U_i\big(\widehat{h},\ \alpha(\widehat{h}),\ \{C\},\ T^*\big|\sigma\big)$. Inequality (5) can therefore be re-written as $(1-\delta)u_i\big(\alpha(\widehat{h}),\ T^*\big) + \delta U_i\big(\widehat{h},\ \alpha(\widehat{h}),\ \{C\},\ T^*\big|\sigma\big) > U_i(\widehat{h}|\sigma)$ for every $i \in C$, which 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 "$(\mathcal{S} \cup \mathcal{N})$-specific punishments" for all coalitions in $\mathcal{S} \cup \mathcal{N}$.

Lemma 6. *For any $u^* \in \mathcal{U}_{CR}(\mathcal{S})$, there exist $(\mathcal{S} \cup \mathcal{N})$-specific punishments $\{u^C : C \in \mathcal{S} \cup \mathcal{N}\} \subseteq \mathcal{U}_{CR}(\mathcal{S})$ such that $\sum_{i \in C} u^C_i < \sum_{i \in C} u^*_i$ for all $C \in \mathcal{S} \cup \mathcal{N}$, and $\sum_{i \in C} u^C_i < \sum_{i \in C} u^{C'}_i$ for all $C, C' \in \mathcal{S} \cup \mathcal{N}, C' \neq C$.*

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

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

Compared to the payoff vector u^*, in u^C every player in C is taxed equally, with a total summing up to ϵ; by contrast, players outside of C are paid equally, with a total 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 the required inequalities. By construction, $\sum_{i \in C} u^C_i = \sum_{i \in C} u^*_i - \epsilon < \sum_{i \in C} u^*_i$ for all $C \in \mathcal{S} \cup \mathcal{N}$, which verifies the first set of inequalities in the lemma.

Now consider two coalitions $C, C' \in \mathcal{S} \cup \mathcal{N}$ with $C \neq C'$. Coalition C can be partitioned as $C = (C \backslash C') \cup (C \cap C')$. Compared to u^*, $u^{C'}$ gives everyone outside C' an extra $\frac{\epsilon}{|N \backslash C'|}$, while lowering the payoff of everyone inside C' by $\frac{\epsilon}{|C'|}$, so

$$\sum_{i \in C} u^{C'}_i = \sum_{i \in C \backslash C'} u^{C'}_i + \sum_{i \in C \cap C'} u^{C'}_i = \Big[\sum_{i \in C \backslash C'} u^*_i + \frac{|C \backslash C'|}{|N \backslash C'|}\epsilon\Big] + \Big[\sum_{i \in C \cap C'} u^*_i - \frac{|C \cap C'|}{|C'|}\epsilon\Big].$$

Combining terms above yields $\sum_{i\in C} u_i^{C'} = \sum_{i\in C} u_i^* - \left[\frac{|C\cap C'|}{|C'|} - \frac{|C\setminus C'|}{|N\setminus C'|}\right]\epsilon$. Note that since $C \neq C'$, either $C\setminus C' \neq \emptyset$ or $C\cap C' \neq C'$ (or both) must be true; in other words, either $\frac{|C\setminus C'|}{|N\setminus C'|} > 0$ or $\frac{|C\cap C'|}{|C'|} < 1$. In either case, $\sum_{i\in C} u_i^{C'} > \sum_{i\in C} u_i^* - \epsilon = \sum_{i\in C} u_i^C$, which gives us the second set of inequalities in the lemma. ■

Proof of Theorem 5. The result comprises two parts.

Part 1: *For all $\delta \geq 0$, public PCEs give each player i at least $\underline{v}_i$ and each coalition $C \in \mathcal{S}$ a total payoff of at least $\underline{u}_C$.*

Part 2: *For every $u \in \mathcal{U}_{CR}(\mathcal{S})$, there $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$, there exists a public PCE supporting u.*

We prove Part 1 here and Part 2 in the Supplementary Appendix. In fact, we establish a statement stronger than Part 1: every public PCE σ guarantees that at every history $h \in \overline{\mathcal{H}}$, $U_i(h|\sigma) \geq \underline{v}_i$ for every $i \in N$ and $\sum_{i\in C} U_i(h|\sigma) \geq \underline{u}_C$ for every $C \in \mathcal{S}$. The fact that $U_i(h|\sigma) \geq \underline{v}_i$ for every $i \in N$ follows from similar arguments as in the proof of Theorem 1. Towards a contradiction, suppose σ is a public plan such that there exists a coalition $C \in \mathcal{S}$ and history $\widehat{h}$ where $\sum_{i\in C} U_i(\widehat{h}|\sigma) < \underline{u}_C$. We prove that σ cannot be a PCE.

To this end, we construct a profitable blocking plan from σ for coalition C. At every history $h \in \overline{\mathcal{H}}$, let $a(h|\sigma)$ denote the default and $\alpha(h) \in \arg\max_{a\in E_C(a(h|\sigma))} \sum_{i\in C} v_i(a)$ be an alternative in coalition C's "best response." By the definition of $\underline{u}_C$, it follows that $\sum_{i\in C} v_i(\alpha(h)) \geq \underline{u}_C > \sum_{i\in C} U_i(\widehat{h}|\sigma)$, so we can find transfers among players in C such that when combined with $\alpha(h)$, these transfers give each $i \in C$ higher payoff 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 $j \in N\setminus C$, 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

$$u_i\Big(\alpha(h),\ \widetilde{T}_C(h), \chi^C\big(T(h|\sigma)\big)\Big) \geq v_i(\alpha(h)) + \sum_{j\in C} \widetilde{T}_{ji}(h) - \sum_{j\in C} \widetilde{T}_{ij}(h) > U_i(\widehat{h}|\sigma),$$

where the weak inequality follows because $\chi^C_{ji}(\,T(h|\sigma)\,) \geq 0$ and $\widetilde{T}_{ij}(h) = 0$ for all $j \in N\setminus C$. Observe that the inequality above can hold 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: $U_i(\widehat{h}|\alpha, \beta) > U_i(\widehat{h}|\sigma)$ for every $i \in C$. Lemma 5 then implies that σ is not a PCE.