Subgame Perfect Coalition Formation

Greg Leo[*]   Yevgeniy Vorobeychik[**]   Myrna Wooders[*]

May 2, 2022

"They used to tell me you have to use your five best players, but I've found that you win with the five who fit together best." Red Auerbach

Abstract: We analyze a dynamic game where players can each make offers to other players to form coalitions. We show that these games have a unique subgame perfect equilibrium outcome that is individually rational and, when players can make enough proposals, Pareto optimal. We also provide sufficient conditions for equilibrium to implement core coalition structures.

1   Introduction

The importance of coalitions to economic and social activities is without doubt; kids split into teams on the playground, countries form alliances, students work in study groups. Coalitions can have significant impacts on the global economy, for example, OPEC, the Middle Eastern oil cartel. The importance of coalitions justifies the large and growing amount of literature studying coalition formation. In this paper, we focus on decentralized coalition formation modelled as a sequential bargaining game[1], called a *Sequential Proposer Game* (SPG).

An SPG proceeds as follows:

1. Players take turns making proposals.[2]

---

[*]Department of Economics, Vanderbilt University.

[**]McKelvey School of Engineering, Washington University in St. Louis.

[1]Sequential bargaining as in our model has its origins in Stahl (1972, 1977) and Rubinstein (1982).

[2]We assume every player has an opportunity to make at least one proposal.

2. The proposer invites a subset of players who are not already part of a coalition to join a coalition with them.

3. Players in the proposed coalition then either accept or reject the invitation. If all accept, the coalition is formed. If a player refuses, the duty of proposing moves to the next person in the exogenous order, which may be the same person. This process continues until all players are part of a coalition, or those players who are not in a coalition have no more opportunities to propose. At this point, those players not in a coalition are each assigned to a coalition consisting of themselves alone.

The extensive form of an SPG is complex. One complicating factor is that, without further assumptions, in a subgame perfect Nash equilibrium (SPNE) of an SPG a player may have their proposal rejected but the same proposal, made subsequently, may be accepted. This is true even though we assume that preferences over coalitions are strict. To simplify analysis, and following Bloch and Diamantoudi (2011), we assume that delay in forming a coalition has some small cost. We show that under this assumption, in every SPNE, all proposals are accepted– there is no delay, and the outcome is unique. With this in hand, we can talk about the outcome properties of an SPG.

As a result of allowing each player the option to reject any proposal they may receive, the equilibrium outcome of any SPG is individually rational. Further, all coalitions that are mutually most preferred by all their members will be formed. This also occurs recursively. Once a mutually preferred coalition has formed, any coalition of the remaining players that is mutually preferred by all its members is also formed, and so on. A procedure with this property is said to *iteratively match soulmates* (Leo et al., 2021).

We note that, in general, the equilibrium of an SPG does not necessarily select a core coalition structure (a partition of the total player set into coalitions). In fact, the core may well be empty. However, a Pareto efficient matching always exists, and it is natural to ask whether equilibrium outcomes are Pareto efficient. This question leads to our most striking result:

When each player is able to make sufficiently many proposals (at least one more than the number of possible coalitions containing the player), the SPNE outcome is always Pareto optimal.

We call games satisfying the proposal count condition *Sequential Repeating Proposer Games* (SRPGs).

In contrast to the well-known model of sequential coalition formation of serial dictatorship of Bogomolnaia and Moulin (2001), to analyze the equilibrium outcomes of a SRPG the game theorist needs to know the preferences of players. In serial dictatorship, players simply dictate, rather than propose, the set of players who will be in a coalition with them. Once a player dictates their chosen coalition, or is chosen to be in a coalition, they have no further choices in the game. Unlike equilibrium of a SPG game, serial dictatorship outcomes are always Pareto efficient but may fail to be individually rational. Serial dictatorship, however, has strong incentive properties as a centralized mechanism. In particular, the game theorist does not need to know the

preferences of individual players to obtain the Pareto efficient outcomes of serial dictatorship. Motivated by this observation, Lou et al. (2022) continue the work in this paper by considering the design of a centralized coalition process that implements the outcome of SPRGs from reported preferences through computational experiments.

To place our results in the literature, we emphasize that our approach is to study generalized coalition formation by starting with a strategic game meant to abstract a natural decentralized formation process.[3] Coalition formation has had a long history in economics going back to Edgeworth (1881), who discussed the concept of the contract curve, now known as the core. In contrast to our paper, many papers studying coalition formation games begin instead with a cooperative game and then design a strategic game whose equilibrium outcome is in the core of the cooperative game or satisfies some other notion of stability from cooperative game theory, such as the stability of the von Newman-Morgenstern stable set. See, for example, Harsanyi (1974); Binmore (1985); Gul (1989); Hart and Mas-Colell (1992); Okada (1996); Perry and Reny (1994); Seidmann and Winter (1998); Bogomolnaia and Jackson (2002); Bloch and Diamantoudi (2011); Herings et al. (2010); Kóczy (2015).

There is a much smaller literature on coalition formation that, like our paper, begins with a strategic game that may or may not have an empty core.[4] For example, Selten and Wooders (1991) introduces a non-cooperative game model that may have an empty core but, over time, as more players join the game, optimally-sized coalitions form. Konishi and Ray (2003) treat a dynamic process of coalition formation and study its absorbing states. Lehrer and Scarsini (2013) consider a game that is a mix of both cooperative and non-cooperative theory and show that, for a sufficiently large discount factor, several concepts of cores are non-empty. Arnold and Wooders (2015) treats a club-formation model, which does not necessarily have a non-empty core and implement concept new to club theory, an ergodic club equilibrium.

One of the key issues raised by our analysis is whether and how outcome properties depend on the proposal sequence. To what extent the number of proposals that can be made by each player affects the outcomes of bargaining games is largely unknown. One paper, however, that particularly focuses on this issue is Moldovanu and Winter (1995), which reanalyzes the game proposed by Selten (1988). In Selten's game, the SPNE depends on the assumed order of moves. Moldovanu and Winter show that an allocation is in the core (of the underlying nontransferable utility game) if and only if, for every possible order of moves, the allocation is an outcome of a stationary subgame perfect equilibrium.[5] Our work emphases the important role that allowing players to

---

[3]There has been an enormous amount of literature studying strategic coalition formation in applications, such as political coalitions (Acemoglu et al., 2005), bargaining games (see Ray and Vohra, 1999, for a survey), public good provision (Liu, 2018), and 2-sided matching games (Sönmez, 1997). This literature is important, but our focus on noncooperative games of coalition formation and the number of proposals required to achieve Pareto efficient outcomes is distinct.

[4]Note that a non-cooperative game can be modelled as a cooperative game by assigning a worth to each possible coalition of players. How this worth is assigned of course affects the outcome. For example, Lehrer and Scarsini (2013) assume that if a coalition deviates during the course of the game, after the deviation the coalition is on its own.

[5]Perry and Reny (1994) go further and assume that time is continuous and any proposer can make a proposal at any time (with some additional conditions). The discussion of the order of proposals in Perry-Reny is insightful and important, but their approach, with continuous time, is outside the scope of this paper.

make multiple offers can have.

We have organized this paper as follows. In section 2, we outline our theoretical environment and model of decentralized coalition formation. In section 3, we provide our theoretical results. In section 4, we conclude.

2 Modeling Decentralized Matching

2.1 Environment

We consider a well known model described in Banerjee et al. (2001) of an environment populated by a set of players $N = \{1, \ldots, n\}$ who are to be partitioned into coalitions. A *coalition* $T \in 2^n$ is a set of players, and a *coalition structure* π is a partition of the total player set into coalitions. For a player i, let π_i be the coalition in the partition π containing i.

In many situations, coalitions face feasibility constraints; for example, coalitions may be constrained to consist of at most k individuals. Generically, let $\mathcal{T}$ denote the set of *feasible* coalitions, which we assume to always include singleton coalitions, $\{i\}$. For a player i, we denote the subset of feasible coalitions that all include i by $\mathcal{T}_i \subset \mathcal{T}$. Each player $i \in N$ has a strict preference ordering $\succ_i$ over $\mathcal{T}_i$. A profile of preferences $\succ$ (or *profile* for short) is a list of preferences for every $i \in N$. Given a profile $\succ = (\succ_1, ..., \succ_i, ..., \succ_n)$, the list of preferences for all players except i is denoted by $\succ_{-i}$. In addition, we assume that players have lexicographic preferences over time, that is, for all $t < t'$, joining a coalition T at time t is strictly preferred to joining T at time t'.[6]

2.2 Sequential Proposer Games

We model the decentralized process of hedonic coalition formation using a natural non-cooperative game with perfect information.[7] In the game, players sequentially propose coalitions that are then accepted or rejected by their prospective members. We term such games *sequential proposer games (SPGs)*.

Formally, an SPG is a game of perfect information in extensive form with player set N, a set of feasible coalitions $\mathcal{T}$, a preference profile $\succ$ *over coalitions*, and an ordered list of players $O = (i_1, i_2, ..., i_m)$, in which each player $i \in N$ is included at least once.[8] The ordering determines the order in which players can make proposals to other players (or to themselves alone) to form coalitions.

The game begins with the first player in the ordering, say i, proposing a coalition $T \in \mathcal{T}_i$. The players in T then sequentially decide whether to accept the proposal. If all players in T accept the proposal, then those players have no decision nodes in the remaining subgame and, in particular, they can no longer make proposals (they loose

---

[6]Time can be measured by the number of actions that are taken to go from a node to a final outcome. Our use of lexicographic preferences was inspired by Bloch and Diamantoudi (2011). Alternatively, we could assume discounting, but any positive level of discounting would lead to the same results.

[7]A hedonic game is simply a game with ordinal preferences over coalitions of membership.

[8]For now, there are no further restrictions on O; for example, if $N = \{1, 2, 3\}$ the ordering O may be $(3, 1, 2)$ or $(1, 1, 3, 2, 3)$.

their places in the ordering).[9] In either case, after this accept/reject phase, we arrive at a new subgame where it is the next player's turn in the ordering to make a proposal (provided that she has not already joined a coalition).

The process continues until there are no more opportunities for coalitions to form – either (a) all players are in coalitions or (b) the last proposer, m, in the order has made their last proposal and players to whom the proposal is made have responded. In case (b), the remaining players, if any, become singleton coalitions. In either case, the outcome is a coalition structure.

We illustrate the mechanics of this game through a simple example.

Example 1. *Consider an SPG with four players $N = \{1, 2, 3, 4\}$, and the order of proposers $O = (1, 2, 3, 4)$ in which the size of each coalition is at most two (roommate problem). Suppose that the profile of preferences is as follows:*

$$
\begin{array}{ccccccc}
1: & \{1,4\} & \succ_1 & \{1,2\} & \succ_1 & \{1,3\} & \succ_1 & \{1\} \\
2: & \{2,1\} & \succ_2 & \{2,4\} & \succ_2 & \{2,3\} & \succ_2 & \{2\} \\
3: & \{3,2\} & \succ_3 & \{3,1\} & \succ_3 & \{3,4\} & \succ_3 & \{3\} \\
4: & \{4,3\} & \succ_4 & \{4,2\} & \succ_4 & \{4,1\} & \succ_4 & \{4\}
\end{array}
$$

The following is an example scenario:

1. *Player 1 proposes to $\{1,4\}$, and 4 rejects the proposal.*

2. *2 proposes to $\{2,1\}$ and 1 accepts the proposal. The group is formed and both 1 and 2 are removed from the game.*

3. *3 proposes to $\{3,4\}$ and 4 accepts the proposal. The group is formed and $3, 4$ are removed.*

The partition that results from this sequence is $\pi = \{\{1,2\}, \{3,4\}\}$.

3 Results

This section is split into three parts. First, we provide our results about the SPNE of sequential proposer games: that any SPNE of an SPG involves no delay and that the SPNE outcome is unique. In this subsection, we also provide several examples demonstrating subtleties of analyzing these games. Next, we prove that the unique outcome of every SPG is individually rational and iteratively matches soulmates. Last, we show that the outcome of subset of SPE that give each player a sufficient number of proposals are also Pareto efficient.

3.1 Equilibrium Properties

As we demonstrate below, there are several important properties that hold in *any* subgame perfect Nash equilibrium of an arbitrary accept-reject game:

---

[9]Informally, we can think of those players who all agree to be in some proposed coalition as leaving the game; their assigned coalition is determined, and they have no further actions in the game.

- *individual rationality* (If a player is matched into a coalition with at least one other player, they prefer that coalition to remaining alone.)

- *iterative matching of soulmates* (Players who most-prefer to be together are matched, even in a more general sense discussed below.)

- *core* (When all players can be matched as soulmates, the outcome is the unique core outcome of the derived cooperative game.)

Pareto optimality is not *necessarily* satisfied by an SPNE outcome, as we show presently.

An outcome π is *Pareto optimal* if there does not exist another feasible outcome π' that is strictly preferred by a nonempty subset of players $N' \subset N$ and to which all other players, $N \setminus N'$ are indifferent. In our context, this means that an outcome is Pareto optimal if there is no collection of players who can all be made better off by a reshuffling of coalition memberships among these players while maintaining the same coalition memberships of all remaining players, if any.

The following example illustrates a SPNE with an outcome that is not Pareto optimal. For simplicity, we consider pairwise matching in the examples but emphasize that our model does not restrict the size of coalitions. What is particularly revealing is that small and seemingly inconsequential changes solely to the order of proposals can have significant changes in equilibrium outcomes.

Example 2. *Consider a roommate problem[10] with a set of 6 players, $\{1, \ldots, 6\}$ who have the following preferences:*

$$
\begin{array}{llllllllllll}
1: & \{1,3\} & \succ_1 & \{1,4\} & \succ_1 & \{1,5\} & \succ_1 & \{1,2\} & \succ_1 & \{1,6\} & \succ_1 & \{1\} \\
2: & \{2,1\} & \succ_2 & \{2,5\} & \succ_2 & \{2,4\} & \succ_2 & \{2,3\} & \succ_2 & \{2,6\} & \succ_2 & \{2\} \\
3: & \{3,2\} & \succ_3 & \{3,4\} & \succ_3 & \{3,1\} & \succ_3 & \{3,5\} & \succ_3 & \{3,6\} & \succ_3 & \{3\} \\
4: & \{4,3\} & \succ_4 & \{4,1\} & \succ_4 & \{4,5\} & \succ_4 & \{4,2\} & \succ_4 & \{4,6\} & \succ_4 & \{4\} \\
5: & \{5,2\} & \succ_5 & \{5,4\} & \succ_5 & \{5,1\} & \succ_5 & \{5,3\} & \succ_5 & \{5,6\} & \succ_5 & \{5\} \\
6: & \{6,5\} & \succ_6 & \{6,1\} & \succ_6 & \{6,2\} & \succ_6 & \{6,3\} & \succ_6 & \{6,4\} & \succ_6 & \{6\}
\end{array}
$$

Suppose the order of proposers is $O = (1, 2, 3, 4, 5, 6)$. The SPNE outcome of this game is $\{\{1,5\}, \{2,4\}, \{3,6\}\}$, as argued in appendix A. This outcome is not Pareto optimal, as $\{\{2,5\}, \{1,4\}, \{3,6\}\}$ is a Pareto improvement. $\square$

Our next example illustrates that with a change in the ordering of players, Pareto-optimality may be achieved. In this example, we make only a minor modification of Example 2: letting player 1 move twice in the very beginning rather than just once.

Example 3. *Specifically, the new order is $O = (1, 1, 2, 3, 4, 5, 6)$; everything else (in particular, the set of players, their preferences, and feasible coalitions) is the same as example 2. We now show that this modification results in coalitions in which players are completely reshuffled. First, it is immediate that if any proposal by 1 is rejected in the very beginning, the subgame becomes identical to the game in Example 2. Having*

---

[10]The stable roommates problem models an environment where a set of players need to be matched into groups of no more than two and was originally introduced by Gale and Shapley (1962)

this in mind, suppose that 1 makes an offer to $\{1,3\}$ at the initial node. If 3 rejects, then 3 is matched with 6 in the resulting subgame. Clearly, 3 strictly prefers to be in a coalition with 1, and would therefore accept. Once the coalition $\{1,3\}$ is formed, 2 and 5 prefer to be with one another rather than with anyone else, and the resulting coalition must be formed as well (see our discussion of this below, in the context of iteratively matching soulmates). Consequently, the SPNE outcome is $\{\{1,3\},\{2,5\},\{4,6\}\}$. It is easy to verify that this outcome is Pareto optimal.

We next proceed to prove several interesting and useful characteristics of subgame prefect Nash equilibria of SPGs, as well as some properties of their equilibrium outcomes. We start with some additional notation. Define a *T-subgame* as the subgame of an SPG in which an offer T has been made and the players $i \in T\backslash\{i\}$ sequentially decide whether to accept or reject this offer. For any proposal T, denote the subgame in which T is rejected by A_{TR} and the subgame in which T is accepted by A_{TA} . Note that each such subgame of an SPG is itself an SPG, with the caveat that we lift the restriction that each player appears at least once in the order O.

Recall that, as in any subgame of a game, if a player does not own any decision nodes in that subgame, then the player has no more choices to make; this holds for all those players who, at prior decision nodes, joined coalitions. A subgame allows the possibility, however, that one or more players may no longer be able to make proposals but still may own decision nodes requiring them to accept or reject proposals.

First, we make a simple observation that is worth stating as a formal remark for the purposes of our proofs below.

Remark 1. *For any strict subgame A, and for any two feasible proposals T, T', $A_{TR} = A_{T'R}$.*

This follows immediately from the fact that if a proposal from a player i is rejected, the outcome is independent of the specific proposal T that was made.

The next lemma serves largely as a tool in subsequent results, but may be interesting in its own right as it addresses the issue of coordination faced by players who had just received a proposal to be on some coalition T and who all prefer T to the outcome that would materialize if this coalition were rejected. We show that in an SPNE such a coalition T will always be accepted, but observe that this is entirely a consequence of the sequential nature of the accept/reject decisions and the assumption of lexicographic preferences. In particular, if players were to decide coalition membership simultaneously, the game becomes one of coordination and a host of "bad" equilibria could emerge in which, for example, a collection of players jointly reject the coalition that is better for all players in the collection. In contrast, with sequential decision-making, the coalition T is selected. Lexicographic preferences ensure that, in this situation, players do not reject a proposal to join T even if T would still be formed in a subsequent subgame.[11]

_____

[11] A player i may make a proposal to all members of T but, if the player makes a proposal that is rejected, they could receive a proposal from another member of the coalition T who appears later in the ordering. Note also that it is possible for a player i to make an offer of coalition $\{i\}$ and then reject the proposal, thus remaining available to join another coalition later in the game. In any case, as we will see, this will not happen in an SPNE.

Lemma 1. *For an arbitrary subgame A, consider a proposal of a coalition T. Let A_R be the subgame in which T is rejected, and let Π_R be the set of SPNE outcomes of A_R. Suppose that $\forall \pi_R \in \Pi_R$ and $\forall i \in T$ either $T \succ_i \pi_{R,i}$, or $T = \pi_{R,i}$.[12] Then all $i \in T$ will accept T in every SPNE of the T-subgame.*

Proof in appendix A.

Next we present one of the main results of this section: all subgame perfect Nash equilibria involve no delay, and result in a unique outcome.

Theorem 1. *In any SPNE of an arbitrary subgame A, all proposals are accepted along the equilibrium path. Moreover, the SPNE outcome is unique.*

The proof of Theorem 1 is obtained by backward induction and is provided in appendix A.

While the above result shows that an SPNE has the property that at every proposer node along the equilibrium path, the SPNE offer is accepted, a strategy must still specify what happens at every other node of the tree, including nodes that would follow a non-SPNE proposal or rejection of am SPNE proposal.

3.2   Outcome Properties

Having characterized the structure of subgame perfect Nash equilibria of an SPG, we now consider whether the unique outcome satisfies important properties. Individual rationality is immediate, since the strategy of rejecting every proposal will, in our game model, leave each player by themselves, and they can therefore do no worse in any subgame perfect Nash equilibrium.

Proposition 1. *Every SPNE outcome of any SPG is individually rational: each player i is at least as well off as in the singleton coalition $\{i\}$.*

Our second result of this section regards whether the outcome is a *core* outcome. The set of players N and their preferences $\succ$ determines a (hedonic) cooperative game of coalition formation. An assignment π of players to coalitions is in the *core* of this cooperative coalition formation game if there does not exist a coalition of players $T \subset N$ with the property that for all $i \in T$, $T \succ_i \pi_i$. As the following example demonstrates, the outcome of and SPNE in and SPG will not always be partition in the core, even when the core is non-empty.

Example 4. *Consider a bipartite matching problem with a set of 6 players, $\{1, \dots, 6\}$ who have the following preferences:*

$$
\begin{array}{llllllllllll}
1: & \{1,4\} & \succ_1 & \{1,5\} & \succ_1 & \{1,6\} & \succ_1 & \{1\} & \succ_1 & \{1,2\} & \succ_1 & \{1,3\} \\
2: & \{2,5\} & \succ_2 & \{2,4\} & \succ_2 & \{2,6\} & \succ_2 & \{2\} & \succ_2 & \{2,1\} & \succ_2 & \{2,3\} \\
3: & \{3,6\} & \succ_3 & \{3,5\} & \succ_3 & \{3,4\} & \succ_3 & \{3\} & \succ_3 & \{3,1\} & \succ_3 & \{3,2\} \\
4: & \{4,3\} & \succ_4 & \{4,2\} & \succ_4 & \{4,1\} & \succ_4 & \{4\} & \succ_4 & \{4,5\} & \succ_4 & \{4,6\} \\
5: & \{5,3\} & \succ_5 & \{5,1\} & \succ_5 & \{5,2\} & \succ_5 & \{5\} & \succ_5 & \{5,4\} & \succ_5 & \{5,6\} \\
6: & \{6,2\} & \succ_6 & \{6,3\} & \succ_6 & \{6,1\} & \succ_6 & \{6\} & \succ_6 & \{6,4\} & \succ_6 & \{6,5\}
\end{array}
$$

---

[12]Where $\pi_{R,i}$ is the coalition to which i is assigned in π_R.

Suppose that all of the above coalitions are admissible, and that the order of proposers is $O = (1, 2, 3, 4, 5, 6)$. The unique SPNE outcome of the SPG is $\{\{1, 5\}, \{2, 6\}, \{3, 4\}\}$. However, $\{3, 5\}$ is a blocking pair, and this game has two core outcomes: $\{\{1, 5\}, \{2, 4\}, \{3, 6\}\}$ and $\{\{1, 4\}, \{2, 5\}, \{3, 4\}\}$.

However, despite not always leading to a core outcome, we now show that SPG equilibria implement another important property, *iterated matching of soulmates (IMS)* (Leo et al., 2021). This property helps provide a sufficient condition to guarantee that SPG outcomes are in the core.

IMS captures the idea of forming coalitions (from the set of players not already in coalitions) which are mutually most-preferred by all of their members. Formally, a coalition T is a coalition of (1st-order) *soulmates* if for all $i \in T$, $T \succ T'$ for all $T' \in \mathcal{T}_i$. IMS is the iterative application of this criterion. In every iteration, match all coalitions consisting of soulmates among players not matched in prior rounds.

Informally, this criterion may be of independent importance because any mechanism, centralized or decentralized, which does not match players who wish to be with one another might be ill received. A more formal motivation is that all coalitions matched by IMS are blocking coalitions (Leo et al., 2021), and players in blocking coalitions may create instability.[13] Moreover, implementing IMS has important consequences for incentive compatibility and core stability. Next, we show that SPG subgame perfect Nash equilibrium outcomes always match soulmates in this iterative sense. More precisely, let $\hat{\mathcal{T}}_{IMS}$ be a collection of coalitions produced by IMS. We say that SPG *implements IMS in SPNE* partition π if $\hat{\mathcal{T}}_{IMS} \subseteq \pi$.

Proposition 2. *Every SPNE of an SPG implements IMS.*

Proof. We prove this by induction.

Base Case: We show that every soulmate coalition must be formed by any SPNE.

Consider a SPNE s in which all proposals are accepted (this is sufficient, since such an SPNE always exists and all SPNE result in a unique outcome by Theorem 1), and let π be the corresponding SPNE outcome. Let T be a coalition of soulmates, and suppose that it is not formed by s. Let $i \in T$ be the earliest proposer in T and let π_i be the coalition to which i is assigned by s. Suppose i proposes to T. By Lemma 1 and the fact that T is a coalition of soulmates, all members of T will accept this proposal. Because $T \succ_i \pi_i$, i strictly prefers to propose T than to propose π_i, s cannot be a SPNE.

Inductive Step: Suppose that all coalitions of kth order soulmates (i.e., from the first k rounds of IMS) are formed. We now show that all soulmate coalitions from $k + 1$st round form as well. We do this by a similar contradiction argument as the base case.

---

[13]As shown by Leo et al. (2021), the assumption that all players can be matched as soulmates is weaker than the top coalition property of Banerjee et al. (2001).

Again, let s be an SPNE with outcome π, and let T be the coalition of $k + 1$st round (conditional) soulmates (i.e., soulmates if all soulmate coalitions from previous k rounds form), and suppose T is not formed. Let $i \in T$ be the earliest proposer in T and let π_i be the coalition to which i is assigned by s. Suppose i proposes to T. By Lemma 1, the fact that T is a coalition of conditional $k + 1$st round soulmates, and the inductive hypothesis, all will accept this proposal (since they cannot possibly be on a coalition with anyone from the first k IMS rounds, and strictly prefer T to all other coalitions). Since $T \succ_i \pi_i$, i strictly prefers to propose T than to propose π_i (which cannot contain any coalitions including soulmates from the first k rounds of IMS), s cannot be a SPNE. $\qquad\square$

As shown by Leo et al. (2021), if IMS matches all players, the resulting outcome is the unique core coalition structure. The following corollary then follows.

Corollary 1. *Suppose that all players can be matched by IMS. Then every SPNE of an arbitrary SPG yields the unique core coalition structure.*

3.3 Pareto Optimality of Sequential Repeating Proposer Games

As we showed in Example 2, SPNE outcomes of an arbitrary SPG need not even be Pareto optimal. Recall, however, that the SPNE outcome of Example 3 *is* Pareto optimal. Thus, as we had observed, ordering over the players can potentially restore Pareto optimality. We now use this insight to devise a restriction of SPGs—specifically, restricting the orderings over proposers—which allows us to guarantee that the outcome is always Pareto optimal.

Specifically, we propose a class of SPGs which we term *Sequential Repeating Proposer Games (SRPGs)*. In an SRPG, the order O over players is such that each player i can make $|\mathcal{T}_i| + 1$ proposals before we move on to another player. It turns out that this condition suffices to guarantee Pareto optimality.[14]

Example 5. *Consider again Example 2, but now let the order allow each proposer to propose seven times, that is, $O = (1, 1, 1, 1, 1, 1, 1, 2, 2, \ldots, 5, 5, 6, 6, 6, 6, 6, 6, 6)$.*

If the very first proposal by 1 is rejected, it is not difficult to show, through a slightly modified argument as in Example 2, that the SPNE outcome is $\{\{1, 5\}, \{2, 4\}, \{3, 6\}\}$. Consequently, as in Example 3, if 1 proposes to $\{1, 3\}$, 3 will accept, and the resulting SPNE outcome of the SRPG is the Pareto optimal outcome $\{\{1, 3\}, \{2, 5\}, \{4, 6\}\}$.

Again, just as the Example 3, the last proposal by 1 serves as a credible threat of the inefficient outcome if the proposal is rejected, which creates the incentive for 3 to accept an offer it would otherwise have rejected.

Theorem 2. *Every SPNE of a SRPG is Pareto optimal.*

Before we prove Theorem 2, we make several formal remarks useful in the proof.

---

[14]Recall that in any SPG, all proposals are accepted. Thus, the size of the set $|\mathcal{T}_i|$ is immaterial here, since in equilibrium, players would only ever make a single proposal. It is only the *potential* of making these proposals that matters.

Remark 2. *Consider a proposer i and consider $k < |\mathcal{T}_i| + 1$ so that i is proposing for the kth time (having been rejected $k-1$ times). Let $\pi_{i,k}$ be the coalition i is assigned to in the SPNE of the game that starts with this kth proposal. Then either $\pi_{i,k} \succ_i \pi_{i,k+1}$ or $\pi_{i,k} = \pi_{i,k+1}$.*

This follows from observing that if $\pi_{i,k+1} \succ_i \pi_{i,k}$, then in SPNE, when i proposes for the kth time, i should make a proposal that will be rejected, contradicting Theorem 1.

Corollary 2. *It follows that there must be some $\bar{k}$ such that $\pi_{i,\bar{k}} = \pi_{i,\bar{k}+1}$, since i can propose more times than there are possible coalitions for i to propose to.*

Remark 3. *If π_k is the SPNE outcome of the game beginning with player i's kth proposal and $\pi_{i,k} = \pi_{i,k+1}$, then $\pi_k = \pi_{k+1}$.*

This follows because the subgame that follows i proposing to $\pi_{i,k}$ and being accepted is the same whether it occurs following i's kth or $k+1$th proposal. Specifically, the next proposer j is the same (the next player in the ordering O who is not in $\pi_{i,k}$) and the set of available players for j to propose to is the same.

Lemma 2. *Let π_1 be the SPNE outcome of a subgame A_1 with player i proposing for the first time, and let π_2 be the SPNE outcome of A_2, the subgame which results if i's first proposal is rejected. Then $\pi_1 = \pi_2$.*

Proof. We will show that if $\pi_{i,k} = \pi_{i,k+1}$, then $\pi_{i,k-i} = \pi_{i,k}$. The result then follows from Remark 3.

Assume $\pi_{i,k} = \pi_{i,k+1}$. From Remark 2, $\pi_{i,k-1} \succ_i \pi_{i,k}$ or $\pi_{i,k-1} = \pi_{i,k}$. If $\pi_{i,k-1} = \pi_{i,k}$, then by Remark 3 $\pi_{k-1} = \pi_k$ and the result follows as shown below.

Assume instead, for contradiction, that $\pi_{i,k-1} \succ_i \pi_{i,k}$. Then, since the coalition $\pi_{i,k-1}$ is accepted by all its members, we have that $\forall\, j \in \pi_{i,k-1}, \pi_{i,k-1} \succ_j \pi_{j,k}$.

Now since $\pi_{i,k-1} \succ_i \pi_{i,k}$, we must have that if i proposes the coalition $\pi_{i,k-1}$ on her kth proposal, it is rejected. Otherwise, if it were accepted in SPNE, i would propose $\pi_{i,k-1}$ and the coalition would be formed, with player i better off as a result, contradicting the fact that it's an SPNE outcome of the game starting with i's kth proposal. This implies that for some $j \in \pi_{i,k-1}, \pi_{j,k+1} \succ_j \pi_{i,k-1}$. But since $\pi_k = \pi_{k+1}$, we have that $\pi_{j,k} \succ_j \pi_{i,k-1} \succ_j \pi_{j,k}$, a contradiction. Thus, it is not the case that $\pi_{i,k-1} \succ_i \pi_{i,k}$, so it must be that $\pi_{i,k-1} = \pi_{i,k}$, implying that $\pi_{k-1} = \pi_k$.

By recursively applying what we have shown thus far, that $\pi_{i,k} = \pi_{i,k+1}$ implies $\pi_{i,k-i} = \pi_{i,k}$, beginning with the $\bar{k}$ from our Corollary to Remark 2, we have that $\pi_1 = \pi_2$, as desired. $\qquad\square$

We now proceed to prove Theorem 2 by contradiction. Let π be the SPNE outcome of an SRPG. Suppose, for contradiction, that the set of coalitions π' is a Pareto-improvement over π. That is, each player j has $\pi'_j \succeq_j \pi_j$ with at least one player having $\pi'_j \succ_j \pi_j$. We will show that this implies π is not an SPNE outcome.

Since $\pi \neq \pi'$, there are some players on different coalitions in π and π'. Let Q be the set of such players, and let $i \in Q$ be the first such player to propose.

It must be that all players in π_i and π'_i are still available when i proposes for the first time. If this were not the case, then there must have been some other player j who

proposed before i who is in π_i' but not π_i (note that all members of π_i are in Q, so i is the first member of π_i to propose, which implies by Theorem 1 that all members of π_i are available when i first proposes). But then $j \in Q$, contradicting the premise that i is the first player in Q to propose.

Now, let A be the subgame starting with i's first opportunity to propose. Since π' is a Pareto-improvement over π, from strict preferences over coalitions, it follows that $\pi_i' \succ_j \pi_j$ for all $j \in \pi_i'$. By Lemma 2, if these players j reject a proposal of π_i', they will be assigned, in SPNE, to π_j in the subgame A_R that begins if i's first proposal is rejected. Thus, if i's first proposal is to the coalition π_i', all members of the coalition will accept. Therefore, since $\pi_i' \succ_i \pi_i$, i will propose to π_i' and this proposal will be accepted. Thus π is not, in fact, an SPNE outcome of the SRPG, since the SPNE outcome is unique by Theorem 1. $\quad\square$

While SRPGs do resolve the Pareto optimality issue, if Example 4 is extended to use the SRPG order, it still results in the same outcome, which is not in the core. Nevertheless, it has been argued that Pareto optimality may itself be a compelling stability property in many coalitional settings (Morrill, 2010).

4   Conclusions

We first summarize our work. We consider sequential non-cooperative coalition formation games with a finite horizon. In these games, players propose coalitions, which are then sequentially accepted or rejected. We analyze subgame perfect Nash equilibria of the resulting perfect information game. Our first key result is that there is a unique SPNE equilibrium outcome, and that all SPNE involve no delay on the equilibrium path (all proposals are accepted).

Our second major positive result is that, in a SPNE, coalitions that are mutually most-preferred, even in a stronger iterative sense, are always formed. While this result is of independent interest, we also use it to provide a sufficient condition for the core outcome to be implemented in an equilibrium of our game.

Our most significant result demonstrates that the number of proposals a player can make may have important effects on the equilibrium outcome of the game. Theorem 2 shows that with a sufficient number of sequential proposals for each player, the equilibrium outcome is Pareto optimal. As Example 2 shows, however, Pareto optimality is not guaranteed if a player can only propose once. Theorem 2 is both novel and non-obvious, with the proof requiring all the results obtained before this theorem.

Given the complexity of our game theoretic model, the driving forces behind why allowing each player to make several proposals ensures Pareto optimality are subtle. One factor, highlighted by our examples, is that, should a proposal of a player be rejected, the outcome of the subgame that follows can act as a credible threat. For instance, at the start of Example 3, player 1 has two proposals. Should the first proposal be rejected, player 1 has only one proposal left, and the game becomes identical to Example 2. The outcome of that game then becomes a credible threat against which player 1 can propose in the first round. If a SPNE outcome were not Pareto efficient, then at some point in the game, players accepted a coalition that should have served as

a credible threat against rejecting whatever coalition they would be matched with in the Pareto improvement. The proof of the final step in Theorem 2 formalizes this intuition.

Our analysis raises several questions for future study. First, as shown by comparing Example 3 to 5, it is not always necessary to give every player a large number of proposals. In fact, giving player 1 two proposals in Example 3 gives the exact same outcome as the full SRPG in Example 5. This raises the question whether the number of proposals required in Theorem 2 can be relaxed. Furthermore, in our SRPGs, each players' proposals come in a block. A player continues to propose until their proposal has been accepted or all of their proposals have been rejected. Our proofs rely on this structure, but it raises the question whether adding proposals would have the same effect if the ordering of proposals did not have this block structure. Finally, we wonder whether there are any welfare comparisons that can be made for individual players as the ordering of proposals are changed within an SRPG. In Example 5, the first proposer is matched with their favorite coalition. Is it always better to propose first?

While Theorem 2 is an interesting result for the class of coalition formation games considered, it also inspires a number of questions. Most important, can similar results be obtained for other classes of games? Are there other situations in which the ability of players to make multiple proposals can lead to Pareto improving outcomes? This may be of particular interest in political environments.

The strong properties of the SRPG equilibrium outcome also raises the question of the extent to which the SRPG can be adapted in practice to create a centralized mechanism that operates with no fore-knowledge of preferences. Our results, however, are obtained under complete information, and a centralized mechanism can only operate under reported preferences. The incentive compatibility of the resulting mechanism then becomes a major concern, but the strategic complexity of these games pose a major theoretical hurdle to analyzing the resulting incentives.

Acknowledgments

References

Daron Acemoglu, Simon Johnson, and James Robinson. The rise of Europe: Atlantic trade, institutional change, and economic growth. *American economic review*, pages

546–579, 2005.

Tone Arnold and Myrna Wooders. Dynamic club formation with coordination. *Journal of Dynamics & Games*, 2(3&4):341, 2015.

Suryapratim Banerjee, Hideo Konishi, and Tayfun Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18(1):135–153, January 2001. ISSN 0176-1714, 1432-217X. doi: 10.1007/s003550000067.

Ken Binmore. Equilibria in extensive games. *The Economic Journal*, 95:51–59, 1985.

Francis Bloch and Effrosyni Diamantoudi. Noncooperative formation of coalitions in hedonic games. *International Journal of Game Theory*, 40(2):263–280, 2011.

Anna Bogomolnaia and Matthew O. Jackson. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior*, 38(2):201–230, February 2002. ISSN 0899-8256. doi: 10.1006/game.2001.0877.

Anna Bogomolnaia and Hervé Moulin. A New Solution to the Random Assignment Problem. *Journal of Economic Theory*, 100(2):295–328, October 2001. doi: 10.1006/jeth.2000.2710.

Francis Ysidro Edgeworth. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. Number 10. C. Kegan Paul & Company, 1881.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Faruk Gul. Bargaining foundations of shapley value. *Econometrica: Journal of the Econometric Society*, pages 81–95, 1989.

John C Harsanyi. An equilibrium-point interpretation of stable sets and a proposed alternative definition. *Management science*, 20(11):1472–1495, 1974.

Sergiu Hart and Andreu Mas-Colell. A non-cooperative interpretation of value and potential. In *Rational interaction*, pages 83–93. Springer, 1992.

P. Herings, Ana Mauleon, and Vincent Vannetelbosch. Coalition formation among farsighted agents. *Games*, 1(3):286–298, 2010.

LászlóA Kóczy. Stationary consistent equilibrium coalition structures constitute the recursive core. *Journal of Mathematical Economics*, 61:104–110, 2015.

Hideo Konishi and Debraj Ray. Coalition formation as a dynamic process. *Journal of Economic theory*, 110(1):1–41, 2003.

Ehud Lehrer and Marco Scarsini. On the core of dynamic cooperative games. *Dynamic Games and Applications*, 3(3):359–373, 2013.

Greg Leo, Jian Lou, Martin Van der Linden, Yevgeniy Vorobeychik, and Myrna Wooders. Matching soulmates. *Journal of Public Economic Theory*, n/a(n/a), 2021. doi: https://doi.org/10.1111/jpet.12542. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jpet.12542.

Weifeng Liu. Global public goods and coalition formation under matching mechanisms. *Journal of Public Economic Theory*, 20(3):325–355, 2018.

Jian Lou, Chen Hajaj, and Yevgeniy Vorobeychik. A rotating proposer mechanism for team formation, 2022. URL https://arxiv.org/abs/2204.04251.

Benny Moldovanu and Eyal Winter. Order independent equilibria. *Games and Economic Behavior*, 9(1):21–34, 1995.

Thayer Morrill. The roommates problem revisited. *Journal of Economic Theory*, 145: 1739–1756, 2010.

Akira Okada. A noncooperative coalitional bargaining game with random proposers. *Games and Economic Behavior*, 16:97–108, 1996.

Motty Perry and Philip J Reny. A noncooperative view of coalition formation and the core. *Econometrica: Journal of the Econometric Society*, pages 795–817, 1994.

Debraj Ray and Rajiv Vohra. A theory of endogenous coalition structures. *Games and Economic Behavior*, 26(2):286–336, 1999.

Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50:97–109, 1982.

Daniel J Seidmann and Eyal Winter. A theory of gradual coalition formation. *The Review of Economic Studies*, 65(4):793–815, 1998.

Reinhard Selten. A noncooperative model of characteristic-function bargaining. In *Models of strategic rationality*, pages 247–267. Springer, 1988.

Reinhard Selten and Myrna H. Wooders. A Game Equilibrium Model of Thin Markets. In Professor Dr Reinhard Selten, editor, *Game Equilibrium Models III*, pages 242–282. Springer Berlin Heidelberg, 1991. ISBN 978-3-642-08110-1 978-3-662-07367-4.

Tayfun Sönmez. Manipulation via capacities in two-sided matching markets. *Journal of Economic theory*, 77(1):197–204, 1997.

Ingolf Stahl. Bargaining theory (stockholm school of economics). *Stockholm, Sweden*, 1972.

Ingolf Stahl. An n-person bargaining game in the extensive form. In *Mathematical economics and game theory*, pages 156–172. Springer, 1977.

A   Additional Proofs

Details of Example 2   Suppose that all of the above coalitions are feasible, and that the order of proposers is $O = (1, 2, 3, 4, 5, 6)$. For example, if 1's offer is rejected, 2 makes an offer. If that gets rejected, then 3 makes an offer, and so on. We now derive the subgame perfect Nash equilibrium outcome of this game (which turns out to be unique, as we show later).

1. Consider any subgame in which player 6 makes an offer. Clearly, every offer will be accepted, since rejection implies that the player who rejects an offer becomes a singleton (and each player in our example prefers to be on a coalition with anyone to being by themselves).

2. Consider a subgame in which players 1-4 have all been rejected, and it is player 5's turn to make an offer. If any offer by 5 is rejected at this point, the outcome will be $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5, 6\}\}$, since 5 is 6's most preferred coalitionmate, and by the preceding logic. Consequently, any offer by 5 to coalitions with players 1-4 will be accepted. Since 5 most prefers 2, who is still available, this is the offer 5 will make, and it will be accepted. Moreover, since 1 is the most preferred remaining player by 6, the outcome in this subgame is $\{\{1, 6\}, \{2, 5\}, \{3\}, \{4\}\}$. **SPNE outcome in this subgame:** $\{\{1, 6\}, \{2, 5\}, \{3\}, \{4\}\}$.

3. Consider a subgame in which players 1-3 have all been rejected, and it's player 4's turn to make an offer. If player 4 makes an offer to $\{3, 4\}$, the coalition $\{3, 4\}$ will form if 3 accepts or coalitions $\{3\}, \{4\}$ will form if it rejects (from subgame (2) above). Since 4 prefers 3 to any others, he can do no better than making an offer to $\{3, 4\}$, with the outcome being $\{\{1, 6\}, \{2, 5\}, \{3, 4\}\}$. It is thus an equilibrium of this subgame for 4 to offer to $\{3, 4\}$, and for 3 to accept. **SPNE outcome in this subgame:** $\{\{1, 6\}, \{2, 5\}, \{3, 4\}\}$.

4. Consider a subgame in which players 1 and 2 have been rejected, and now it's player 3's turn. If player 3 makes an offer to $\{2, 3\}$, 2 prefers to reject, because 2 prefers to be with 5 (the outcome of subgame (3)) than with 3. If player 3 makes an offer to $\{3, 4\}$, this offer is accepted, and the outcome is again $\{\{1, 6\}, \{2, 5\}, \{3, 4\}\}$. Making any other offer cannot improve 3's utility. **SPNE outcome in this subgame:** $\{\{1, 6\}, \{2, 5\}, \{3, 4\}\}$.

5. Consider a subgame in which player 1 was rejected, and player 2 now makes an offer. If 2 makes an offer to $\{1, 2\}$, 1 will accept, because if 1 rejects, they end up paired with 6 (subgame (4)), and 1 prefers being with 2. Since 1 is the most prefered pick by 2, 2 would strictly prefer making this offer to any other. Thus, coalition $\{1, 2\}$ will form. Once this happens, $\{3, 4\}$ will coalition up since they are then conditional soulmates, which implies that $\{5, 6\}$ will coalition up as well. **SPNE outcome in this subgame:** $\{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$.

6. Now, consider player 1's options. If 1 makes an offer to $\{1, 3\}$ or $\{1, 4\}$, it will be rejected, because both 3 and 4 prefer to be with each other than to be with 1 (and they end up together if they reject 1). If 1 makes an offer to $\{1, 5\}$, 5 will

accept, since 5 prefers to be with 1 than to be with 6 (which is the outcome if 5 rejects 1's offer). Consequently, 1 will make an offer to $\{1, 5\}$ in equilibrium, and 5 will accept, forming the coalition $\{1, 5\}$. Now, by the time 2 gets to move, 1 and 5 are off the market. Suppose that 2 and 3 then make offers which are rejected. If 4 then makes an offer to $\{3, 4\}$, 3 will accept, because otherwise both will end up by themselves (since 6 will make an offer to $\{2, 6\}$). Since 3 accepts, the coalitions $\{3, 4\}$ and $\{2, 6\}$ form in this subgame, with the resulting SPNE outcome in this subgame being $\{\{1, 5\}, \{3, 4\}, \{2, 6\}\}$. Backing up, suppose it's 3's turn to make an offer. If 3 offers to $\{2, 3\}$, 2 will accept, because otherwise 2 ends up with 6. Since 2 is 3's most preferred player, the coalition $\{2, 3\}$ will then form. Consequently, the SPNE of the subgame in which 2 is rejected after 1 and 5 coalition up is $\{\{1, 5\}, \{2, 3\}, \{4, 6\}\}$. Finally, suppose that 2 makes an offer to $\{2, 4\}$, its most preferred remaining coalitionmate. 4 will then accept, since rejecting the offer will cause 4 to be coalitioned up with 6, who is less preferred than 2. Consequently, the coalitions $\{2, 4\}$ and $\{3, 6\}$ will form. This means that the following outcome is a **SPNE outcome of the full game:** $\{\{1, 5\}, \{2, 4\}, \{3, 6\}\}$.

Proof of Lemma 1   We prove this by induction, after noting that A_R is unique by Remark 1.

Base Case: Suppose that the coalition T has been proposed. Consider an arbitrary sequential order of accept/reject decisions for players in T. Suppose that i is last in that order and all players before i have accepted. Then i will clearly accept since for any $\pi_R \in \Pi_R$, by assumption either $T \succ_i \pi_{R,i}$ or, if $T = \pi_{R,i}$ and, from lexicographic time preferences this holds even if, in a further subgame, another proposer proposes T and it is accepted.

Inductive Step: Consider a player i such that none of the players $k < i$ in the accept/reject order have rejected. Our inductive hypothesis is that if i accepts, then in every SPNE of the residual T-subgame all players $k' > i$ (which follow i in the order) accept. It is immediate that i's unique optimal strategy is then to accept, since for any $\pi_R \in \Pi_R$ either $T \succ_i \pi_{R,i}$, or $T = \pi_{R,i}$, and acceptance is preferred by lexicographic time preferences. The final step is to observe that when i is the first player in the order, none of the players before i have rejected, because no one precedes i.

Proof of Theorem 1   We prove this by showing the result for a subgame with only one remaining proposer and then appealing to backward induction.

Base Case: Consider an arbitrary subgame with only one player, i, who can still make a proposal and the set of feasible coalitions for i, denoted by $\mathcal{T}_i$ (none of the others matter). We show that in this subgame in every SPNE all proposals are accepted and result in a unique outcome. First, define $\mathcal{T}_i^{IR} = \{T \in \mathcal{T}_i | T \succ_j \{j\} \forall\, j \in T\} \cup \{i\}$, that is, a subset of feasible coalitions in which every coalition is preferred by all its members over being by themselves unioned with $\{i\}$. Clearly, every coalition offer $T \in \mathcal{T}_i^{IR}$ other than $\{i\}$ will be accepted. Let T_i^* be i's most preferred coalition in $\mathcal{T}_i^{IR}$. If $T_i^* = \{i\}$, by lexicographic preferences i strictly prefers to propose to and to accept coalition $\{i\}$. Otherwise, $T_i^* \succ_i \{i\}$. Because all $j \in T_i^*$ accept and form

a coalition, coalitions which have been formed thus far are fixed, and any remaining players become singletons, the subgame has a unique SPNE outcome.

Now consider the player who is the next to last proposer. Standard backward induction for extensive games with perfect information can now be applied and the above result holds for the "rolled back" game. This can be continued until the first player in the ordering O is to make an offer, which proves the result.[15]

---

[15]It is interesting to note some differences between this Theorem and Zermelo's Theorem and its extensions presenting uniqueness results for SPNE of extensive form games with perfect information. Zermelo's Theorem requires strict preferences and each terminal node of the game is unique. We do not necessarily have uniqueness of each terminal node and players may be indifferent between some terminal nodes – those that assign them to the same coalition.