Highlights (for review)

1. We introduce a dynamic game of coalition formation (with the property, that if players are to join a particular coalition then they prefer to join it sooner rather than later) and show that the SPNE coalition structure is unique and matches soulmates.
2. Surprisingly, with a simple condition on the number of proposals that can be made by each player, the SPNE outcome is both Pareto-optimal and individually rational.
3. Sufficient conditions are provided for the SPNE coalition structure to coincide with the core of the induced cooperative game.
4. We design a mechanism to implement equilibrium of the game and provide sufficient conditions to ensure truthful reporting of preferences is a strong Nash equilibrium.
5. We show empirically that players rarely have an incentive to misreport preferences more generally.
6. For the problem with cardinal preferences, we show empirically that the resulting mechanism results in significantly higher social welfare than serial dictatorship, and the outcomes are highly equitable

Manuscript

Non-Cooperative Team Formation and a Team Formation Mechanism

Matthew Chambers[b], Chen Hajaj[a], Greg Leo[b], Jian Lou[a], Martin Van der Linden[b], Yevgeniy Vorobeychik[a], Myrna Wooders[b]

[a] *Department of Electrical Engineering and Computer Science, Vanderbilt University*
[b] *Department of Economics, Vanderbilt University*

Abstract

We model decentralized team formation as a game in which players make offers to potential teams whose members then either accept or reject the offers. The games induce no-delay subgame perfect equilibria with unique outcomes that are individually rational and match soulmates. We provide sufficient conditions for equilibria to implement core coalition structures. When each player can make a sufficiently large number of proposals, we obtain the novel and surprising result that outcomes are Pareto optimal. We then design a mechanism to implement equilibrium of this game and provide sufficient conditions to ensure that truthful reporting of preferences is a strong ex post Nash equilibrium. Moreover, we show empirically that players rarely have an incentive to misreport preferences more generally. Furthermore, for the problem with cardinal preferences, we show empirically that the resulting mechanism results in significantly higher social welfare than serial dictatorship, and the outcomes are highly equitable.

JEL classification: C72, C63, C71, C78, C62

Keywords: Team Formation, Coalition Formation, Mechanism Design, Subgame Perfection, Pareto Optimality

☆*E-mail addresses:* matthew.l.chambers@vanderbilt.edu (M. Chambers), chen.hajaj@vanderbilt.edu (C. Hajaj), g.leo@vanderbilt.edu (G. Leo), jian.lou@vanderbilt.edu (J. Lou), martin.van.der.linden@vanderbilt.edu (M. Van der Linden), yevgeniy.vorobeychik@vanderbilt.edu (Y. Vorobeychik), myrna.wooders@vanderbilt.edu (M. Wooders).

"*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

1. Introduction

Whether business, social, or recreational, activity is often organized into groups through informal, decentralized processes. In school, students find their own study groups, roommates, project partners, and homecoming dates. Similar informal procedures are often responsible for matching business partners, research collaborators, and tennis doubles teams.

We model decentralized matching as a sequential bargaining game[1], with the restriction that our game has non-transferable utility. In our game, a proposer invites a subset of players to join her in a team. Players in the proposed team then sequentially accept or reject the invitation. If all accept, the team is formed, and all associated players are removed from the game. Then another player (or potentially the same player if her team was rejected) becomes the proposer, and the process continues for a predefined number of rounds or until no players remain who have not become members of a team. At a terminal node, players who remain unassigned to a team at that node, are assigned to singleton teams.

We analyze this game under the condition that players have perfect information about each other's preferences. First, we demonstrate two general properties that such games possess: there is a unique subgame perfect Nash equilibrium (SPNE) team structure and every SPNE involves no delay in the formation of equilibrium teams. Both are quite surprising, given that SPNE tends to be a weak criterion in most prior game models (necessitating refinements, such as stationarity). Moreover, equilibrium outcomes are always individually rational. Our final general positive result for games with an arbitrary exogenous order of proposers, is that every SPNE *matches soulmates*, that is, any team that is most preferred by all its members is formed. Indeed, our result is stronger: it matches soulmates even in a recursive sense, where upon removing a

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

set of matched soulmates, new soulmate teams arise once player preferences no longer include matched players, and so on, until no soulmate teams remain; all such teams are formed in every SPNE of our game, independent of order over proposers. An important consequence of this is that whenever all players are so matched, the SPNE outcome coincides with the unique core outcome. In general, however, the games we consider do not select a core coalition structure.

Our results for Pareto optimal outcomes are especially interesting. We find that a restriction on our game model dictating that any proposer in the specific order must be able to propose at least $n + 1$ times, where n is the number of teams containing that proposer, ensures Pareto optimality of outcomes. We term the resulting restricted set of games "Rotating Proposer Games (RPGs)".

Our games bear resemblance to a number of previously studied models of non-cooperative coalition formation games, for example, Seidmann and Winter (1998); Ray and Vohra (1999); Bloch (1996); Chatterjee et al. (1993); Okada (1996); Bloch and Diamantoudi (2006). Most of these are TU games (Chatterjee et al., 1993; Okada, 1996; Seidmann and Winter, 1998; Ray and Vohra, 1999), and all assume an infinite horizon. Moreover, most make a consequential assumption about the order of proposers where the first player to reject a proposal becomes the proposer in the next round. In contrast, we consider finite-horizon games in which the proposer order is fully exogenous.

As an alternative to fully decentralized coalition formation, a planner may design a centralized matching process in which players report their hedonic preferences over teams, and a centralized mechanism returns a partition of players into teams and, if need be, singleton sets. A significant advantage of the centralized process is that it is now natural to allow for incomplete information about player preferences. This, however, poses significant theoretical challenges, discussed below.

A natural option, which is unlike most approaches to the problem in prior literature, is to simply centralize a decentralized mechanism that has good equilibrium properties. In instances where preference reports are truthful, the outcome properties are the same as those of the equilibrium of the decentralized mechanism.

We take this approach and introduce a new mechanism, the Rotating Proposer Mechanism (RPM), that centralizes the rotating proposer game. RPM allows us to achieve Pareto efficiency, individual rationality, and IMS. Our empirical work demonstrates that these can be achieved with only a small relaxation of incentive compatibility. While RPM involves a substantial computational burden to implement exactly, we develop several approximations that, by construction, maintain individual rationality and IMS (details provided in Supplementary Materials). Using an algorithm for finding an upper bound on the number of untruthful players (also in Supplementary Materials), we show that RPM and its approximate versions introduce few incentives for manipulation in several classes of the roommate problem, as well as settings with 3-player teams. In extensive experiments, we evaluate the properties of RPM in both exact and approximate versions, in terms of social welfare (a much stronger notion than Pareto optimality, using cardinal preferences over coalitions) and fairness (using several natural notions thereof). We show that in comparison with random serial dictatorship, which serves as a calibration baseline for empirical results, RPM achieves high social welfare and has desirable equity properties.

To place our contribution in the literature and highlight some of the problems faced, we briefly note some of the literature. Alcalde and Barberà prove that without restrictions on the sets of admissible preferences, there is no matching mechanism that is Pareto optimal, individually rational, and strategyproof, even in two-sided problems (a special case of our team-formation environment) (Alcalde and Barberà, 1994). Rodriguez-Alvarez shows that any mechanism that is strategyproof and individually rational must either be bossy or put restrictions on which partitions can form (Rodriguez-Alvarez, 2004). Moreover, Leo et al. prove that any mechanism that matches soulmates cannot be strategyproof on general preference domains (Leo et al., 2017).

A number of mechanisms have been proposed to achieve subsets of the desired properties of efficiency, incentive compatibility, individual rationality, and matching of soulmates. Aziz et al. present a class of mechanisms that are Pareto optimal and individually rational (Aziz et al., 2013). Specific instances of this class can be selected to effect other properties, such as improved fairness (without formal guarantees) and even strategyproofness in restricted settings. In the context of roommate problems (teams of two), Biro et al. exhibits mechanisms which are both Pareto optimal and implement iterative matching of soulmates (IMS) (Biro et al., 2016).[2] Pareto efficiency and individual rationality in roommate problems can also be achieved either by almost stable matchings (Abraham et al., 2006) or least-unpopular matchings (Manlove, 2013). Wright and Vorobeychik empirically evaluated several mechanisms for team formation, but offer few theoretical guarantees (Wright and Vorobeychik, 2015). The literature thus shows that guaranteeing even individually rationality and efficiency requires relaxing incentive compatibility, justifying our interest in the "small" relaxation of incentive compatibility.

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 $N = \{1, \dots, n\}$ who must be partitioned into teams.[3] A *team* $T \in 2^n$ is a set of players, and a *team structure* π is a partition of the total player set into teams. For a player i, let π_i be the team in the partition π containing i.

In many situations teams face some feasibility constraints; for example, teams may be constrained to consist of at most k individuals. Generically, let $\mathcal{T}$ denote the set of *feasible* teams, which we assume to always include singleton teams, $\{i\}$. For a player i, we denote the subset of feasible teams that 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 team T at time t is strictly preferred to joining T at time t'.[4]

[2] Actually, these satisfy a more general property of maximum irreversibility.

[3] We use the terms teams and coalitions interchangeably throughout.

[4] 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 (2006).

2.2. Non-Cooperative Coalition Formation Game

We model the decentralized process of hedonic coalition formation using a natural non-cooperative game with perfect information.[5] In the game, players sequentially propose teams that are then accepted or rejected by their prospective members. We term such games *accept-reject games (ARGs)*.

Formally, an ARG is *a game of perfect information in extensive form with player set* N, a set of feasible teams $\mathcal{T}$, a preference profile $\succ$ *over team structures*, 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.[6] The ordering determines the order in which players can make proposals to other players (or to themselves alone) to form teams. Each proposal and its responses lead to a subgame. The game begins with the first player in the ordering, say i, proposing a team $T \in \mathcal{T}_i$. 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 their places in the ordering).[7] If one or more players in T reject the proposal, we arrive at a subgame in which any players in T who still had proposals to make can do so, when it is their turn in the ordering O, and can still accept or reject proposals made to them to join teams. In either case, 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 team).

The game proceeds to the next proposer in the ordering O. The process continues until there are no more opportunities for teams to form – either (a) all players are in teams or (b) the m_{th} player in the order has made a proposal and players to whom the proposal is made have responded. In case (b), the remaining players, if any, become singleton teams. In either case, the outcome is a team structure.

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

[5]A hedonic game is simply a game with ordinal preferences over teams of membership.

[6]The order of the players in O is arbitrary; for example, if $N = \{1, 2, 3\}$ the ordering O may be $(3, 1, 2)$.

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

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

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

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 $\pi = \{\{1, 2\}, \{3, 4\}\}$. Note that this partition is also a subgame perfect Nash equilibrium (SPNE) of this ARG.

2.3. Equilibrium Properties of ARGs

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

- individual rationality (players are not a part of any team if they would prefer to be by themselves),
- matching of soulmates (players who all prefer to be together are matched, even in a more general sense discussed below), and
- when the game is "IMS-complete" (see below), the outcomes are in the core of the derived cooperative game.

Surprisingly, 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, $N \backslash N'$ 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 team memberships among these players while maintaining the same team memberships of all remaining players, if any.

We begin our analysis with two examples that illustrate the subtleties involved in analyzing ARGs. What is particularly revealing is that small and seemingly inconsequential changes solely to the order of proposals can effect significant changes in equilibrium outcomes. The following example illustrates a SPNE with an outcome that is not Pareto optimal.

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

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

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.

Example 3. *Consider now a very minor modification of Example 2: we let player* 1 *move twice in the very beginning rather than just once. 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 teams) remains the same. We now show that this modification results in teams 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 this in mind, suppose that* 1 *makes an offer to* 3 *at the very beginning. If* 3 *rejects, it is teamed up with* 6 *in the resulting subgame. Clearly,* 3 *strictly prefers to be on a team with* 1*, and would therefore accept. Once the team* $(1, 3)$ *is formed,* 2 *and* 5 *prefer to be with one another rather than with anyone else, and the resulting team 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.*

It is instructive to observe that in the above example the SPNE outcome, if the first proposal by 1 were rejected, serves as a kind of credible threat to player 3. This turns out to have significant consequences for optimality, as we show below.

We next proceed to prove several interesting and useful characteristics of subgame prefect Nash equilibria of ARGs, as well as some properties of their equilibrium outcomes. We start with some additional notation. Define a *T-subgame* as the subgame of an ARG 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 ARG is itself an ARG, 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 simply has no more choices to make; this holds for all those players who, at prior decision nodes, joined teams. 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.

Observation 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 team T and who all prefer T to the outcome that would materialize if this team were rejected. We show that in an SPNE such a team 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 team 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 team that is better for all players in the collection. In contrast, with sequential decision-making the team 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.[8]

Lemma 1. *Consider a T-subgame of an arbitrary subgame A for a proposal of team 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 $T = \pi_{R,i}$.[9] Then all $i \in T$ will accept T in every SPNE of the T-subgame.*

We relegate the proof of Lemma 1 to 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. It is an immediate corollary to the following Theorem.

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

The proof of Theorem 1 is obtained by backward induction and is relegated to Appendix A.

Remark: The above result shows that an SPNE has the property that at every proposer node, the SPNE offer is accepted. Of course a strategy must still specify what

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

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

happens at every other node of the tree, including nodes that would follow a rejection of a proposal.

2.4. The Coalitional Game

Having characterized the structure of subgame perfect Nash equilibria of an ARG, we now consider whether the unique outcome satisfies important properties of the corresponding coalitional game. The first of these is individual rationality. This, it turns out, 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. *In every SPNE outcome of any ARG, each player* i *is at least as well off as in the singleton team* $\{i\}$.

The set of players N and their preferences $\succ$ determines a (hedonic) cooperative game of coalition formation. An assignment π of players to teams is in the *core* of this cooperative team formation game if there does not exist a coalition of players $T \subset N$ with the property that for all $i \in T$, $T \succ_i \pi_i$. An interesting question is whether, if the core of cooperative game that can be formed from the information on preferences is nonempty, equilibrium outcomes always result in a core coalition structure. As the following example demonstrates, this is not the case, even for bipartite matching problems.

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

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

Suppose that all of the above teams are feasible, and that the order of proposers is $O = (1, 2, 3, 4, 5, 6)$. *It is not difficult to ascertain that the unique SPNE outcome (under lexicographic preferences) of this game 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, we now show that ARG equilibria implement another important property, *iterated matching of soulmates (IMS)* (Leo et al., 2017). This, it turns out, leads to a sufficient condition to guarantee that ARG outcomes are in the core.

IMS captures the idea that a set of players who, among the set of players not already in teams, all prefer to be with each other, are natually matched. Formally, a team T is a team of (1st order) *soulmates* if for all $i \in T$, $T \succ T'$ for all $T' \in \mathcal{T}_i$. Iteratively applying this criterion we obtain IMS: in every iteration, we match all teams 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 perceived.[10] A more formal motivation is that all teams matched by IMS are blocking coalitions (Leo et al., 2017), and players in blocking coalitions may create instability.[11] Moreover, implementing IMS has important consequences for incentive compatibility and core stability. Next, we show that ARG subgame perfect Nash equilibrium outcomes always match soulmates in this iterative sense. More precisely, let $\hat{\mathcal{T}}_{IMS}$ be a collection of teams produced by IMS. We say that ARG *implements IMS in SPNE* partition π if $\hat{\mathcal{T}}_{IMS} \subseteq \pi$.

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

Proof. We prove this by induction.

[10]Of course, in some situations, it may not be desirable to match soulmates. For example, in forming sports teams in a school, a "captains mechanism," in which two captains sequentially choose team members, may be preferable.

[11]As shown by Leo et al. (2017), the assumption that all players can be matched as soulmates is weaker than the top coalition property of Banerjee et al. (2001).

Base Case: We show that every soulmate team must be formed by any SPNE. We prove this by contradiction.

Consider a SPNE s in which all proposals are accepted (sufficient, since such a 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 team 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 team to which i is assigned by s. Suppose i proposes to T. By Lemma 1 and the fact that T is a team 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 teams of kth order soulmates (i.e., from the first k rounds of IMS) are formed. We now show that all soulmate teams from $k+1$st round form as well. We do this by a similar contradiction argument as the base case.

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

As shown by Leo et al. (2017), 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 are matched by IMS. Then every SPNE of an arbitrary ARG yields the unique core coalition structure.*

2.5. The Rotating Proposer Game

As we showed in Example 2, SPNE outcomes of an arbitrary ARG 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 ARGs—specifically, restricting the orderings over proposers—which allows us to guarantee that the outcome is always Pareto optimal.

Specifically, we propose a class of ARGs which we term *rotating proposer games (RPGs)*. In an RPG, 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 efficiency.[12]

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 same SPNE outcome obtains as in that example (i.e.,* $\{(1, 5), (2, 4), (3, 6)\}$*). Consequently, as in Example 3, if* 1 *proposes to* 3*,* 3 *will accept, and the resulting SPNE outcome of the RPG 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 RPG is Pareto optimal.*

Before we prove Theorem 2, we make several observations.

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

This follows from observing that if $\pi_{i,k+1} \succ_i \pi_{ik}$, then in SPNE, when i proposes

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

for the kth time, she should make a proposal which 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 teams for her to propose to.*

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

This follows because the subgame that follows i proposing to π_{ik} 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 π_{ik}) 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 $\pi_{ik} = \pi_{i,k+1}$, then $\pi_{i,k-i} = \pi_{ik}$. The result then follows from Observation 3. Assume $\pi_{ik} = \pi_{i,k+1}$.

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

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

Now since $\pi_{i,k-1} \succ_i \pi_{ik}$, we must have that if i proposes the team $\pi_{i,k-1}$ on her kth proposal, it is rejected. Otherwise, she would propose π_{k-1} and be accepted, making herself better off. So 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 $\pi_{jk} \succ_j \pi_{i,k-1} \succ_j \pi_{jk}$, a contradiction. Thus, it is not the case that $\pi_{i,k-1} \succ_i \pi_{ik}$, so it must be that $\pi_{i,k-1} = \pi_{ik}$, implying that $\pi_{i,k-1} = \pi_{ik}$.

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

We now proceed to prove Theorem 2 by contradiction. Let π be the SPNE outcome of an RPG. Suppose, for contradiction, that the set of teams 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 teams in π and π'. Let Q be the set of such players, and let $i \in Q$ be the first such player to propose.

Claim. 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 teams 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 team π'_i, all members of the team 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 RPG, since the SPNE outcome is unique by Theorem 1. $\square$

However, while RPGs do resolve the Pareto optimality issue, Example 4 can be extended to use the RPG order and 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).

In summary, RPG equilibrium outcomes are individually rational and implement IMS (inherited from general ARGs), and, in addition, are Pareto optimal. Moreover, all equilibria have no delay of forming teams, and result in a unique outcome. From the perspective of decentralized hedonic coalition formation with complete information, this is a strong set of properties. However, complete information is a strong assumption, one we would in practice wish to relax. To do so, we proceed to consider a *centralized* (mechanism design) approach to team formation that turns RPGs into a direct mechanism by implementing a SPNE in which, given a collection of *reported* preferences, all offers are accepted.

3. The Rotating Proposer Mechanism

In order to move to a centralized team formation setting, we need to formally define a *team formation mechanism*. A *team formation mechanism* M maps every preference profile $\succ$ to a partition π, i.e. $\pi = M(\succ)$. Our goal is to exhibit such a mechanism, and analyze its properties. The mechanism, termed Rotating Proposer Mechanism (RPM), implements the subgame perfect Nash equilibrium of the RPG in which all proposals are accepted. In this equilibrium, whenever it's a player i's turn to propose, i makes a proposal to her most preferred team among those that would be accepted.

For any profile, if all players report their preferences truthfully, equilibrium outcomes of the game have a number of good properties which are thereby inhereted by RPM. Of particular note is that RPM is individually rational, Pareto optimal, and implements IMS. However, it is also immediate from known results that the RPM mechanism is not in general strategyproof (this would conflict with individual rationality and implementing IMS (Leo et al., 2017)).

The loss of incentive compatibility seems problematic. However, one side-effect of RPM implementing IMS is that RPM is strongly incentive compatible[13] and yields a unique core team structure on a restricted class of preference domains for which IMS always matches all players (Leo et al., 2017). As an example, these domains include other well-known restrictions on preferences, such as top coalition (Banerjee et al., 2001) and common ranking (Farrell and Scotchmer, 1988) properties.

This, however, would seem to limit its practical consideration, as such restrictions can rarely be guaranteed or verified. Moreover, we wish to make stronger efficiency claims than Pareto optimality, and also view fairness as an important criterion. For the former, we are particularly interested in *utilitarian social welfare*, a much stronger criterion than Pareto efficiency. We will also consider several notions of fairness discussed

[13]More precisely, truth telling is a strong ex post Nash equilibrium.

below.

While we cannot make strong theoretical guarantees about these for broad realistic preference domains, we consider such properties empirically.

3.1. Empirical Methodology

In our empirical assessments, we use both synthetic and real hedonic preference data. In both cases, preferences were generated based on a social network structure in which a player i is represented as a node and the total order over neighbors is then generated randomly. Non-neighbors represent undesirable teammates (i would prefer being alone to being teamed up with them).

The networks used for our experiments were generated using the following models:

- **Scale-free network:** We adapt the Barabási-Albert model (Albert and Barabási, 2002) to generate scale-free networks. For each (n, m), where n is the number of players, m denotes the density of the network, we generate $1,000$ instances of networks and profiles.
- **Karate-Club Network (Zachary, 1977):** This network represents an actual social network of friendships between 34 members of a karate club at a US university, where links correspond to neighbors. We generate 100 preference profiles based on the network.

Finally, we used a *Newfrat* dataset (Nordlie, 1958) that contains 15 matrices recording weekly sociometric preference rankings from 17 men attending the University of Michigan. In order to quantitatively evaluate both the exact and approximate variants of RPM, the ordinal preferences $\succ_i$ have to be converted to cardinal ones $u_i(\cdot)$, upon which both mechanisms operate. For this purpose, we introduce a *scoring function* suggested by Bouveret and Lang (2011) to measure a player's utility. To compute a player i's utility of player j we adopt *normalized Borda scoring function*, defined as $u_i(j) = g(r) = 2(k - r + 1)/k - 1$, where k is the number of i's neighbors, and $r \in \{1, \ldots, k\}$ is the rank of j in i's preference list. Without loss of generality, for every player i we set the utility of being a singleton $u_i(i) = 0$. We assume that the preferences of players are additively separable (Banerjee et al., 2001), which means that a player i's utility of a team T is $u_i(T) = \sum_{j \in T} u_i(j)$.

We begin by discussing computational issues associated with RPM, and then proceed to consider its economic properties.

3.2. Computational Considerations for Implementing RPM

While RPM is a rather intuitive mechanism, it is quite challenging to implement the associated subgame perfect Nash equilibrium. In particular, the size of the backward induction search tree is $O(2^{\sum_{i=1}^{n} |\mathcal{T}_i|})$. Even in the roommate problem, in which the size of teams is at most two, computing SPNE is $O(2^{n^2})$. We address this challenge in three ways: (1) preprocessing and pruning to reduce the search space, (2) approximation for the roommate problem, and (3) a general heuristic implementation.

3.3. Summary of Empirical Results

We now summarize the main empirical results. Full details can be found in Supplementary Materials.

Under RPM, incentives to misreport preferences are rare. For the roommate problem, we find that typically 1% of players or fewer have an incentive to misrepresent their preferencs, and fewer than 3% of all randomly generated profiles have *any* such players. Approximate versions of RPM (which enable implementation of this mechanism at a larger scale) do not much degrade these results. With teams of (at most) three, our experiment reflect only approximate RPM, and we find that no more than 5% of players have an incentive to misreport preferences.

RPM is highly efficient. We compare utilitarian social welfare of RPM (using a cardinal transformation of ordinal preferences) and its approximations to serial dicatorship (which is also Pareto optimal). We observe that in all experiments RPM yields much higher social welfare, with improvements typically ranging between 15 and 20%.

RPM yields equitable outcomes. As is well known, serial dictatorship results in highly inequitable outcomes (in the ex post sense). We observe that RPM yields outcomes far more equitable, with significant improvements in terms of the Gini coefficient, and a dramatically lower correlation between a random proposer order and utility (for example, correlation in some experiments drops from over 0.4 to well below 0.05).

4. Discussion and Conclusions

While the issue of bargaining for coalition formation has received considerable attention, much of this has focused on games with transferable utility and almost all on games with infinite horizons. In contrast, we consider sequential NTU games with a finite horizon, in which players iteratively propose teams, 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 an essentially unique no-delay equilibrium (all proposals are accepted in every equilibrium), and the equilibrium outcome is unique. Our second major positive result is that in a subgame perfect Nash equilibrium teams involving soulmates, even in a stronger iterative sense, are always formed. While this result 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. Finally, we exhibit a restricted class of games, where the restriction is on the exogenously specified order of proposers, in which equilibrium outcomes are Pareto optimal.

The finiteness of the game enables us to convert the perfect information decentralized team formation setting into a mechanism in which players submit their preferences, are randomly ordered, and the unique equilibrium outcome is implemented. The resulting mechanism thereby always yields individual rationality, Pareto optimality and, in an important special case, is incentive compatible and yields the unique core allocation. Moreover, we experimentally illustrate that this mechanism yields high social welfare compared to Random Serial Dictatorship (which is also Pareto optimal) when utilities are cardinal, and is significantly more equitable. As significantly, we show that while the mechanism is not in general incentive compatible, incentives to misreport preferences are relatively rare. Finally, we address the computational challenge of implementing the resulting equilibrium outcome by devising an approximate implementation of this mechanism which performs well in experiments, and retains most of the theoretical guarantees.

Acknowledgments

Professors Vorobeychik and Wooders grateful acknowledge support for this research from the NSF Grant IIS-1526860. Wooders also acknowledges support from the Douglas Grey Fund for Research in Economics, and Vorobeychik also acknowledges partial support from the NSF Grant IIS-1649972 and ONR Grant N00014-15-1-2621. Dr. Van der Linden was a PhD student in Vanderbilt Economics before moving to Utah State. We are grateful to participants at the conferences and universities where this work was presented.

References

Abraham, D.J., Biro, P., Manlove, D.F., 2006. "almost stable" matchings in the roommates problem, in: International Workshop on Approximation and Online Algorithms, pp. 1–14.

Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97. URL: `http://link.aps.org/doi/10.1103/RevModPhys.74.47`, doi:`10.1103/RevModPhys.74.47`.

Alcalde, J., Barberà, S., 1994. Top dominance and the possibility of strategy-proof stable solutions to matching problems. Economic Theory 4, 417–435. URL: `http://dx.doi.org/10.1007/BF01215380`, doi:`10.1007/BF01215380`.

Aziz, H., Brandt, F., Harrenstein, P., 2013. Pareto optimality in coalition formation. Games and Economic Behavior 82, 562–581.

Bade, S., 2015. Serial dictatorship: The unique optimal allocation rule when information is endogenous. Theoretical Economics 10, 385–410.

Banerjee, S., Konishi, H., Sönmez, T., 2001. Core in a simple coalition formation game. Social Choice and Welfare 18, 135–153. doi:10.1007/s003550000067.

Biro, P., Inarra, E., Molis, E., 2016. A new solution concept for the roommate problem: Q-stable matchings. Mathematical Social Sciences 79, 74–82.

Bloch, F., 1996. Sequential formation of coalitions with fixed payoff division and externalities. Games and Economic Behavior 14, 90–123.

Bloch, F., Diamantoudi, E., 2006. Noncooperative formation of coalitions in hedonic games. International Journal of Game Theory 40, 263–280.

Bouveret, S., Lang, J., 2011. A general elicitation-free protocol for allocating indivisible goods, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume One, AAAI Press. pp. 73–78. URL: http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-024, doi:10.5591/978-1-57735-516-8/IJCAI11-024.

Chatterjee, K., Dutta, B., Ray, D., Sengupta, K., 1993. A noncooperative theory of coalitional bargaining. The Review of Economic Studies 60, 463–477.

Farrell, J., Scotchmer, S., 1988. Partnerships. Quarterly Journal of Economics 103, 279–297.

Leo, G., Lou, J., Van der Linden, M., Yevgeniy, Wooders, M., 2017. Matching soulmates. Working paper, Available at SSRN: https://ssrn.com/abstract=2833553.

Lorenz, M.O., 1905. Methods of measuring the concentration of wealth. Publications of the American statistical association 9, 209–219.

Manlove, D.F., 2013. Algorithmics of Matching Under Preferences. World Scientific Publishing Company.

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

Nordlie, P., 1958. A longitudinal study of interpersonal attraction in a natural group setting. Ph.D. thesis. University of Michigan.

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

Ray, D., Vohra, R., 1999. A theory of endogenous coalition structures. Games and Economic Behavior 26, 286–336.

Rodriguez-Alvarez, C., 2004. On the Impossibility of Strategy-Proof Coalition Formation Rules. Economics Bulletin 4, 1–8. URL: `https://ideas.repec.org/a/ebl/ecbull/eb-04d70004.html`.

Rubinstein, A., 1982. Perfect equilibrium in a bargaining model. Econometrica 50, 97–109.

Seidmann, D.J., Winter, E., 1998. A theory of gradual coalition formation. Review of Economic Studies 65, 793–815.

Stahl, I., 1972. Bargaining Theory. Stokholm School of Economics.

Stahl, I., 1977. An n-person bargaining game in the extensive form, in: Mathematical economics and game theory. Springer, pp. 156–172.

Wright, M., Vorobeychik, Y., 2015. Mechanism design for team formation, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), pp. 1050–1056. URL: `http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9902`.

Zachary, W., 1977. An information flow model for conflict and fission in small groups. Journal of Anthropological Research 33, 452–473.

Zermelo, E., 1913. Über eine anwendung der mengenlehre auf die theorie des schachspiels, in: Proceedings of the fifth international congress of mathematicians, II, Cambridge UP, Cambridge. pp. 501–504.

Appendix A. Additional Proofs

Details of Example 2. Suppose that all of the above teams 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 team 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 teammate, and by the preceding logic. Consequently, any offer by 5 to 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, the team $(3, 4)$ will form if 3 accepts or teams $(3), (4)$ 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, with the outcome being $\{(1, 6), (2, 5), (3, 4)\}$. It is thus an equilibrium of this subgame for 4 to offer to 3, 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, 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 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, 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, team $(1,2)$ will form. Once this happens, $(3,4)$ will team up since they are then conditional soulmates, which implies that $(5,6)$ will team 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 3 or 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 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 5 in equilibrium, and 5 will accept, forming the team $(1,5)$. 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, 3 will accept, because otherwise both will end up by themselves (since 6 will make an offer to 2). Since 3 accepts, the teams $(3,4)$ 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, 2 will accept, because otherwise 2 ends up with 6. Since 2 is 3's most preferred player, the team $(2,3)$ will then form. Consequently, the SPNE of the subgame in which 2 is rejected after 1 and 5 team up is $\{(1,5),(2,3),(4,6)\}$. Finally, suppose that 2 makes an offer to 4, its most preferred remaining teammate. 4 will then accept, since rejecting the offer will cause 4 to be teamed up with 6, who is less preferred than 2. Consequently, the teams $(2,4)$ 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 Observation 1.

Base Case: Suppose that the team 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 teams 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 teams in which every team is preferred by all its members over being by themselves unioned with $\{i\}$. Clearly, every team offer $T \in \mathcal{T}_i^{IR}$ other than $\{i\}$ will be accepted. Let T_i^* be i's most preferred team in $\mathcal{T}_i^{IR}$. If $T_i^* = \{i\}$, by lexicographic preferences i strictly prefers to propose to and to accept team $\{i\}$. Otherwise, $T_i^* \succ_i \{i\}$. Because all $j \in T_i^*$ accept and form a team, teams 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.[14]

[14] 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 Zermelo (1913) 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 team.

Supplementary Material

Click here to download Supplementary Material: supplement (2).pdf