Decentralized Multilateral Bargaining

Yuan Ju[*] Juan Vidal-Puga[†]

November 21, 2023

Abstract

We present a decentralized mechanism of multilateral negotiation that allows every player to make a proposal as well as accommodates counteroffers and partial agreements. Only local unanimity is required for reaching an agreement and players are not excluded even if their proposals have been rejected, both being key relevant features in most real-life negotiations. The role of planner becomes minimal in our mechanism compared to those in the literature. This leads to a new solution theory that synthesizes the alternating-offer bargaining model *a la Rubinstein (1982)* and the general non-transferable utility environment with n players, which strategically establishes the Nash bargaining solution for pure bargaining problems, the Shapley value for transferable utility games, and in general, the Shapley NTU value for nontransferable utility games.

Keywords: Shapley NTU value, nontransferable utility game, subgame perfect equilibrium, bargaining

1 Introduction

In this paper, we present a new strategic mechanism of multilateral bargaining to study the fundamental economic problem how benefits from cooperation will be distributed among the economic agents. The seminal work of Hart and Mas-Colell (1996) opened the door of the analysis of multilateral bargaining and advanced our understanding of the Shapley value (Shapley, 1953) and the consistent value (Maschler and Owen, 1989, 1992)

[*]Department of Economics and Related Studies, University of York, York, YO10 5DD, UK. yuan.ju@york.ac.uk

[†]Economics, Society and Territory (ECOSOT) and Departamento de Estatística e IO. Universidade de Vigo. 36200 Vigo (Pontevedra), Spain. vidalpuga@uvigo.es

from a non-cooperative perspective. Our starting point is to develop a more natural and practical mechanism, inspired by the actual negotiation processes in international politics, that requires the minimal role of a planner to allow for truly decentralized bargaining. In doing so, the paper offers three original contributions to the literature.

Firstly, we construct a genuinely decentralized bargaining model where players have much more freedom in making proposals and forming coalitions. Unlike the conventional approach in the literature that rarely accommodates partial agreements, our model has a weaker requirement on unanimity and allows for players to directly choose to join and form a certain coalition. Moreover, players will never be forced to leave the negotiation table even though their proposals have not been accepted by the others, which further makes our bargaining protocol independent from a planner. Indeed, when players can freely choose whom to form a coalition and are not forced to drop out from the negotiation after their proposals were rejected, the planner of the underlying mechanism has virtually no influence on the choices of the players. Thus, it would be a truly decentralized mechanism. By contrast, the role of a planner is much stronger in existing models such as Hart and Mas-Colell (1996) and Pérez-Castrillo and Wettstein (2001), where, for example, if a player's proposal is rejected, she will be (eventually) excluded from future negotiations.

Secondly, we offer a bargaining theory that synthesizes the alternating-offer bargaining model a la Rubinstein (1982) and the general non-transferable utility environment with n players. The alternating-offer bargaining model is widely accepted as a natural and robust bargaining protocol, and has been tremendously influential in the economic literature, particularly so in the development of the theory of the firm and labour economics (cf. Stole and Zwiebel (1996)). Alternating offers seems prevalent in all sorts of real-world negotiations, yet a systematic and effective modelling of offers being alternatively made in an n-player coalitional context has been an open problem for decades. Note that the alternating-offer feature hardly appears in the two notable multilateral bargaining models (and the variations) of Hart and Mas-Colell (1996) and Pérez-Castrillo and Wettstein (2001) that adopt a top-down perspective that effectively leave out the possibilities for players to make counter-offers: It starts immediately with the negotiation towards the grand coalition. When a player i makes an offer, all others can only agree or disagree. If rejected, player i will effectively leave the game and the others will negotiate only among themselves. We offer a constructive and bottom-up negotiation protocol where all players are allowed to make proposals for the rest to approve or reject. If one does not agree with the others, she can make an alternative proposal such that the others can consider it. This proposal could further be abandoned if "better" proposals are available. Thus, all players can make proposals, have opportunities to review others' proposals, and take up new proposals so long as they wish. So counter-offers are truly preserved and the interaction among players is much more extensive in our model.

Finally, through this multilateral alternating-offer bargaining mechanism we establish an integrated solution theory that has the Nash bargaining solution for pure bargaining problems, the Shapley value for transferable utility games, and in particular, the Shapley NTU value for the general non-transferable utility games. To the best of our knowledge, this paper is the first one that offers a non-cooperative bargaining foundation for the Shapley NTU value in the most general case, despite a previous attempt by Vidal-Puga (2008) that focuses on a restricted situation. Thus, the robustness of the Shapley's solutions is confirmed from both axiomatic and strategic perspectives, and its connection to the Nash bargaining solution is installed in a unified bargaining framework. Such a framework makes it possible for us to comprehend and review all these major solution concepts for different game environments in the same context, and helps discover the strategic difference between the Shapley NTU value and the consistent value (Maschler and Owen, 1989, 1992). Moreover, a natural extension to games with coalition structure is provided, which yields the Owen value Owen (1977).

Other desirable features of the mechanism include allowing for coalition formation and players proposing rules (called payoff configurations in Hart and Mas-Colell (1996)) instead of payoffs, which makes the mechanism general and flexible. By agreeing to a rule, players form a certain coalition, and consequently, a partition of players is formed, which can proceed in negotiation with other players. Hence, the model itself does not place any force towards a grant coalition. By having players proposing rules, we effectively admit proposing payoffs since the latter can be viewed as the outcome of a specific rule. More importantly, this is closer to real world negotiations. It is common that players may not be restricted to a specific payoff, but find a range of payoffs, or general policies, a set of principles or regulations, and solutions potentially agreeable. Confining to final payoffs may restrict the scope of negotiation and affect the likelihood of reaching an agreement. Proposing rules removes such an imposed condition and makes the modelled negotiation practical. It is common to see that in real life countries negotiate on terms or clauses, rather than directly on payoffs, in order to reach an agreement. In the literature, proposing rules can be found as early as in, among others, van Damme (1986) that designs a two-player bargaining game in which players' offers are solution concepts such as the Nash bargaining solution or the Kalai-Smorodinsky solution. Vidal-Puga (2008) also applies this idea in the general n-player NTU environment, and de Clippel and Serrano

(2008) characterize a payoff configuration for partition function form games on the basis of the balanced contributions.

The paper is organized as follows. Section 2 presents the basic notation. Section 3 presents the model, where we construct two decentralized multilateral negotiation mechanisms both possessing the appealing features mentioned above. To help illustrate the details of the mechanisms, Section 4 offers an example that was well studied in the literature. Section 5 contains the main results. Section 6 provides some concluding remarks.

2 Preliminaries

Let $N = \{1, ..., n\}$ be a finite set of players. A coalition is a subset of N. Given $S \subset N$ and $x \in \mathbb{R}^N$, we define $x_S \in \mathbb{R}^N$ as the restriction of x to the coordinates in S. Given $S, T \subset N$ with $S \cap T = \emptyset$, $x \in \mathbb{R}^S, y \in \mathbb{R}^T$, we define $x \times y \in \mathbb{R}^{S \cup T}$ as $(x \times y)_i = x_i$ for all $i \in S$ and $(x \times y)_i = y_i$ for all $i \in T$, and given $X \subset \mathbb{R}^S$ and $Y \subset \mathbb{R}^T$, we define

$$X \times Y = \{x \times y : x \in X, y \in Y\}.$$

Let Π denote the set of all orders of players in N with generic element $\pi = (\pi_1, \ldots, \pi_n)$. Given $\pi \in \Pi$ and $i = \pi_k \in N$, let P_i^π be the set of predecessors of player i in π, i.e.,

$$P_i^\pi = \{\pi_l \in N : l < k\}$$

and let $\overline{P_i^\pi} = P_i^\pi \cup \{i\}$. Given $\mu \geq 0$, we define Λ_μ^N as the set of normalized N-dimensional vectors whose coordinates are bounded from below by μ, i.e.,

$$\Lambda_\mu^N = \left\{ \lambda \in \mathbb{R}^N : \sum_{i \in N} \lambda_i = 1, \lambda_i \geq \mu \forall i \in N \right\}.$$

A non-transferable utility (NTU) game is a mapping V that assigns to each coalition $S \in 2^N \setminus \{\emptyset\}$ a subset $V(S)$ of $\mathbb{R}^S$ such that the following properties hold:

(P1) $V(S)$ is a non-empty closed subset of $\mathbb{R}^S$.

(P2) $V(S)$ is comprehensive, i.e., if $x \in V(S)$ and $y \in \mathbb{R}^S$ such that $y \leq x$, then $y \in V(S)$.

(P3) $V(S)$ is bounded from above, i.e., for each $x \in \mathbb{R}^S$, the set $\{y \in V(S) : y \geq x\}$ is compact.

(P4) $V(N)$ is strictly non-leveled and smooth[1], i.e., there exists $\mu(V) > 0$ such that for each x in the Pareto frontier of $V(N)$ there exists a unique $\lambda \in \Lambda^N_{\mu(V)}$ outwards to $V(N)$ in x.

Without loss of generality, we normalize $\max V(\{i\}) = 0$ for all $i \in N$,.

We say that V is a *pure bargaining problem* if $0_N \in V(N)$ and 0_S is on the Pareto frontier of $V(S)$ for all $S \subset N$, where 0_S is defined as $(0_S)_i = 0$ for all $i \in S$.

Definition 2.1 *The* Nash solution *of a (normalized) pure bargaining problem given by* V *is the payoff allocation* $Nash(V) \in V(N)$ *that maximizes the product of its coordinates, i.e.,*

$$Nash(V) \in \arg \max_{x \in V(N)} \prod_{i \in N} x_i.$$

The Nash solution is always Pareto efficient, i.e., $Nash(V)$ belongs to the frontier of $V(N)$.

We say that V is a *transferable utility (TU) game* if there exists a mapping v that assigns to each coalition S a real number $v(S) \in \mathbb{R}$ such that

$$V(S) = \left\{ x \in \mathbb{R}^S : \sum_{i \in S} x_i \leq v(S) \right\} \tag{1}$$

for all $S \subseteq N$.

Properties (P1)-(P4) follow from (1). For example, (P4) holds with $\lambda_i = 1$ irrespectively of the Pareto allocation x.

Definition 2.2 *The* Shapley value *of a TU game given by* v *is the payoff allocation* $Sh(v) \in \mathbb{R}^N$ *defined as*

$$Sh(v)_i = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \left(v\left(\overline{P_i^\pi}\right) - v\left(P_i^\pi\right) \right)$$

for all $i \in N$.

The Shapley value is always Pareto efficient, i.e., $Sh(v)$ belongs to the frontier of $V(N)$ (or, equivalently, $\sum_{i \in N} Sh(v)_i = v(N)$).

For each $i \in N$ and $x \in \mathbb{R}^{N \setminus \{i\}}$ let $\tau_i(x)$ be defined as

$$\tau_i(x) = \max\{t \in \mathbb{R} : x \times (t) \in V(N)\}.$$

[1] Our results are not affected if we require non-levelness and/or smoothness for each $V(S)$.

Under smoothness and non-levelness (P4), τ_i is a well-defined, derivable, and strictly decreasing in each coordinate function from $\mathbb{R}^{N\setminus\{i\}}$ to $\mathbb{R}$.

Assume the utility of players can be transfered at a constant rate given by $\lambda \in \Lambda_0^N$. Then, a linear transformation of players' utilities leads to a TU game. If the Shapley value of this new game is feasible in the former game without the constant-rate transfer assumption, then we say that it is a Shapley NTU value.

Formaly, given $\lambda \in \Lambda_0^N$, we define the associated game V^λ as

$$V^\lambda(S) = \left\{ x \in \mathbb{R}^S : \sum_{i \in S} \lambda_i x_i \leq v^\lambda(S) \right\}$$

where

$$v^\lambda(S) = \max_{y \in V(S)} \sum_{i \in S} \lambda_i y_i$$

for all $S \subseteq N$.

Definition 2.3 *A payoff allocation $x \in \mathbb{R}^N$ is a Shapley NTU value of V if $x \in V(N)$ and there exists $\lambda \in \Lambda_0^N$ such that $\lambda_i x_i = Sh(v^\lambda)_i$ for all $i \in N$.*

An additional typical property for NTU games is the following:

(P5) (*Zero-monotonicity*) Given $i \in N \setminus S$, it holds $x \times y \in V(S \cup \{i\})$ for each $x \in V(S), y \in V(\{i\})$

Zero-monotoniciy in a TU game is equivalent to $v(S) \leq v(S \cup \{i\})$ for all $i \in N \setminus S$.

Under our normalization, zero-monotonicity can be re-stated as $x \times (0) \in V(S \cup \{i\})$ for each $S \subset N$ and $i \in N \setminus S$.

Under (P1)-(P5), we show that the Shapley NTU value always exists. As far as we know, there are no previous results of existence of the Shapley NTU value in such a general setting. We are also unaware of results of uniqueness, apart from some trivial cases such as pure bargaining problems and *TU* games. In general case uniqueness fails, see for example Hart and Mas-Colell (1996) for a non super-additive game.

It is well-known that the (unique) Shapley NTU value coincides with the Nash bargaining solution for pure bargaining problems and with the Shapley value for TU games.

3 Decentralized multilateral negotiation mechanism

Fix a zero-monotonic NTU game V. Players will decide how to share the benefit of their mutual collaboration, and they can choose whom to collaborate with, refusing to collaborate at all being always an option. Their payoffs will only depend on the identity of the players who actually collaborate.

Given $T \subseteq N$, we define a *rule supported by* T as a function f^T which assigns to each coalition $S \supseteq T$ a feasible payoff allocation for S, i.e., $f^T(S) \in V(S)$ for all $S \supseteq T$. Thus, a rule determines a *payoff configuration* (see Hart and Mas-Colell (1996)). However, a rule should not be interpreted as a set of payoff allocations, one for every coalition, but as an index that indicates the approved payoff allocation when a particular coalition of players has agreed to collaborate. We denote the set of all such rules as $\mathcal{F}^T$ and $\mathcal{F} = \bigcup_{T \subseteq N} \mathcal{F}^T$.

The strategic bargaining model that we are going to propose closely follows the spirit of Rubinstein's alternating offer protocol (Rubinstein, 1982), albeit necessarily proceeding in a more sophisticated manner concerning the n-person non-transferable utility environment where coalitional bargaining is allowed. Before we formally introduce our model, it would be useful to provide an informal and intuitive description to present the essential idea.

A random ordering of players is formed. For convenience, say it is $(1, 2, \ldots, n)$. Every player will be allowed to express their opinions, in the form of making proposals (i.e., rules as defined above) about how to share payoffs among them or simply accepting others' proposals to form coalitions. Firstly, player 1 makes a proposal $f^{\{1\}} \in \mathcal{F}^{\{1\}}$.

Then, it is for player 2 to express her opinion. If player 2 concurs with player 1's opinion, then she can accept $f^{\{1\}}$ and join player $\{1\}$ to form coalition $\{1, 2\}$, with their joint proposal supported by $\{1, 2\}$ to be $f^{\{1,2\}} \in \mathcal{F}^{\{1,2\}}$, where $f^{\{1,2\}}(S) = f^{\{1\}}(S)$ for all $S \supseteq \{1, 2\}$. If player 2 does not accept the proposal of player 1, then she will make a proposal $f^{\{2\}} \in \mathcal{F}^{\{2\}}$. After that, player 1 can say yes or no to $f^{\{2\}}$, in the spirit of alternating offer. If player 1 accepts player 2's proposal, then they form a coalition with the joint proposal $f^{\{1,2\}} \in \mathcal{F}^{\{1,2\}}$, where $f^{\{1,2\}}(S) = f^{\{2\}}(S)$ for all $S \supseteq \{1, 2\}$. If player 1 rejects it, then the two players stay apart for the moment holding their own proposals, respectively.

Obviously, up to now there will be two cases for players 1 and 2: either they form a coalition with an agreed proposal or they separate. It is worth noting here that even though they may not reach an agreement as in the latter case, neither player 1 nor player 2 will be excluded from future negotiations.

By the ordering, in the next step it will be for player 3 to take actions. There are three cases.

1. Coalition $\{1, 2\}$ was formed and player 3 accepts $f^{\{1,2\}}$. Then, the three players form a coalition with the joint proposal to be defined as $f^{\{1,2,3\}}(S) = f^{\{1,2\}}(S)$ for

all $S \supseteq \{1, 2, 3\}$.

2. Coalition $\{1, 2\}$ was formed, but player 3 does not accept $f^{\{1,2\}}$ and proposes $f^{\{3\}} \in \mathcal{F}^{\{3\}}$ instead. Then, coalition $\{1, 2\}$ will decide to accept or reject player 3's proposal. Players 1 and 2 will vote sequentially.

   (a) If *both* players in $\{1, 2\}$ accept $f^{\{3\}}$, then these three players form a coalition with the joint proposal defined as $f^{\{1,2,3\}}(S) = f^{\{3\}}(S)$ for all $S \supseteq \{1, 2, 3\}$.

   (b) Otherwise, players 1 and 2 keep their coalition $\{1, 2\}$ with their proposal while player 3 stands alone for now keeping her own proposal.

3. Players 1 and 2 did not form a coalition but set apart with their own proposals. Player 3 can decide to join either of them to form a coalition if she accepts the corresponding proposal. If player 3 agrees on neither $f^{\{1\}}$ nor $f^{\{2\}}$ but proposes $f^{\{3\}}$, then players 1 and 2 sequentially decide to accept or reject $f^{\{3\}}$.

   (a) If both reject it, then the three players set apart from each other while holding their own proposals.

   (b) If both accept it, then they form the coalition $\{1, 2, 3\}$ with the joint proposal defined as $f^{\{1,2,3\}}(S) = f^{3}(S)$ for all $S \supseteq \{1, 2, 3\}$.

   (c) If one player, e.g., player 1, accepts it while the other player rejects it, then coalition $\{1, 3\}$ is formed with the joint proposal defined as $f^{\{1,3\}}(S) = f^{\{3\}}(S)$ for all $S \supseteq \{1, 3\}$ while the other player remains alone with her own proposal.

Note that we can have a variant mechanism such that whenever an instance of acceptance happens, thereby a new coalition is formed, the rest of the players are allowed to vote on the corresponding proposal to possibly join the coalition. It will not change the subgame perfect equilibrium payoff allocations of the model, but it would make the mechanism unnecessarily cumbersome, while adding neither any substantially new feature nor real insight. Hence, we will not further analyze this variant here.

In this way, the negotiation proceeds to player n to make a proposal, with possibilities of finally forming the grand coalition or partial coalitions (including the possible case of all singleton coalitions). If the grand coalition is formed, then the game ends while the corresponding payoff sharing schemes are worked out and implemented accordingly. Otherwise, if the grand coalition is not formed, with some probability they repeat the process from the very beginning with a (new) random order of all the players, and with the remaining probability the game stops as it is with players to receive payoffs according to the corresponding proposals.

Note that the mechanism would allow for all the players to express their opinions before possibly starting playing the game again.

Below we formally describe the decentralized multilateral negotiation mechanism associated to V and a fixed parameter $\rho \in [0,1)$. All players in N are assumed to be rational, completely informed, and expected utility maximizers. The mechanism proceeds in the following steps.

Step 0: All players in N form a randomized order, with all orderings to be formed equally likely. For convenience of expression, we assume the order is, without loss of generality, $\pi = (1, \ldots, n)$. Go to the next step.

Step 1: Only player 1 is active in this step while all other players take no action but simply wait. Player 1 announces a rule $f^{\{1\}} \in \mathcal{F}^{\{1\}}$ and the *present situation* is a pair (R^1, F^1) with $R^1 = \{\{1\}\}$ and $F^1 = \left\{f^{\{1\}}\right\}$. It proceeds to the next step.

Step s $(s = 2, \ldots, n)$: By induction we know the present situation of the set of players P_s^π and denote it as (R^{s-1}, F^{s-1}), where R^{s-1} is a partition of P_s^π whose elements are nonempty sets of P_s^π and F^{s-1} is a set of rules supported by the coalitions of R^{s-1}, respectively. Let $R^{s-1} = \left\{R_1^{s-1}, \ldots, R_K^{s-1}\right\}$ and $F^{s-1} = \left\{f^{R_1^{s-1}}, \ldots, f^{R_K^{s-1}}\right\}$ with $K \geq 1$. There are two cases in this step.

Case 1. Player s *accepts* a particular $f^{R_k^{s-1}}$ and joins R_k^{s-1} to form coalition $R_k^{s-1} \cup \{s\}$ with its supported rule defined as $f^{R_k^{s-1} \cup \{s\}}(S) = f^{R_k^{s-1}}(S)$ for all $S \supseteq R_k^{s-1} \cup \{s\}$. It then leads to the present situation (R^s, F^s), where $R^s = \{R_1^s, \ldots, R_K^s\}$ with $R_k^s = R_k^{s-1} \cup \{s\}$ and $R_l^s = R_l^{s-1}$ for all $l \in \{1, ..., K\} \setminus \{k\}$, and $F^s = \left\{f^{R_1^s}, \ldots, f^{R_K^s}\right\}$ with $f^{R_k^s} = f^{R_k^{s-1} \cup \{s\}}$ and $f^{R_l^s} = f^{R_l^{s-1}}$ for all $l \in \{1, ..., K\} \setminus \{k\}$. Go to the next step.

Case 2. Player s does not accept any rule in F^{s-1} but proposes a rule $f^{\{s\}} \in \mathcal{F}^{\{s\}}$. All players in P_s^π sequentially vote for or against $f^{\{s\}}$. Let $A \subseteq P_s^\pi$ be the set of players who vote in favour, and let $B = \bigcup_{R_k^{s-1} \subseteq A} R_k^{s-1}$ be the set of players who belong to a coalition in which *all* its members vote in favour. Then, all players of B join player s and form coalition $B \cup \{s\}$, which leads to the present situation (R^s, F^s), where $R^s = \{R_1^s, \ldots, R_{K'}^s\}$ with $R_{K'}^s = B \cup \{s\}$ and $\left\{R_1^s, \ldots, R_{K'-1}^s\right\} = \{P \in R^{s-1} : P \nsubseteq B\}$, and $F^s = \left\{f^{R_1^s}, \ldots, f^{R_{K'}^s}\right\}$ with $f^{R_{K'}^s}$ defined as $f^{R_{K'}^s}(S) = f^{\{s\}}(S)$ for all $S \supseteq R_{K'}^s \cup \{s\}$, and $f^{R_l^s} = f^{R_{l'}^{s-1}}$ when $R_l^s = R_{l'}^{s-1}$. Go to the next step.

Step $n+1$: It has two cases.

Case 1. $R^n = \{N\}$. That is, the grand coalition N is formed and F^n contains only one rule. The game ends with (R^n, F^n) being implemented, i.e., each player $i \in N$ receives $f^N(N)_i$.

Case 2. $R^n \neq \{N\}$. Then, two things may happen:

(a) With probability ρ, the game goes back to Step 0 and the whole process is repeated from the beginning.

(b) With probability $1-\rho$, the game stops with (R^n, F^n) being implemented, i.e., for each $R_k^n \in R^n$, each player $i \in R_k^n$ receives $f^{R_k^n}(R_k^n)_i$.

Proceeding in the above protocol, finally players in N will reach[2] a situation (R^n, F^n), where $R^n = \{R_1^n, \ldots, R_K^n\}$ and $F^n = \left\{f^{R_1^n}, \ldots, f^{R_K^n}\right\}$. The final payoff allocation will be $f^{R_k^n}\left(R_k^n\right)_i$ for all $i \in R_k^n \in R^n$.

We work with stationary strategies. A strategy is stationary if it only depends on the elements present on the negotiation table, and not on the history that yields these elements. Note that a stationary subgame perfect equilibrium is also optimal against deviation strategies that are non-stationary. From now on, when we say subgame perfect equilibrium, we mean stationary subgame perfect equilibrium.

We also assume that, in case of indifference between voting in favour or against a proposal, players vote in favour and, in case of indifference between joining a coalition or proposing a new rule, players join the coalition. These are standard assumptions in the literature. For example, Moldovanu and Winter (1994) "assume that each player prefers to be a member of *large* coalitions than smaller ones provided that he earns the same payoff in the two agreements", and Hart and Mas-Colell (1996) "assume that both proposers and respondents break ties in favor of quick termination of the game."

It is useful to describe a game for two players, say 1 and 2. In this case, we are in a pure bargaining problem.

Example 3.1 *Assume $N = \{1,2\}$ and V is the unanimity 2-player game given by $V(\{1\}) = V(\{2\}) = (-\infty, 0]$ and $V(N) = \{x \in \mathbb{R}^N : x_1 + x_2 \leq 1\}$. Assume player 1 is first chosen. Now, player 1 announces some $f^{\{1\}} \in \mathcal{F}^{\{1\}}$. Say, for example, $f^{\{1\}}(\{1\}) = 0$ and $f^{\{1\}}(N) = \left(\frac{1}{2}, \frac{1}{2}\right)$. Player 2 observes $f^{\{1\}}$ and makes her decision.*

- *If player 2 agrees to join player, the game ends and the final payoff allocation is $\left(\frac{1}{2}, \frac{1}{2}\right)$.*

[2]Since $\rho < 1$, any coalition of players will agree on a proposal with probability 1.

- *If player 2 disagrees, she proposes some $f^{\{2\}} \in \mathcal{F}^{\{2\}}$. Say, for example, $f^{\{2\}}(\{2\}) = 0$ and $f^{\{2\}}(N) = \left(\frac{\rho}{1+\rho}, \frac{1}{1+\rho}\right)$. Then, player 1 should choose whether to accept or reject this proposal.*
  - *If player 1 accepts, the final payoff allocation is $\left(\frac{\rho}{1+\rho}, \frac{1}{1+\rho}\right)$.*
  - *If player 1 rejects, with probability ρ the whole process is repeated, and with probability $1-\rho$ the final payoff allocation is $f^{\{1\}}(\{1\}) \times f^{\{2\}}(\{2\}) = (0,0)$.*

In Example 3.1, the second case (i.e., disagreeing) is an optimal response for player 2, and her proposal will be accepted by player 1. Player 1 can propose some $f^{\{1\}}$ with $f^{\{1\}}(N)_2 \geq \frac{1}{1+\rho}$ in order to be accepted. Hence, there is an equilibrium path in which player 1 proposes an acceptable rule. In any case, the final payoff allocation in equilibrium is $\left(\frac{\rho}{1+\rho}, \frac{1}{1+\rho}\right)$.

Notice that player 1's proposal is innocuous from a strategic point of view. Thus, the game for two players essentially coincides with the random-version of Rubinstein's bargaining model first studied by Hart and Mas-Colell (1996), and thus yielding the Nash bargaining solution as ρ approaches 1.

This analysis can be extended for more than two players when the game is a pure bargaining problem.

Thus, we have (cf. Theorem 3 in Hart and Mas-Colell (1996)) the following result:

Theorem 3.1 *For a pure bargaining problem, for each ρ there is at least one equilibrium. Moreover, any equilibrium payoff allocation converges to the Nash bargaining solution as ρ approaches 1.*

For more than two players, the partial proposals are not innocuous, as they allow a reassignment of utilities among those that leave the player indifferent. See Example 5 in Vidal-Puga (2008).

Our main result, which comprises Proposition 5.1, Proposition 5.2, Proposition 5.3, Proposition 5.4, and Proposition 5.5 in Section 5, is the following:

Theorem 3.2 *For each ρ there is at least one subgame perfect equilibrium, and any equilibrium payoff allocation converges to a Shapley NTU value as ρ approaches 1. Moreover, any equilibrium payoff allocation coincides with the Shapley NTU value in expected terms when $V(N)$ is delimited by a hyperplane.*

4 Detailed example with three players

In this section, we focus on an example which belongs to a family first described by Roth (1980) and also commented by Myerson (1980); Harsanyi (1980); Hart and Kurz (1983); Hart and Mas-Colell (1996); Aumann (1985).

Consider a Parliament with 101 seats and three political parties, 1, 2, and 3. Party 1 has 49 votes. Parties 2 and 3 have 26 votes each. A Government can only be established with a majority (51) of votes. Moreover, the Parliament regulation says that, in case of a coalition of parties forming the government, the number of ministers should be shared between those parties proportionally to their respective number of seats. Assuming that the parties are only interested in the percentage of ministers they have, the resulting NTU game is the following:

$$
\begin{aligned}
V(\{i\}) &= \{(0)\} - \mathbb{R}_+^{\{i\}} \text{ for all } i \in \{1,2,3\} \\
V(\{1,i\}) &= \left\{\left(\frac{49}{75}, \frac{26}{75}\right)\right\} - \mathbb{R}_+^{\{1,i\}} \text{ for all } i \in \{2,3\} \\
V(\{2,3\}) &= \left\{\left(\frac{1}{2}, \frac{1}{2}\right)\right\} - \mathbb{R}_+^{\{2,3\}} \\
V(\{1,2,3\}) &= \left\langle \left\{ \left(\frac{49}{75}, \frac{26}{75}, 0\right), \left(\frac{49}{75}, 0, \frac{26}{75}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right), \left(\frac{49}{101}, \frac{26}{101}, \frac{26}{101}\right) \right\} \right\rangle - \mathbb{R}_+^{\{1,2,3\}}.
\end{aligned}
$$

where $\langle X \rangle$ represents the convex hull of the elements in set X. Let $N = \{1,2,3\}$. Any element in $V(N)$ can be achieved by the three parties by agreeing on dividing the political term into several governments with different coalitions. Moreover, since $\left(\frac{49}{101}, \frac{26}{101}, \frac{26}{101}\right)$ belongs to the convex hull of the other three possible payoff allocations, $V(N)$ can also be written as

$$
V(N) = \left\langle \left\{ \left(\frac{49}{75}, \frac{26}{75}, 0\right), \left(\frac{49}{75}, 0, \frac{26}{75}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right) \right\} \right\rangle - \mathbb{R}_+^{N}.
$$

In any case, this game does not satisfy one of the conditions we require to NTU games. In particular, it does not satisfy smoothness and non-levelness (in each point on the surface there exists a unique orthonormal vector with all its coordinates being away above from zero). This property will be irrelevant in this example. In any case, we can make the game smooth and non-level by assuming that a party in the government can propose as ministers candidates from a continuum of independent people with increasing autonomous vote. Independent candidates imply a smooth decrease in the party's utility. The idea is that these independent candidates would drastically decrease the utility in one of the parties in exchange of a small (but positive) increase in the utility of the other party (see Figure 1).

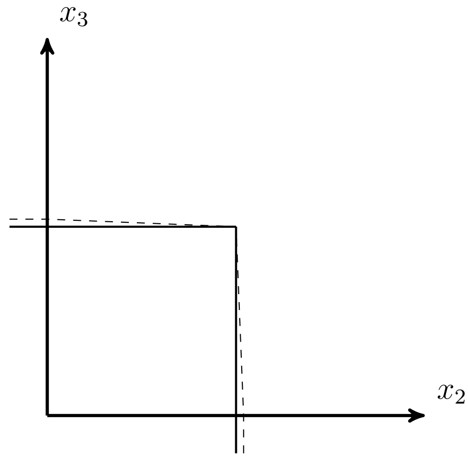

Figure 1: Regular lines represent $V(\{2,3\})$. Dashed lines represent the effect of independent candidates.

Roth (1980) argues that the only reasonable solution for this game is $(0, 0.5, 0.5)$. This payoff allocation can be obtained by applying the Harsanyi (1963) value, which yields two possible payoff allocations: $(0.23, 0.38, 0.38)$ and $(0, 0.5, 0.5)$. Aumann (1985) justifies the Shapley NTU value of this game, which in this case is unique: $(0.33, 0.33, 0.33)$. Alternatively, by applying the Hart and Mas-Colell (1996) model of negotiation, one can obtain the consistent value (Maschler and Owen, 1989, 1992), which in this case is also unique: $(0.44, 0.28, 0.28)$.

We check how our non-cooperative game yields the Shapley NTU value $(0.33, 0.33, 0.33)$ in expected terms.

The proposal $f^{\{i\}} \in \mathcal{F}^{\{i\}}$ of each party $i \in N$, in any stage in which they should propose a rule, may satisfy (in equilibrium) $f^{\{i\}}(\{i\}) = (0)$. Hence, we assume it throughout the whole section. However, the other $f^{\{i\}}(S)$ would depend on the stage of the game.

Assume first that $x^\rho \in V(N)$ is the expected final payoff allocation when parties play a (stationary) subgame perfect equilibrium. Let x^π be the expected final payoff allocation when order π is chosen by Nature at stage 0.

In order to compute $x^{(1,2,3)}$, we should take into account what would happen if players 1 and 2 do not agree and propose $\left(R^{\{1,2\}}, F^{\{1,2\}}\right)$ with $R^{\{1,2\}} = \{\{1\},\{2\}\}$. Then, an optimal strategy for party 3 is to either choose a party that maximizes her final payoff (by joining either of them to form a coalition) or to propose $f^{\{3\}}$ with

$$f^{\{3\}}(N) = (\rho x_1^\rho, \rho x_2^\rho, 1 - \rho\,(x_1^\rho + x_2^\rho)).$$

Hence, party 3 final payoff will be

$$\max\left\{\rho x_3^\rho + (1-\rho) f_3^{\{1\}}(\{1,3\}), \rho x_3^\rho + (1-\rho) f_3^{\{2\}}(\{2,3\}), 1 - \rho\,(x_1^\rho + x_2^\rho)\right\}.$$

In case parties 1 and 2 agree and propose $\left(R^{\{1,2\}}, F^{\{1,2\}}\right)$ with $R^{\{1,2\}} = \{\{1,2\}\}$, then an optimal strategy for party 3 is to choose the more profitable of these options:

- to reject this proposal, propose an acceptable $f^{\{3\}}$ with

$$\begin{aligned}
f^{\{3\}}(N)_1 &= \rho x_1^\rho + (1-\rho) f^{\{1,2\}}(\{1,2\})_1 \\
f^{\{3\}}(N)_2 &= \rho x_2^\rho + (1-\rho) f^{\{1,2\}}(\{1,2\})_2 \\
f^{\{3\}}(N)_3 &= 1 - \rho(x_1^\rho + x_2^\rho) - (1-\rho)\left(f^{\{1,2\}}(\{1,2\})_1 + f^{\{1,2\}}(\{1,2\})_2\right)
\end{aligned}$$

  and obtain $\alpha\left(f^{\{1,2\}}\right)_3 := 1 - \rho(x_2^\rho + x_3^\rho) - (1-\rho)(f^{\{1,2\}}(\{1,2\})_1 + f^{\{1,2\}}(\{1,2\})_2)$.

  We will see that the resulting x^ρ is such that this $f^{\{3\}}(N)$ is indeed contained into $V(N)$ for any $\rho \in [0,1)$.

In any case, party 3 can assure itself $\alpha\left(f^{\{1,2\}}\right)_3$, which is bounded below by ρx_3^ρ:

$$\begin{aligned}
\alpha\left(f^{\{1,2\}}\right)_3 &= 1 - \rho(x_1^\rho + x_2^\rho) - (1-\rho)\left(f^{\{1,2\}}(\{1,2\})_1 + f^{\{1,2\}}(\{1,2\})_2\right) \\
&\geq 1 - \rho(x_1^\rho + x\rho_2) - (1-\rho) \sup_{f \in \mathcal{F}^{\{1,2\}}} \left(f(\{1,2\})_1 + f(\{1,2\})_2\right) \\
&= 1 - \rho(x_1^\rho + x_2^\rho) - (1-\rho)\left(\frac{49}{75} + \frac{26}{75}\right) \\
&= 1 - \rho(x_1^\rho + x_2^\rho) - (1-\rho) = \rho(1 - x_1^\rho - x_2^\rho) \geq \rho x_3^\rho.
\end{aligned}$$

Party 2 can anticipate party 1's response and choose its optimal strategy. The optimal strategy for party 2 is to choose the most profitable of the following options:

- to accept party 1's proposal (by joining coalition $\{1\}$) and obtain either $f^{\{1\}}(N)_2$ or $\rho x_2^\rho + (1-\rho) f^{\{1\}}(\{1,2\})_2$, depending on whether $f^{\{1\}}(N)_3$ is greater or smaller than $\alpha\left(f^{\{1\}}\right)_3$, or

- to reject this proposal, propose an acceptable $f^{\{2\}}$ with

$$\begin{aligned}
f^{\{2\}}(\{1,2\}) &= \left(\frac{49}{75}, \frac{26}{75}\right) \\
f^{\{2\}}(N) &= (\rho x_1^\rho, \rho x_2^\rho + 1 - \rho, \rho(1 - x_1^\rho - x_2^\rho))
\end{aligned}$$

  and obtain $\alpha(f^{\{2\}})_2 := f^{\{2\}}(N)_2 = \rho x_2^\rho + 1 - \rho$.

For party 2, the option to reject party 1's proposal and propose an $f^{\{2\}}$ unacceptable by party 1 and acceptable by party 3 is out of the table. To see why, notice that party 2 should present $f^{\{2\}}$ with $f^{\{2\}}(\{2\}) = (0)$ and $f^{\{2\}}(\{2,3\}) = \left(\frac{1}{2}, \frac{1}{2}\right)$, but since $1-\rho(x_1^\rho + x_2^\rho)$

is always greater than $\rho x_3^\rho + (1-\rho)\frac{1}{2}$ (due to the fact that $x_1^\rho + x_2^\rho + x_3^\rho \geq 1$ and $\rho \in [0,1)$), party 3 would not accept to join party 2.

In any case, party 2 can assure itself at least ρx_2^ρ.

Now, party 1 can anticipate party 2 and party 3's responses and choose its optimal strategy. An optimal strategy for party 1, for ρ close enough[3] to 1, is to propose an acceptable $f^{\{1\}}$ with

$$\begin{aligned} f^{\{1\}}(\{1,2\}) &= \left(\frac{49}{75}, \frac{26}{75}\right) \\ f^{\{1\}}(N) &= \left(\rho x_1^\rho, \rho x_2^\rho + 1 - \rho, \rho(1 - x_1^\rho - x_2^\rho)\right). \end{aligned}$$

Knowing this, we conclude that, for ρ close enough to 1,

$$x^{(1,2,3)} = \left(\rho x_1^\rho, \rho x_2^\rho + 1 - \rho, \rho(1 - x_1^\rho - x_2^\rho)\right).$$

Analogously,

$$\begin{aligned} x^{(1,3,2)} &= (\rho x_1^\rho, \rho(1 - x_1^\rho - x_3^\rho), \rho x_3^\rho + 1 - \rho) \\ x^{(2,1,3)} &= (\rho x_1^\rho + 1 - \rho, \rho x_2^\rho, \rho(1 - x_1^\rho - x_2^\rho)) \end{aligned}$$

an so on. In general,

$$\begin{aligned} x_{\pi_1}^\pi &= \rho x_{\pi_1}^\rho \\ x_{\pi_2}^\pi &= \rho x_{\pi_2}^\rho + 1 - \rho \\ x_{\pi_3}^\pi &= \rho(1 - x_{\pi_1}^\rho - x_{\pi_2}^\rho). \end{aligned}$$

Since x^ρ is the average of these x^π, and all of them satisfy $x_1^\pi + x_2^\pi + x_3^\pi = 1$, we conclude $x_1^\rho + x_2^\rho + x_3^\rho = 1$. By symmetry,

$$x^\rho = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

and hence

$$\begin{aligned} x_{\pi_1}^\pi &= \frac{\rho}{3} \\ x_{\pi_2}^\pi &= 1 - \frac{2\rho}{3} \\ x_{\pi_3}^\pi &= \frac{\rho}{3}. \end{aligned}$$

[3] In particular, $\rho \in \left[\frac{147}{173}, 1\right) \approx [0.8497, 1)$.

Example 4.1 *Assume Nature chooses order* $(1, 2, 3)$*. In equilibrium, party* 1 *may propose* $f^{\{1\}}$ *with*

$$\begin{aligned} f^{\{1\}}(\{1\}) &= (0) \\ f^{\{1\}}(\{1,2\}) &= \left(\frac{49}{75}, \frac{26}{75}\right) \\ f^{\{1\}}(\{1,3\}) &= \left(\frac{49}{75}, \frac{26}{75}\right) \\ f^{\{1\}}(N) &= \left(\frac{\rho}{3}, 1 - \frac{2\rho}{3}, \frac{\rho}{3}\right), \end{aligned}$$

which is equivalent to stating that

> *I (party* 1*) propose to form government all the term with either party* 2 *or party* 3 *alone, but I would also agree to divide the term so that the government is held by the three-party coalition government for* X *part of the term, by parties* 1 *and* 3 *for* Y *part of term, by parties* 2 *and* 3 *for* Z *part of term, and by parties* 1 *and* 2 *for the remaining part of term.*

The values of X*,* Y *and* Z *depend on* ρ*. For example, for* $\rho = 0.9$*, it could be:*

> *I (party* 1*) propose to form government all the term with either party* 2 *or party* 3 *alone, but I would also agree to divide the term so that the government is held by parties* 1 *and* 3 *for* 9% *of term, by parties* 2 *and* 3 *for* 54% *of term, and by parties* 1 *and* 2 *for the remaining* 37% *of term.*

Now, party 2 *agrees and party* 1 *also agrees. The final payoff allocation is then*

$$\left(\frac{\rho}{3}, 1 - \frac{2\rho}{3}, \frac{\rho}{3}\right)$$

which converges to $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ *as* ρ *approaches* 1*.*

5 Main results

Given $\pi \in \Pi$ and $x \in V(N)$, we define $\psi^{\rho,\pi}(x) \in V(N)$ inductively as follows:

$$\psi^{\rho,\pi}(x)_{\pi_n} = \min_{y \in V(N \setminus \{\pi_n\})} \left\{ \tau_{\pi_n} \left(\rho x_{N \setminus \{\pi_n\}} + (1-\rho) y \right) \right\}.$$

Assume $\psi^{\rho,\pi}(x)_{\pi_l} \in \mathbb{R}$ is defined for each $l > k$. Equivalently, $\psi^{\rho,\pi}(x)_{N \setminus \overline{P^{\pi}_{\pi_k}}} \in \mathbb{R}^{N \setminus \overline{P^{\pi}_{\pi_k}}}$ is defined. We then define

$$\psi^{\rho,\pi}(x)_{\pi_k} = \min_{y \in V\left(P^{\pi}_{\pi_k}\right)} \left\{ \tau_{\pi_k} \left(\left(\rho x_{P^{\pi}_{\pi_k}} + (1-\rho) y \right) \times \psi^{\rho,\pi}(x)_{N \setminus \overline{P^{\pi}_{\pi_k}}} \right) \right\}.$$

Notice that $\psi^{\rho,\pi}(x)$ does not depend on x_{π_n}.

Proposition 5.1 *For each $\rho \in [0, 1)$, the proposals corresponding to a stationary subgame perfect equilibrium are always accepted, and the final payoff allocations in the subgame where the order is given by $\pi \in \Pi$ are characterized by $\psi^{\rho,\pi}(x^\rho)$ for all $\pi \in \Pi$, where*

$$x^\rho = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \psi^{\rho,\pi}(x^\rho).$$

Moreover, these proposals are always nonnegative, i.e, $\psi^{\rho,\pi}(x) \in \mathbb{R}^N_+$ for all $\pi \in \Pi$.

Proof. Assume we are in a subgame perfect equilibrium. For each $\pi \in \Pi$, let x^π be the final payoff allocation in the subgame that begins when order π is chosen. Let x^* be the average of these x^π over Π. Assume, w.l.o.g., order $\pi = (1, \dots, n)$ is chosen. We prove the following Claim:

Claim 5.1 *Each player $i \in N$ can assure herself at least $\psi^{\rho,\pi}(x^*)_i$, i.e., $x^\pi_i \geq \psi^{\rho,\pi}(x^*)_i$. Moreover, this assured payoff can be done by making a unanimously acceptable offer (i.e. all players in P^{π_i} vote in favour in equilibrium).*

In order to prove this claim, we proceed by backwards induction. Assume we are at step n and player n faces proposal (R^{n-1}, F^{n-1}) with $R^{n-1} = (R^{n-1}_1, \dots, R^{n-1}_K)$ and $F^{n-1} = \left(f^{R^{n-1}_1}, \dots, f^{R^{n-1}_K}\right)$. For notational simplicity, from now on we write f^k instead of $f^{R^{n-1}_k}$. Let $y^n \in \mathbb{R}^{P^\pi_n}$ be defined as

$$y^n = f^1(R^{n-1}_1) \times \cdots \times f^K\left(R^{n-1}_K\right).$$

By super-additivity of V, $y^n \in V(P^\pi_n)$. Thus,

$$\psi^{\rho,\pi}(x^*)_n \leq \tau_n\left(\rho x^*_{P^\pi_n} + (1-\rho)y^n\right).$$

Then, player n can assure herself at least $\psi^{\rho,\pi}(x^*)_n$ by not accepting any rule in F^{n-1} and proposing $f^{\{n\}}$ with

$$f^{\{n\}}(N) = \left(\rho x^*_{P^\pi_n} + (1-\rho)y^n\right) \times \tau_n\left(\rho x^*_{P^\pi_n} + (1-\rho)y^n\right).$$

Then, in case there exist incompatible rules between the players at the end of the process (Case 2 in Step $n+1$), the final expected payoff for each player $i \in R^{n-1}_k$ is $\rho x^*_i + (1 - \rho) f^k\left(R^{n-1}_k\right)$. Under our assumption of agreement in case of indifference, in equilibrium all players in P^π_n will vote in favour so that the final payoff allocation is $f^{\{n\}}(N)$. Moreover, by definition of τ_n, this is an optimal strategy for player n when making a unanimously acceptable offer.

Assume now the result holds for each player $i > s$, and we are at step s. Player s faces proposal (R^{s-1}, F^{s-1}) with $R^{s-1} = (R_1^{s-1}, \ldots, R_K^{s-1})$ and $F^{s-1} = \left(f^{R_1^{s-1}}, \ldots, f^{R_K^{s-1}}\right)$. Again, for notational simplicity, from now on we write f^k instead of $f^{R_k^{s-1}}$. Let y^s be defined as

$$y^s = f^1(R_1^{s-1}) \times \cdots \times f^K\left(R_K^{s-1}\right).$$

By super-additivity of V, $y^s \in V\left(P_s^\pi\right)$. Thus,

$$\psi^{\rho,\pi}(x^*)_s \leq \tau_s\left(\left(\rho x^*_{P_s^\pi} + (1-\rho)y^s\right) \times \psi^{\rho,\pi}(x^*)_{N\setminus\overline{P_s^\pi}}\right).$$

Then, player s can assure herself at least $\psi^{\rho,\pi}(x^*)_s$ by not accepting any rule in F^{s-1} and proposing $f^{\{s\}}$ with

$$f^{\{s\}}\left(P_i^\pi\right) \in \arg\min\left\{\tau_i\left(\left(\rho x^*_{P_i^\pi} + (1-\rho)y\right) \times \psi^{\rho,\pi}(x^*)_{N\setminus\overline{P_s^\pi}}\right) : y \in V\left(P_i^\pi\right)\right\}$$

for all $i > s$ and

$$f^{\{s\}}(N) = \left(\rho x^*_{P_s^\pi} + (1-\rho)y^s\right) \times \psi^{\rho,\pi}(x^*)_{N\setminus\overline{P_s^\pi}} \times \tau_s\left(\left(\rho x^*_{P_s^\pi} + (1-\rho)y^s\right) \times \psi^{\rho,\pi}(x^*)_{N\setminus\overline{P_s^\pi}}\right).$$

Under our assumption of agreement in case of indifference, in equilibrium all players in P_s^π will vote in favour and all players in $N \setminus \overline{P_s^\pi}$ will accept the rule in the growing coalition so that the final payoff allocation is $f^{\{s\}}(N)$. Moreover, by definition of τ_s, this is an optimal strategy for player s when making a unanimity acceptable offer. This conclude the proof of Claim 5.1.

Under Claim 5.1, $x^\pi \geq \psi^{\rho,\pi}(x^*)$. By definition, $\psi^{\rho,\pi}(x^*)$ belongs to the Pareto frontier of $V(N)$, and hence $x^\pi = \psi^{\rho,\pi}(x^*)$. This implies $x^\rho = x^*$ and hence the first part of the result. Moreover, since we are in equilibrium, these proposals are nonnegative because each player $i \in N$ has the strategy of voting always against any proposal and propose $f^{\{i\}}$ with $f^{\{i\}}(S) = 0_S$ for all $S \supseteq \{i\}$, which assures her a zero payoff. ■

Proposition 5.2 *For each $\rho \in [0,1)$, $\pi \in \Pi$, and $x \in V(N) \cap \mathbb{R}^N_+$, $\psi^{\rho,\pi}(x)$ is characterized by*

1. $\psi^{\rho,\pi}(x) \in \partial V(N)$
2. $\lambda_i \psi^{\rho,\pi}(x)_i = \rho\lambda_i x_i + (1-\rho)\left(v^\lambda\left(\overline{P_i^\pi}\right) - v^\lambda\left(P_i^\pi\right)\right)$ *for all $i \in N \setminus \{\pi_n\}$*

for some $\lambda \in \Lambda^N_{\mu(V)}$.

Proof. Fix $\pi \in \Pi$, $\rho \in [0,1)$ and $x \in V(N)$. We assume, w.l.o.g., $\pi = (1,2,\ldots,n)$. For each $i \in N$, take

$$y^i \in \arg\min\left\{\tau_i\left(\left(\rho x_{P_i^\pi} + (1-\rho)y\right) \times \psi^{\rho,\pi}(x)_{N\setminus\overline{P_i^\pi}}\right) : y \in V\left(P_i^\pi\right)\right\}$$

and

$$z^i = y^i \times \psi^{\rho,\pi}(x)_{N\setminus\overline{P_i^\pi}} \times \tau_i\left(y^i \times \psi^{\rho,\pi}(x)_{N\setminus\overline{P_i^\pi}}\right) \in \partial V(N).$$

In particular, $z^1 = \psi^{\rho,\pi}(x)$. Since all points in $(z^i)_{i\in N}$ belong to the Pareto frontier of $V(N)$, there exists some $\lambda \in \Lambda^N_{\mu(V)}$ such that all points $(z^i)_{i\in N}$ belong to the hyperplane

$$\left\{y \in \mathbb{R}^N : \sum_{i\in N}\lambda_i y_i = \sum_{i\in N}\lambda_i z_i^1\right\}.$$

Then, $\psi^{\rho,\pi}(x)$ coincides in both V and the NTU game W defined as

$$W(N) = \left\{y \in \mathbb{R}^N : \sum_{i\in N}\lambda_i y_i \le \sum_{i\in N}\lambda_i z_i^1\right\}$$

and $W(S) = V(S)$ otherwise. Hence,

$$\begin{aligned}
\lambda_i\psi^{\rho,\pi}(x)_i &= \lambda_i \min\left\{\tau_i\left(\left(\rho x_{P_i^\pi} + (1-\rho)y\right) \times \psi^{\rho,\pi}(x)_{N\setminus\overline{P_i^\pi}}\right) : y \in V\left(P_i^\pi\right)\right\} \\
&= \min\left\{\lambda_i\tau_i\left(\left(\rho x_{P_i^\pi} + (1-\rho)y\right) \times \psi^{\rho,\pi}(x)_{N\setminus\overline{P_i^\pi}}\right) : y \in V\left(P_i^\pi\right)\right\} \\
&= \min\left\{\sum_{i\in N}\lambda_i z_i^1 - \rho\sum_{j<i}\lambda_j x_j - (1-\rho)\sum_{j<i}\lambda_j y_j - \sum_{j>i}\lambda_j\psi^{\rho,\pi}(x)_j : y \in V\left(P_i^\pi\right)\right\} \\
&= \sum_{i\in N}\lambda_i z_i^1 - \rho\sum_{j<i}\lambda_j x_j - (1-\rho)v^\lambda\left(P_i^\pi\right) - \sum_{j>i}\lambda_j\psi^{\rho,\pi}(x)_j \\
&= \sum_{j\le i}\lambda_j\psi^{\rho,\pi}(x)_j - \rho\sum_{j<i}\lambda_j x_j - (1-\rho)v^\lambda\left(P_i^\pi\right)
\end{aligned}$$

for all $i \in N$. Rearranging terms:

$$\sum_{j<i}\lambda_j\psi^{\rho,\pi}(x)_j = \rho\sum_{j<i}\lambda_j x_j + (1-\rho)v^\lambda\left(P_i^\pi\right)$$

for all $i \in N$. Hence, given $i \in N\setminus\{n\}$,

$$\sum_{j<i+1}\lambda_j\psi^{\rho,\pi}(x)_j = \rho\sum_{j<i+1}\lambda_j x_j + (1-\rho)v^\lambda\left(P_{i+1}^\pi\right)$$

and thus

$$\begin{aligned}
\lambda_i\psi^{\rho,\pi}(x)_i &= \rho\lambda_i x_i + (1-\rho)\left(v^\lambda\left(P_{i+1}^\pi\right) - v^\lambda\left(P_i^\pi\right)\right) \\
&= \rho\lambda_i x_i + (1-\rho)\left(v^\lambda\left(\overline{P_i^\pi}\right) - v^\lambda\left(P_i^\pi\right)\right).
\end{aligned}$$

■

Corollary 5.1 *There exists $M \in \mathbb{R}_+$ such that $\left|\psi^{\rho,\pi}(x^\rho)_i - x_i^\rho\right| \leq (1-\rho)M$ for all $\rho \in [0,1)$, $\pi \in \Pi$ and $i \in N$.*

Proof. From Proposition 5.1 we deduce that $x^\rho \in V(N) \cap \mathbb{R}_+^N$. Fix $\rho \in [0,1)$, $\pi \in \Pi$, and $i \in N$. From Proposition 5.2, we deduce that there exists some $\lambda^\rho \in \Lambda_{\mu(V)}^N$ such that

$$\psi^{\rho,\pi}(x^\rho)_i - x_i^\rho = (1-\rho)\left(x_i^\rho - \frac{1}{\lambda_i^\rho}\left(v^{\lambda^\rho}\left(\overline{P_i^\pi}\right) - v^{\lambda^\rho}\left(P_i^\pi\right)\right)\right) \tag{2}$$

for all $i \in N\backslash\{\pi_n\}$. As for π_n, let W be the NTU game defined in the proof of Proposition 5.2. Then, $w^{\lambda^\rho}(N) = \sum_{i\in N} \lambda_i^\rho x_i^\rho$. Hence,

$$\begin{aligned}
\psi^{\rho,\pi}(x^\rho)_{\pi_n} - x_{\pi_n}^\rho &= \frac{1}{\lambda_{\pi_n}^\rho}\left(w^{\lambda^\rho}(N) - \rho\sum_{l<n}\lambda_{\pi_l}^\rho x_{\pi_l}^\rho - (1-\rho)v^{\lambda^\rho}\left(N\setminus\{n\}\right)\right) - x_{\pi_n}^\rho \\
&= \frac{1}{\lambda_{\pi_n}^\rho}\left(w^{\lambda^\rho}(N) - \rho\sum_{l<n}\lambda_{\pi_l}^\rho x_{\pi_l}^\rho - (1-\rho)v^{\lambda^\rho}\left(N\setminus\{n\}\right) - \lambda_{\pi_n}^\rho x_{\pi_n}^\rho\right) \\
&= \frac{1}{\lambda_{\pi_n}^\rho}\left(\sum_{l<n}\lambda_{\pi_l}^\rho x_{\pi_l}^\rho - \rho\sum_{l<n}\lambda_{\pi_l}^\rho x_{\pi_l}^\rho - (1-\rho)v^{\lambda^\rho}\left(N\setminus\{n\}\right)\right) \\
&= \frac{1-\rho}{\lambda_{\pi_n}^\rho}\left(\sum_{l<n}\lambda_{\pi_l}^\rho x_{\pi_l}^\rho - v^{\lambda^\rho}\left(N\setminus\{n\}\right)\right) \\
&= (1-\rho)\left(-x_{\pi_n}^\rho + \frac{1}{\lambda_1^\rho}\left(w^{\lambda^\rho}(N) - v^{\lambda^\rho}\left(N\setminus\{n\}\right)\right)\right).
\end{aligned} \tag{3}$$

From (2) and (3), and taking into account that $0 \leq w^{\lambda^\rho}(N) \leq v^{\lambda^\rho}(N)$, we deduce that the result holds for

$$M = \frac{1}{\mu(V)}\max\left\{v^\lambda(S) : S \subseteq N, \lambda \in \Lambda_{\mu(V)}^N\right\}.$$

■

As we will see, these $\psi^{\rho,\pi}(x^\rho)$ determine the payoff allocations in the subgames where π is the chosen order. Hence, Corollary 5.1 implies that for ρ close to 1 (i.e., low risk of breakdown), the final subgame perfect equilibrium payoff allocations in the subgames are very similar to the expected final payoff allocation, which is at least close to Pareto optimal (because the $\psi^{\rho,\pi}(x)$ are always Pareto optimal), and there is not substantial advantage or disadvantage in being at a certain position in the order; the random effect vanishes.

Proposition 5.3 *For each ρ there is at least one subgame perfect equilibrium.*

Proof. Under Proposition 5.1, any subgame perfect equilibrium payoff allocation is characterized by $x^\rho = \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\psi^{\rho,\pi}(x^\rho)$, where $\psi^{\rho,\pi}(x^\rho) \in \mathbb{R}^N_+$ is the payoff allocation in the subgame that begins when $\pi \in \Pi$ is the chosen order. To prove the existence of such x^ρ, consider the continuous function

$$\psi^\rho(x) = \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\psi^{\rho,\pi}(x)$$

for each $x \in V(N)\cap\mathbb{R}^N_+$. We will prove that ψ^ρ has a fixed-point. Under our conditions, $V(N)\cap\mathbb{R}^N_+$ is a non-empty, compact convex set. In order to apply Brouwer's fixed-point theorem, it is enough to prove that $\psi^{\rho,\pi}(x) \in V(N)\cap\mathbb{R}^N_+$ for all $x \in V(N)\cap\mathbb{R}^N_+$, so that convexity of $V(N)\cap\mathbb{R}^N_+$ assures that their average belongs to $V(N)\cap\mathbb{R}^N_+$ too. By definition, $\psi^{\rho,\pi}(x) \in V(N)$ for all $x \in \mathbb{R}^N$. Hence, we just need to prove that $\psi^{\rho,\pi}(x)_i \geq 0$ for all $i \in N$ and all $x \in V(N)\cap\mathbb{R}^N_+$. Fix $x \in V(N)\cap\mathbb{R}^N_+$ and fix $\sigma \in \Pi$. We assume w.l.o.g. $\sigma = (1, 2, \ldots, n)$. Let

$$y^n \in \arg\min_{y\in V(P^\sigma_n)}\left\{\tau_n\left(\rho x_{P^\sigma_n} + (1-\rho)y\right)\right\}$$

and let

$$z^n = \left(\rho x_{P^\sigma_n} + (1-\rho)y^n\right)\times\left(\tau_n\left(\rho x_{P^\sigma_n} + (1-\rho)y^n\right)\right).$$

By zero-monotonicity, $y^n \times (0) \in V(N)$. By convexity of $V(N)$, $\rho x + (1-\rho)\left(y^n \times (0)\right) \in V(N)$. Hence,

$$\psi^{\rho,\sigma}(x)_n = z^n_n = \tau_n\left(\rho x_{P^\sigma_n} + (1-\rho)\,y^n\right) \geq \left(\rho x + (1-\rho)(y^n\times(0))\right)_n = \rho x_n \geq 0.$$

By definition, z^n belongs to the Pareto frontier of $V(N)$. Assume we have defined z^j in the Pareto frontier of $V(N)$ for each $j > i$ and, moreover, $z^j_k = \psi^{\rho,\sigma}(x)_k$ for all $i < j \leq k$. In particular, let $z^*_k = z^j_k$ for all $i < j \leq k$, so that $z^*_{N\setminus\overline{P^\sigma_i}} \in \mathbb{R}^{N\setminus\overline{P^\sigma_i}}$ is well-defined. For $i > 1$, let

$$y^i \in \arg\min_{y\in V(P^\sigma_i)}\left\{\tau_i\left(\left(\rho x_{P^\sigma_i} + (1-\rho)y\right)\times\left(z^*_{N\setminus\overline{P^\sigma_i}}\right)\right)\right\}$$

and let

$$z^i = \left(\rho x_{P^\sigma_i} + (1-\rho)y^i\right)\times\left(\tau_i\left(\rho x_{P^\sigma_i} + (1-\rho)y^i\times\left(z^*_{N\setminus\overline{P^\sigma_i}}\right)\right)\right)\times\left(z^*_{N\setminus\overline{P^\sigma_i}}\right).$$

Finally, let $z^1 = x_{N\setminus\{n\}}\times(\tau_n(x_{N\setminus\{n\}}))$. We have then n vectors $(z^i)_{i\in N}$ on the Pareto frontier of $V(N)$. Hence, there exists $\lambda^{\rho,\sigma} \in \Lambda^N_{\mu(V)}$ and $\alpha \in \mathbb{R}_+$ such that $\sum_{j\in N}\lambda^{\rho,\sigma}_j z^i_j = \alpha$ for all $i \in N$. Given such $\lambda^{\rho,\sigma} \in \Lambda^N_{\mu(V)}$, consider the NTU game $(N, V^{\rho,\sigma})$ defined as

$$V^{\rho,\sigma}(N) = \left\{y \in \mathbb{R}^N : \sum_{i\in N}\lambda^{\rho,\sigma}_i y_i \leq \alpha\right\}$$

and $V^{\rho,\sigma}(S) = V(S)$ for all $S \subset N$. By definition, $z^1 \in V^{\rho,\sigma}(N)$ and $x \leq z^1$, so that comprehensiveness (P2) implies $x \in V^{\rho,\sigma}(N)$. Following the above procedure for x, ρ and σ, functions τ_i yield the same values in both V and $V^{\rho,\sigma}$ for all $i \in N$. Since $\psi^{\rho,\sigma}(x)$ does not depend on x_n, we have, for each $i < n$:

$$\lambda_i^{\rho,\sigma}\psi^{\rho,\sigma}(x)_i = \lambda_i^{\rho,\sigma}\psi^{\rho,\sigma}\left(x_{N\setminus\{n\}} \times \tau_n(x_{N\setminus\{n\}})\right)_i$$

under Proposition 5.2:

$$= \rho\lambda_i^{\rho,\sigma}x_i + (1-\rho)\left(v^{\lambda^{\rho,\sigma}}\left(\overline{P_i^\sigma}\right) - v^{\lambda^{\rho,\sigma}}\left(P_i^\sigma\right)\right).$$

Zero-monotonicity of V implies (for $i < n$) that $v^{\lambda^{\rho,\sigma}}\left(\overline{P_i^\sigma}\right) \geq v^{\lambda^{\rho,\sigma}}\left(P_i^\sigma\right)$. Hence, $\psi^{\rho,\sigma}(x)_i \geq 0$ for all $i \in N$. Therefore, there exists a fixed-point x^ρ, which completes the proof of existence of subgame perfect equilibria. ■

Proposition 5.4 *If the frontier of $V(N)$ is flat, i.e., there exist $\lambda \in \Lambda_0^N$ and $\alpha \in \mathbb{R}$ such that*

$$V(N) = \left\{x \in \mathbb{R}^N : \sum_{i\in N}\lambda_i x_i \leq \alpha\right\},$$

then, in the subgame that begins after order $\pi \in \Pi$ is chosen, there exists a unique subgame perfect equilibrium payoff allocation given by $\psi^{\rho,\pi}(x^\rho)$ and characterized by

$$\lambda_i\psi^{\rho,\pi}\left(x^\rho\right)_i = \rho\lambda_i x_i^\rho + (1-\rho)\left(v^\lambda\left(\overline{P_i^\pi}\right) - v^\lambda\left(P_i^\pi\right)\right)$$

for all $i \in N$. Moreover, the unique expected subgame perfect equilibrium payoff allocation x^ρ is the (unique) Shapley NTU value, i.e.,

$$\lambda_i x_i^\rho = Sh_i(N, v^\lambda)$$

for all $i \in N$ and all $\rho \in [0,1)$.

Proof. Existence is guaranteed by Proposition 5.3. We now prove uniqueness. Given the definition of $V(N)$, $\alpha = v^\lambda(N)$. Moreover, for each $\pi \in \Pi$ and $i = \pi_k \in N$,

$$\lambda_i\tau_i(x) = \lambda_i \max\left\{t \in \mathbb{R} : \lambda_i t + \sum_{j\in N\setminus\{i\}}\lambda_j x_j \leq v^\lambda(N)\right\} = v^\lambda(N) - \sum_{j\in N\setminus\{i\}}\lambda_j x_j$$

and hence

$$
\begin{aligned}
\lambda_i \psi^{\rho,\pi}(x)_i &= \min \left\{ \lambda_i \tau_i \left(\rho x_{P_i^\pi} + (1-\rho) y \times \psi^{\rho,\pi}(x)_{N \setminus \overline{P_i^\pi}} \right) : y \in V\left(P_i^\pi\right) \right\} \\
&= \min_{y \in V(P_i^\pi)} \left\{ v^\lambda(N) - \sum_{j \in P_i^\pi} \lambda_j \left(\rho x_j + (1-\rho) y_j\right) - \sum_{j \in N \setminus \overline{P_i^\pi}} \lambda_j \psi^{\rho,\pi}(x)_j \right\} \\
&= v^\lambda(N) - \sum_{j \in N \setminus \overline{P_i^\pi}} \lambda_j \psi^{\rho,\pi}(x)_j - \max_{y \in V(P_i^\pi)} \left\{ \sum_{j \in P_i^\pi} \lambda_j \left(\rho x_j + (1-\rho) y_j\right) \right\} \\
&= v^\lambda(N) - \sum_{j \in N \setminus \overline{P_i^\pi}} \lambda_j \psi^{\rho,\pi}(x)_j - \max_{y \in V(P_i^\pi)} \left\{ \sum_{j \in P_i^\pi} \lambda_j \rho x_j + \sum_{j \in P_i^\pi} \lambda_j (1-\rho) y_j \right\} \\
&= v^\lambda(N) - \sum_{j \in N \setminus \overline{P_i^\pi}} \lambda_j \psi^{\rho,\pi}(x)_j - \rho \sum_{j \in P_i^\pi} \lambda_j x_j - (1-\rho) \max_{y \in V(P_i^\pi)} \sum_{j \in P_i^\pi} \lambda_j y_j \\
&= v^\lambda(N) - \sum_{j \in N \setminus \overline{P_i^\pi}} \lambda_j \psi^{\rho,\pi}(x)_j - \rho \sum_{j \in P_i^\pi} \lambda_j x_j - (1-\rho) v^\lambda(P_i^\pi).
\end{aligned}
$$

By definition, $\psi^{\rho,\pi}(x)$ is Pareto efficient, and thus the above equality can be rewritten as

$$
\sum_{j \in P_i^\pi} \lambda_j \psi^{\rho,\pi}(x)_j = \rho \sum_{j \in P_i^\pi} \lambda_j x_j + (1-\rho) v^\lambda(P_i^\pi). \tag{4}
$$

Analogously, when $k < n$,, by taking $i = \pi_{k+1}$,

$$
\sum_{j \in \overline{P_i^\pi}} \lambda_j \psi^{\rho,\pi}(x)_j = \rho \sum_{j \in \overline{P_i^\pi}} \lambda_j x_j + (1-\rho) v^\lambda\left(\overline{P_i^\pi}\right). \tag{5}
$$

Notice that (4) and (5) hold for any $i \in N$ and any $\pi \in \Pi$. By substracting (4) from (5), we obtain, for each $i \in N$ and $\pi \in \Pi$,

$$
\lambda_i \psi^{\rho,\pi}(x)_i = \rho \lambda_i x_i + (1-\rho) \left(v^\lambda\left(\overline{P_i^\pi}\right) - v^\lambda(P_i^\pi) \right). \tag{6}
$$

Let x^ρ be a subgame perfect equilibrium payoff allocation, i.e.,

$$
x^\rho = \frac{1}{n!} \sum_{\pi \in \Pi} \psi^{\rho,\pi}\left(x^\rho\right).
$$

By taking $x = x^\rho$ and averaging on π in (6), we obtain, for each $i \in N$,

$$
\lambda_i x_i^\rho = \rho \lambda_i x_i^\rho + (1-\rho) Sh_i(v^\lambda).
$$

Since $\rho < 1$, de deduce that $\lambda_i x_i^\rho = Sh_i(v^\lambda)$ for all $i \in N$, and hence x^ρ does not depend on ρ and it is the Shapley NTU value of V. ■

Proposition 5.5 *Let $(x^\rho)_{\rho\in[0,1)}$ such that, for each $\rho \in [0,1)$, $x^\rho \in V(N)$ is a subgame perfect equilibrium payoff allocation and there exists $\lim_{\rho\to 1} x^\rho = x^1$, then x^1 is a Shapley NTU value of V.*

Proof. Under Corollary 5.1, as $\rho \to 1$, x^ρ approaches each $\psi^{\rho,\pi}(x^\rho)$, all of them on the Pareto surface of $V(N)$. Closedness of $V(N)$ (given by (P1)) assures that x^1 is Pareto efficient also. Let $\lambda^1 \in \Lambda^N_{\mu(V)}$ be the outward unit normal vector to x^1 on $\partial V(N)$. We then associate with each $\rho \in [0,1)$ and each $\pi \in \Pi$ an NTU game $V^{\rho,\pi}$ with $V^{\rho,\pi}(N)$ flat. Let $\lambda^{\rho,\pi} \in \Lambda^N_{\mu(V)}$ be the outward unit normal to the hyperplane passing through the (Pareto efficient) payoff allocations $(z^i)_{i\in N}$ defined as in the proof of Proposition 5.3. Let $V^{\rho,\pi}(N)$ be the resulting half-space. If the hyperplane is not unique, following Hart and Mas-Colell (1996), we choose $\lambda^{\rho,\pi}$ the closest possible to λ^1. We also define $V^{\rho,\pi}(S) = V(S)$ for all $S \subset N$. Since $\psi^{\rho,\pi}(x^\rho)$ approaches x^1 as $\rho \to 1$ (by Corollary 5.1), the smoothness of $\partial V(N)$ (given by (P4)) implies that $\lambda^{\rho,\pi} \to \lambda^1$ for each $\pi \in \Pi$. Therefore, as $\rho \to 1$, each $V^{\rho,\pi}$ approaches the game V^1 defined as $V^1(N) = V^{\lambda^1}(N)$ and $V^1(S) = V(S)$ otherwise. Consider the variation of the mechanism so that, once order π is chosen, the agents play on game $V^{\rho,\pi}$. By construction, for each $\pi \in \Pi$, $\psi^{\rho,\pi}(x^\rho)$ coincides in both V and V^π. Because of the characterization of Proposition 5.1 and Proposition 5.2, $\psi^{\rho,\pi}(x^\rho)$ remains as the (unique) subgame perfect equilibrium payoff allocation for each $V^{\rho,\pi}$. Fix $i \in N$. Under Proposition 5.4,

$$\psi^{\rho,\pi}(x^\rho)_i = \rho x^\rho_i + (1-\rho)\frac{\left(v^{\rho,\pi}\right)^{\lambda^{\rho,\pi}}\left(\overline{P^\pi_i}\right) - \left(v^{\rho,\pi}\right)^{\lambda^{\rho,\pi}}\left(P^\pi_i\right)}{\lambda^{\rho,\pi}_i}.$$

Hence,

$$x^\rho_i = \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\psi^{\rho,\pi}(x^\rho)_i = \rho x^\rho_i + (1-\rho)\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\frac{\left(v^{\rho,\pi}\right)^{\lambda^{\rho,\pi}}\left(\overline{P^\pi_i}\right) - \left(v^{\rho,\pi}\right)^{\lambda^{\rho,\pi}}\left(P^\pi_i\right)}{\lambda^{\rho,\pi}_i}.$$

Since $\rho < 1$, rearranging terms and dividing by $1-\rho$ we get:

$$x^\rho_i = \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\frac{\left(v^{\rho,\pi}\right)^{\lambda^{\rho,\pi}}\left(\overline{P^\pi_i}\right) - \left(v^{\rho,\pi}\right)^{\lambda^{\rho,\pi}}\left(P^\pi_i\right)}{\lambda^{\rho,\pi}_i}.$$

Taking the limit as $\rho \to 1$, the continuity of the marginal contributions with respect to the hyperplanes implies that

$$x^1_i = \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\frac{\left(v^1\right)^{\lambda^1}\left(\overline{P^\pi_i}\right) - \left(v^1\right)^{\lambda^1}\left(P^\pi_i\right)}{\lambda^1_i} = \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\frac{v^{\lambda^1}\left(\overline{P^\pi_i}\right) - v^{\lambda^1}\left(P^\pi_i\right)}{\lambda^1_i}$$

and thus x^1 is a Shapley NTU value of V. ■

6 Variations

6.1 Player-specific rules

Assume that when player i makes a proposal given by a rule f, this rule should apply not only to coalitions that contain player i, but to any coalition, containing or not player i. Our results are not affected by this change, even though the interpretation fundamentally different. From a technical point of view, the two options would mean the same thing and imply the same result, with only minor difference that two rules might be effectively compatible but incompatible for coalitions without the two proposing players.

Our approach is to allow a rule to apply only to coalitions that contain the proposer, not to all possible coalitions. However, even in the latter case, as a matter of fact, the rule proposed by player i has no binding power on coalitions without containing player i.

6.2 The role of (dis)agreement with new proposals

What happens when players simply propose a rule and a compatible one is randomly chosen at the end of the process (i.e., there are not votes for agreeing or disagreeing with posterior rules)? The game will move on with probability ρ at the end of the last stage if there is no unanimity on the chosen rule, but it is not allowed for other players to have opportunities to accept new proposals from their successors in the order.

Such non-cooperative game implements the Shapley value in TU games and the Nash solution in pure bargaining problems, but not the Shapley NTU value in the general case. When $V(S)$ are hyperplanes in the positive quadrant, we get the consistent value. In the general case, we get the consistent value if we allow the coalitions inside the final partition to renegociate among them in case of breakdown.

6.3 Full randomization

Our mechanism requires to form a randomized order of players at Step 1. Once this randomization is over, the order of the players is fixed until Step $n+1$. We can keep the structure and all the details of the mechanism but make one modification that requires randomization in each step. That is, in each step, a player is randomly selected from the set of remaining players (i.e., those who were not selected in the previous steps), who will either accept a proposal from the present situation of the previous step or propose her rule. We can still implement the Shapley NTU value as players take contingent strategies. That is, a player's proposal would be a list of rules, each corresponding to an ordering of players. Such a modification of the mechanism can only unnecessarily complicate the process.

6.4 Breakdown under unanimity

Step $n+1$ in our mechanism states that, if unanimity is achieved (Case 1), the game is over. This statement is a very reasonable outcome when universal unanimity is obtained. However, we can avoid this case and assume that, independently of what the situation is, the game goes back to Step 0 with probability ρ (Case 2). With this variation, the final payoff allocation in the subgame that begins when order $\pi \in \Pi$ is chosen is not $\psi^{\rho,\pi}(x^\rho)$, but $\rho x^\rho + (1-\rho)\psi^{\rho,\pi}(x^\rho)$. Apart from that, the main results do not change with this variation, that just implies an innocuous delay in the completion of the game. In particular, the expected subgame perfect equilibrium payoff allocation remains the same:

$$\frac{1}{|\Pi|}\sum_{\pi\in\Pi}(\rho x^\rho + (1-\rho)\psi^{\rho,\pi}(x^\rho)) = \rho x^\rho + (1-\rho)\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\psi^{\rho,\pi}(x^\rho) = \rho x^\rho + (1-\rho)x^\rho = x^\rho.$$

7 Concluding remarks

7.1 Universal unanimity and partial unanimity

One important advantage of our mechanism is its robustness, as we do not require "universal" unanimity in reaching an agreement. To appreciate its benefit, consider a game where one player, say player n, is a null player. If this null player prefers to behave in a way to be harmful to the others whenever possible, then we would still be able to implement the Shapley NTU value with our mechanism, even though this null player is not rational. By contrast, in such cases those existing results in Hart and Mas-Colell (1996) or Pérez-Castrillo and Wettstein (2001) will not hold. In Hart and Mas-Colell (1996), unless player n is offered some positive payoff, he will always reject any offer made by other players. Since it is necessary to have unanimous agreement for any offer to be accepted and implemented, it is impossible to obtain the Shapley or consistent value in subgame perfect equilibrium. Likewise, in Pérez-Castrillo and Wettstein (2001), this null player can manage to be a non-proposer by making very negative bids to others (effectively demanding payments from others) and then she will always reject any non-positive offer made by the proposer. Therefore, the Shapley value cannot be implemented in subgame perfect equilibrium, either.

However, in our mechanism, when a non-null player makes a proposal (i.e., suggests a rule), she can simply ignore player n but makes a proposal that is acceptable by all other players, which will still lead to the Shapley NTU value in subgame perfect equilibrium. Player n may still reject the proposal, but it will not affect the outcome because we do not require the universal unanimity to accept a rule. In practice, the presence of player n results in the variation of the mechanism explained in Subsection 6.4. So long as those in a coalition all accept a rule, then this coalition can join the player who proposed the rule to form a coalition. Hence, player n will finally be abandoned and get 0 payoff at the end while the others will get their Shapley NTU value payoffs.

Acknowledgments

Juan Vidal-Puga acknowledges support from Grant TED2021-130241A-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, and also from Spanish Ministerio de Economía y Competitividad (ECO2014-52616-R), Ministerio de Economía, Industria y Competitividad (ECO2017-82241-R), and Xunta de Galicia (GRC 2015/014).

References

Aumann (1985). On the non-transferable utility value: A comment on the Roth-Shafer examples. *Econometrica*, 53:667–677.

Harsanyi (1980). Comments on Roth's paper, 'Values for games without side payments'. *Econometrica*, 48:477–477.

Harsanyi, J. C. (1963). A simplified bargaining model for the n-person cooperative game. *International Economic Review*, 4(2):194–220.

Hart and Kurz (1983). Endogenous formation of coalitions. *Econometrica*, 51:1047–1064.

Hart, S. and Mas-Colell, A. (1996). Bargaining and value. *Econometrica*, 64(2):357–380.

Maschler and Owen (1992). The consistent Shapley value for games without side payments. In Selten, R., editor, *Rational iteration*, pages 5–12. Springer-Verlag, New York.

Maschler, M. and Owen, G. (1989). The consistent Shapley value for hyperplane games. *International Journal of Game Theory*, 18(4):389–407.

Moldovanu, B. and Winter, E. (1994). Core implementation and increase returns to scale for cooperation. *Journal of Mathematical Economics*, 23(6):533–548.

Myerson, R. B. (1980). Conference structures and fair allocation rules. *International Journal of Game Theory*, 9(3):169–182.

Owen, G. (1977). Values of games with a priori unions. In Henn, R. and Moeschlin, O., editors, *Mathematical Economics and Game Theory*, volume 141 of *Lecture Notes in Economics and Mathematical Systems*, pages 76–88. Springer-Verlag, Berlin.

Pérez-Castrillo, D. and Wettstein, D. (2001). Bidding for the surplus: a non-cooperative approach to the Shapley value. *Journal of Economic Theory*, 100(2):274–294.

Roth, A. E. (1980). Values of games without side payments: some difficulties with current concepts. *Econometrica*, 48(2):457–465.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. and Tucker, A., editors, *Contributions to the theory of games*, volume II of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, Princeton NJ.

Stole and Zwiebel (1996). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies*, 63(3):375–410.

van Damme (1986). The Nash bargaining solution is optimal. *Journal of Economic Theory*, 38:78–100.

Vidal-Puga, J. (2008). Forming coalitions and the Shapley NTU value. *European Journal of Operational Research*, 190(3):659–671.