Reaching consensus through approval bargaining

Author:
Laslier, JF; Núñez, M; Pimienta, C; Gonzalez Pimienta, Jose Carlos

Publication details:
Games and Economic Behavior
v. 104
pp. 241 - 251
0899-8256 (ISSN); 1090-2473 (ISSN)

Publication Date:
2017-07-01

Publisher DOI:
https://doi.org/10.1016/j.geb.2017.04.002

REACHING CONSENSUS THROUGH APPROVAL BARGAINING*

JEAN-FRANÇOIS LASLIER[a], MATÍAS NÚÑEZ[b], AND CARLOS PIMIENTA[c]

ABSTRACT. In the Approval Bargaining game, two players bargain over a finite set of alternatives. To this end, each one simultaneously submits a utility function u jointly with a real number α; by doing so she approves the lotteries whose expected utility according to u is at least α. The lottery to be implemented is randomly selected among the most approved ones. We first prove that there is an equilibrium where players truthfully reveal their utility function. We also show that, in any equilibrium, the equilibrium outcome is approved by both players. Finally, every equilibrium is sincere and Pareto efficient as long as both players are partially honest.

KEY WORDS. Approval voting, bargaining, partial honesty, consensual equilibrium.

JEL CLASSIFICATION. C70, C72.

1. INTRODUCTION

An elementary version of the bargaining problem involves two players with complete information who have to decide on the terms of a possible cooperation. The outcome is either an agreement about such terms, or else a conflict, in the case that no agreement is reached. While dynamic bargaining has been extensively explored and often leads to desirable outcomes (in models à la Rubinstein, 1982), the literature on simultaneous bargaining is scant. It has been argued (see, e.g., Osborne and Rubinstein, 1990) that, not to leave room for renegotiation, the bargaining outcome

* We thank Gani Aldashev, Arnaud Dellis, Marc Fleurbaey, Thibault Gajdos, Laurence Jacquet, Don Keenan, Marcus Pivato, William Thomson, Alain Trannoy, Dimitrios Xefteris and seminar participants for useful comments. We also thank the Associate editor and the referees for their useful advice. Jean-François and Matías acknowledge financial support from the projects ANR-14-CE24-0007-01 CoCoRICo-CoDec, ANR-11-LBX-0023-01 and ANR-10-LBX-93. Carlos acknowledges financial support from the Australian Research Council's Discovery Projects funding scheme DP140102426.

[a] PARIS SCHOOL OF ECONOMICS & CNRS, FRANCE.

[b] UNIVERSITÉ PARIS DAUPHINE, PSL RESEARCH UNIVERSITY, CNRS, UMR [7243], LAMSADE, 75016 PARIS, FRANCE.

[c] SCHOOL OF ECONOMICS, UNSW BUSINESS, THE UNIVERSITY OF NEW SOUTH WALES, SYDNEY, AUSTRALIA.

Email addresses: jean-francois.laslier@ens.fr, matias.nunez@dauphine.fr, c.pimienta@unsw.edu.au.

Corresponding Author: Carlos Pimienta.

Date: February 21, 2017.

should be Pareto optimal. Furthermore, if both players are to participate in the bargaining mechanism, then the outcome should not be worse than disagreement.

We design a two-player simultaneous model of bargaining such that, in equilibrium, parties always reach an agreement. Moreover, as long as agents are *partially honest* (Dutta and Sen, 2012) every equilibrium outcome is Pareto efficient. Partial honesty has been recently analyzed by the mechanism design literature and it captures a mild form of preference for honesty. A partially honest agent prefers being sincere over lying whenever sincerity does not lead to a worse outcome.[1] In our model, each player simultaneously approves of a *set of lotteries over the pure alternatives*. A player does so announcing a utility function u and a real number α; by doing so she approves of the lotteries whose expected utility according to u is at least α. A player's announcement is sincere if its utility component coincides with her true utility function.[2] And a partially honest player prefers a sincere announcement if she cannot do better by lying.

The outcome induced by a strategy profiles is determined as follows. If some lotteries are approved by both players, then the two approved sets intersect. We define the *winning set* to be such an intersection and we say that the winning set is *consensual*. If no lottery is approved by both players, then we define the winning set to be the set of lotteries that are approved by at least one agent. In this case, we say that the winning set is *non-consensual*. Finally, the mechanism selects a lottery at random using the uniform probability over the winning set. The alternative to be implemented is decided by this selected lottery.[3] Thus, in the same vein as Babichenko and Schulman (2015) and Núñez and Laslier (2015), one can think of our model as a reinterpretation of approval voting (Brams and Fishburn, 1983; Laslier and Sanver, 2010) as a bargaining mechanism when there are just two voters.[4] Hence, in the sequel, we refer to our bargaining mechanism as *approval bargaining*.

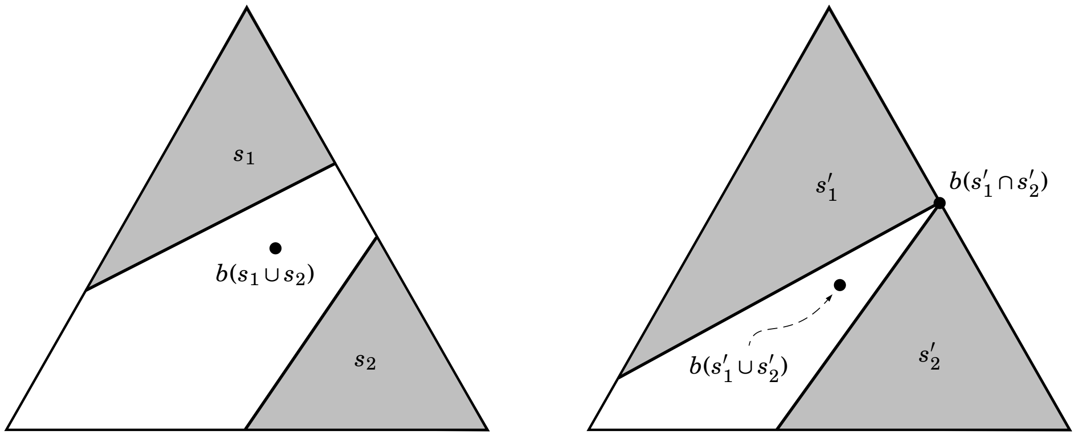
In some sense, our approval bargaining game is similar to Nash's (1953) demand game. In the demand game, two players make simultaneous demands and each one

[1] We discuss how our paper relates to the mechanism design literature at the end of the Introduction. We do not attempt to give a review on the bargaining literature and simply refer the reader to Serrano (2008).

[2] Thus, sincerity in this context implies the sincerity notion used in the approval voting literature in which a strategy is *sincere* if, whenever it contains an alternative, it also contains the other alternatives that the player prefers to it. See Merill and Nagel (1987), Brams (2008), and Núñez (2014) for works dealing with sincerity under approval voting.

[3] Núñez (2015) analyzes a voting rule in this fashion and shows that it leads to type-revelation with many voters.

[4] The difference is that in Babichenko and Schulman (2015) and Núñez and Laslier (2015) agents approve alternatives while, in the current model, agents approve lotteries over alternatives. This is the main driving force behind of the efficiency properties of our model.



(A) Non-consensual strategy profile.

(B) Consensual strategy profile.

FIGURE 1. Strategy profiles in the approval bargaining game.

receives the payoff she requests if both payoffs are jointly feasible and nothing otherwise. Our model is more complex since strategies are not unidimensional and the threat point is decided endogenously. Figure 1 illustrates this in a bargaining situation with three alternatives, each one represented at the corresponding vertex of the simplex. Figure 1a depicts the non-consensual and sincere strategy profile (s_1, s_2) while Figure 1b shows the consensual and sincere strategy profile (s'_1, s'_2). Under (s_1, s_2), player 1 (resp. player 2) approves every lottery in the closed subset labeled s_1 (resp. s_2). The outcome induced by (s_1, s_2) is the uniform probability measure over $s_1 \cup s_2$ and the expectation of such a measure is the barycenter $b(s_1 \cup s_2)$. In Figure 1b, the strategies s'_1 and s'_2 intersect so that the induced outcome is $b(s'_1 \cap s'_2)$. Note that either players can deviate to some non-consensual strategy that induces an outcome arbitrarily close to $b(s'_1 \cup s'_2)$. Hence, $b(s'_1 \cup s'_2)$ is the endogenous threat point that sustains the equilibrium outcome $b(s'_1 \cap s'_2)$.

This example suggests that, under a non-consensual strategy profile, players have two joint incentives: (1) approving a large subset of lotteries so that the induced expected outcome is as close as possible to it and, consequently, (2) playing some sincere strategy that approves every lottery in the upper contour set of some indifference curve. These two incentives work together so that both players approve bigger and bigger sets. The consequence is that a non-consensual strategy profile cannot be an equilibrium. In Section 3, we prove that every equilibrium strategy profile has a nonempty intersection in the same way as in Figure 1b. Finally, note that in this figure either player can deviate to a non-sincere consensual strategy and still induce outcome $b(s'_1 \cap s'_2)$ as long as the resulting intersection is also $b(s'_1 \cap s'_2)$. But partial honesty guarantees that players would rather play the sincere strategy.

Building on these observations, we prove that the approval bargaining game has the following properties.

(1) *Existence of equilibrium*: Every game has an equilibrium in sincere strategies.

(2) *Consensual equilibria*: In every equilibrium, players agree on some subset of lotteries.
(3) *Sincerity and Pareto efficiency*: If players are partially honest, every equilibrium outcome is in sincere strategies and Pareto efficient.
(4) *Balanced equilibrium outcomes*: If p is an equilibrium outcome then a player's upper contour set of p cannot be "too small" unless both players agree on what the best alternative is.

More precisely, the last property takes the relative size of the upper contour sets as a measure of how much the equilibrium outcome favors one player over the other. The smaller a player's upper contour set, the closer the corresponding outcome is to the player's most preferred alternative. Every equilibrium outcome of the game is *balanced* in the following sense: If p is an equilibrium outcome and C_i is the set of lotteries that player i prefers to p, then both players prefer p to selecting a lottery uniformly from the set $C_1 \cup C_2$. Furthermore, Properties (3) and (4) characterize the set of equilibrium payoffs (and, therefore, the set of equilibria): If p is Pareto efficient and balanced, then $u_1(p)$ and $u_2(p)$ are equilibrium payoffs.

The rest of the paper is structured as follows. After providing a brief account of relevant known results in mechanism design, Section 2 presents the model. Equilibrium properties are derived in Section 3 and the characterization of equilibria is presented in Section 4. The Appendix proves two lemmata needed to establish the results, briefly considers the game without assuming partial honesty, and contains the proof of existence of equilibria.

Relationship with the mechanism design literature

Maskin (1999) proves that a two-player, Pareto optimal rule defined on the domain of all strong orderings is Nash implementable if and only if it is dictatorial. In view of this result, Moore and Repullo (1990) and Dutta and Sen (1991) characterize Nash implementation with two agents and use such a characterization to find domain restrictions that yield positive results (see also Busetto and Colognato, 2009). In particular, Dutta and Sen (1991) show that if the set of outcomes is the probability simplex over finitely many alternatives and players have von Neumann-Morgenstern utility functions satisfying suitable conditions, then the correspondence that selects the set of Pareto efficient and individually rational lotteries can be implemented.[5]

Bagnoli and Lipman (1989) argue that most games introduced by the implementation literature are built to be applicable to very general settings rather than for their plausibility. For this reason, these mechanisms are often quite complex. For

[5] Dutta and Sen assume that no agent is indifferent between two pure alternatives, and that there is no affine transformation u_1, u_2 of their utility functions that satisfies either $u_1 = u_2$ or $u_1 = -u_2$. Apart from Proposition 3 in Appendix B, we do not impose any such restriction.

example, the generalized mechanism used to prove the sufficiency result of the characterization mentioned above uses an "integer game". The unappealing features of integer games (Jackson, 1992), among other reasons, has stimulated researchers to investigate the implementation problem using different approaches. A recent one explores the scope for implementation when players are partially honest (see Matsushima, 2008b; Ortner, 2015). Under partial honesty, a player prefers a truthful message when it does not lead to a strictly worse outcome than what she would obtain otherwise. Dutta and Sen (2012) find necessary and sufficient conditions for implementation under partial honesty when there are two players. However, the existing results do not apply to our setting for different reasons. Some need more than two players (Matsushima, 2008b), some use monetary transfers (Matsushima, 2008a; Kartik et al., 2014) and some propose mechanisms that do not seem suitable to be understood as bargaining protocols (for example Dutta and Sen, 2012 and Kartik and Tercieux, 2012 also use integer games).

2. The game

Consider two *players* indexed by $i = 1,2$ and a set of *alternatives* $X := \{x_1,\dots,x_K\}$ with at least two elements. Let $\Delta := \{p \in \mathbb{R}^K_+ \mid \sum p_i = 1\}$ denote the probability simplex over X. We identify an alternative $x \in X$ with the degenerate lottery that assigns probability one to x. Each player i has preferences over Δ represented by a von Neumann-Morgenstern utility function u_i. Let $\mathscr{U}$ be the set of all linear functions defined on Δ. We assume that a player's best and worst alternatives are associated to different utility values. Utilities are normalized so that, for each i, $\max_{x\in X} u_i(x) = 1$ and $\min_{x\in X} u_i(x) = 0$.

A player's strategy is an announcement of a function in $\mathscr{U}$ together with a real number in $[0,1]$. Hence, players have the common strategy set $S := \mathscr{U} \times [0,1]$. Formally, a strategy $s_i := \{\omega_i, \alpha_i\}$ is the subset of lotteries $\{p \in \Delta : \omega_i(p) \geq \alpha_i\}$. That is, s_i is the subset of lotteries in Δ that give a player with utility function ω_i a level of utility larger or equal to α_i. We interpret a strategy as the set of lotteries that a player "approves". A strategy $s_i = \{\omega_i, \alpha_i\} \in S$ is *sincere* for player i if ω_i is equal to player i's true utility function u_i. Player i's set of sincere strategies is denoted $\mathscr{S}_i \subset S$.

We introduce some preliminary facts to describe the outcome induced by a strategy profile (s_1, s_2). Recall that, given a convex subset $A \in \Delta$, its affine hull aff(A) is the smallest affine set containing A. The dimension of a nonempty convex subset A, denoted by dim(A), is the dimension of its affine hull (see Rockafellar, 1997). The dimension of a finite union of convex sets $\bigcup_{z\in Z} A_z$ is equal to $\max_{z\in Z} \dim(A_z)$. Let λ_n be the Lebesgue measure in $\mathbb{R}^n$. Since we work in the probability simplex over X, we often refer to λ_{K-1}. For simplicity, we simply write λ instead of λ_{K-1}. For any n-dimensional set $A \in \Delta$ that can be written as the finite union of convex sets, the uniform measure with support A is given by $\mu(\cdot \mid A) = \lambda_n(\cdot)/\lambda_n(A)$. For any n-dimensional set $A \in \Delta$, its barycenter $b(A)$ is

$$b(A) := \int_A p \, d\mu(p \mid A).$$

By convention, we let $b(\varnothing) = b(\Delta)$. The strategy profile $s = (s_1, s_2) \in S$ induces the outcome $\theta(s_1, s_2)$ defined as:

$$\theta(s_1, s_2) := \begin{cases} b(s_1 \cap s_2) & \text{if} \quad s_1 \cap s_2 \neq \varnothing, \\ b(s_1 \cup s_2) & \text{otherwise.} \end{cases}$$

If $s_1 \cap s_2 \neq \varnothing$ there are lotteries that both players approve and we say that the strategy profile (s_1, s_2) and the induced outcome $b(s_1 \cap s_2)$ are *consensual*. If $s_1 \cap s_2 = \varnothing$ we say that the strategy profile (s_1, s_2) and the induced outcome $b(s_1 \cup s_2)$ are *non-consensual*. In either case, the relevant set has a well-defined dimension so that $\theta(s_1, s_2)$ is always well-defined.

We assume that players are *partially honest*, that is, they prefer playing sincere strategies as long as they cannot obtain a better outcome by not doing so. We follow the formal definition of partial honesty given by Dutta and Sen (2012). We denote by $\succeq_i$ player i's ordering over the set of strategy profiles $S \times S$ when she is partially honest. Its asymmetric component is denoted by $\succ_i$.[6]

Definition 1. Player i is partially honest if for each $s_i, s'_i, s_j \in S$,

(1) If $u_i(\theta(s_i, s_j)) \geq u_i(\theta(s'_i, s_j))$, $s_i \in \mathscr{S}_i$, and $s'_i \notin \mathscr{S}_i$, then $(s_i, s_j) \succ_i (s'_i, s_j)$.
(2) In all other cases, $(s_i, s_j) \succeq_i (s'_i, s_j)$ if and only if $u_i(s_i, s_j) \geq u_i(s'_i, s_j)$.

The first part of the definition represents the player's partial preference for honesty. She strictly prefers the strategy profile (s_i, s_j) to (s'_i, s_j) when s_i is a sincere strategy and s'_i is not, provided that the outcome $\theta(s_i, s_j)$ is at least as good as $\theta(s'_i, s_j)$. The second part of the definition implies that, in every other case, the player's preference ordering over strategy profiles is the same as the one induced by her preference ordering over lotteries.

These rules describe the simultaneous approval bargaining game $\Phi = (S, S, u_1, u_2)$. We focus on equilibrium in pure strategies.

Definition 2. A strategy profile (s_1, s_2) is an equilibrium if for every player i and every $s'_i \in S$ we have $(s_i, s_j) \succeq_i (s'_i, s_j)$.

Our first result is about equilibrium existence. In the Appendix, we prove that the game admits an equilibrium in sincere strategies for any specification of the players' utilities. Note that the outcome function is discontinuous so that existence does not follow from standard fixed-point theorems. Indeed, the outcome $\theta(s_1, s_2)$ might "jump" discontinuously whenever the limit of a sequence of non-consensual

[6] Hereinafter, once we introduce player i we let player j be the other player so that $i \neq j$.

strategy profiles is a consensual strategy profile. (See, for instance, Figure 1b. The outcome induced by any non-consensual strategy profile close to (s'_1, s'_2) is close to $b(s'_1 \cup s'_2)$, far away from $\theta(s'_1, s'_2) = b(s'_1 \cap s'_2)$.)

Theorem 1. *Every approval bargaining game Φ has an equilibrium in sincere strategies.*

3. Consensus, sincerity and Pareto efficiency

Consider some strategy s_j of player j and let $\alpha_i^* := \max_{p \in s_j} u_i(p)$ be the utility associated to the best lottery in s_j from player i's viewpoint. Player i's sincere strategy $s_i^* = \{u_i, \alpha_i^*\}$ induces outcome $b(s_i^* \cap s_j)$ and yields utility $u_i(b(s_i^* \cap s_j)) = \alpha_i^*$ to player i. Therefore, s_i^* is player i's best *consensual response* to s_j, i.e., no other consensual response to s_j does better than s_i^*. Every other consensual and sincere response to s_j does strictly worse because it approves an open subset of lotteries in s_j that yield player i a payoff strictly lower than α_i^*. And every other consensual and non-sincere response to s_j can only induce an outcome no better than $b(s_i^* \cap s_j)$. Since players are partially honest we conclude that s_i^* is the *unique* best consensual response to s_j.

But even if s_i^* is player i's best consensual response to s_j, it may not be the best response overall. Hence assume that s_i is player i's best response to s_j and that the strategy profile (s_1, s_2) is non-consensual. If s_j is full dimensional (that is, of dimension $K-1$), then s_i is the *unique* best response to s_j. To see why, note that, provided that both s_i and s_j are full dimensional, the induced outcome is a convex combination between $b(s_i)$ and $b(s_j)$. Such a convex combination depends on the relative measures of s_1 and s_2. Therefore, player i's best response s_i must be "big" relative to s_j so that the induced outcome is as close as possible to $b(s_i)$. It must also contain only lotteries that give player i a sufficiently high payoff so that player i's payoff from $b(s_i)$ is as high as possible. More precisely, if s_i is player i's best response to s_j, then s_i must coincide with the lotteries she prefers to $b(s_i \cup s_j)$.

Lemma 1. *Let $s_j \in S$ be a full dimensional strategy and let $s_i \in S$ be a non-consensual best response to s_j. Then $s_i = \{u_i, u_i(b(s_i \cup s_j))\}$.*

The proof is in the Appendix. It is important that strategy s_j be full dimensional. Otherwise, player i need not have a best response, as the next example shows.

Example 1. Let $X := \{x, y, z\}$ and let x and y be, respectively, player 1's and player 2's unique most preferred alternatives. In particular, we have $u_1(x) = u_2(y) = 1$ and $u_1(y) < 1$. Consider player 2's strategy $s_2 = \{u_2, 1\}$ and note that it is not full dimensional. No consensual strategy of player 1 is a best response to s_2 since any such strategy gives her utility $u_1(y)$, whereas strategy $\{u_1, 1\}$ gives her utility $\frac{1}{2} + \frac{1}{2}u_1(y) > u_1(y)$. In turn, take $\varepsilon > 0$ small enough and consider player 1's strategy $s_1^\varepsilon = \{u_1, 1-\varepsilon\}$. Since $\lambda(s_2) = 0$ and $s_1^\varepsilon \cap s_2 = \emptyset$ for ε small enough, $u_1(s_1^\varepsilon, s_2) =$

$u_1(b(s_1^\varepsilon \cup s_2)) = u_1(b(s_1^\varepsilon))$. Therefore, $u_1(s_1^\varepsilon, s_2)$ converges to 1 as ε decreases. But $u_1(s_1^0, s_2) = \frac{1}{2} + \frac{1}{2}u_1(y) < 1$. That is, player 1 has no best response to s_2.

If both players agree on what the best alternative is, it is easy to construct a consensual strategy profile such that both players coordinate to implement this alternative. If players do not have a common best alternative and play a non-consensual strategy profile, then they would want to approve a large set relative to the other player to "attract" the outcome in the direction where their utility increase as much as possible. The end result is that both strategies would be large enough to intersect. Therefore, we obtain:

Proposition 1. *Every equilibrium is consensual.*

Proof. Suppose to the contrary that there is a non-consensual equilibrium (s_1, s_2). If both s_1 and s_2 are full dimensional then, by Lemma 1, $\theta(s_1, s_2)$ belongs to both s_1 and s_2. Hence, (s_1, s_2) is a consensual strategy profile.

Assume now that player j's strategy s_j is lower dimensional. Player i's payoff under (s_1, s_2) is 1 because she can obtain a payoff as close to 1 as she wants by choosing $\varepsilon > 0$ and playing $s_i^\varepsilon := \{u_i, 1-\varepsilon\}$. But then, player i's strategy is lower-dimensional as well because otherwise $u_i(\theta(s_1, s_2)) = u_i(b(s_i)) \neq 1$ (recall that there is a best and a worst alternative so preferences are locally nonsatiated). Repeating the argument, player j's equilibrium utility is also 1. Now, if s_j has the same dimension as $s_i \cup s_j$ then every lottery approved in s_j is among player i's best lotteries. Therefore, player i's payoff cannot decrease if she plays the sincere strategy $\{u_i, 1\}$ against s_j. Thus, by partial honesty, s_i is also sincere and satisfies $\{u_i, 1\} \subset s_i$. But then (s_1, s_2) is a consensual strategy profile as we wanted. □

Since every every equilibrium is consensual, in equilibrium, players play their best consensual response against each other. As noticed above, partial honesty implies that the unique best consensual response to the opponent's strategy is a sincere strategy. Thus, players only play sincere strategies in equilibrium. In turn, this implies that only Pareto efficient lotteries are equilibrium outcomes. Of course, in our setting, a lottery $p \in \Delta$ is *Pareto efficient* if there is no other lottery $q \in \Delta$ such that $u_i(q) \geq u_i(p)$ for $i = 1, 2$ with at least one being a strict inequality.

Theorem 2. *Every equilibrium is sincere and Pareto efficient.*

Proof. Let (s_1, s_2) be an equilibrium. From Theorem 1 we have $s_1 \cap s_2 \neq \emptyset$. Furthermore, such an intersection only contains player i's most preferred lotteries in s_j. Since player i can induce such an outcome playing a sincere strategy, partial honesty implies that every equilibrium is in sincere strategies.

To prove Pareto Efficiency, let $\alpha_i^* := \max_{p \in s_j} u_i(p)$ so that we can write $s_i = \{u_i, \alpha_i^*\}$. Suppose that there is $q \in \Delta$ such that $u_i(q) \geq \alpha_i^*$ for $i = 1, 2$, with strict inequality for at least one player. Then q belong to both s_1 and s_2 because they are sincere strategies. But this contradicts the definition of α_i^* for at least one $i = 1,2$. Therefore, every lottery in the winning set of an equilibrium is Pareto efficient. □

Since only Pareto efficient lotteries are equilibrium outcomes, only Pareto efficient alternatives are in the support of such outcomes. If, say, alternative x_1 is Pareto dominated by alternative x_2, a lottery p with $p_1 > 0$ is Pareto dominated by the lottery q defined by:

$$\begin{aligned} q'_1 &= 0, \\ q'_2 &= p_1 + p_2, \text{ and} \\ q'_k &= p_k \text{ for } k = 3, \ldots, K. \end{aligned}$$

This shows that any lottery that assigns positive probability to inefficient alternatives is inefficient. Since only efficient lotteries are equilibrium outcomes of the game then, ex-post, players never have a common incentive to renegotiate once the equilibrium outcome has been realized into some alternative in X.

4. A CHARACTERIZATION OF EQUILIBRIUM OUTCOMES

As we have seen so far, the approval bargaining game only admits sincere and consensual equilibria, which lead to efficient outcomes. If players have a common best alternative, then there is a unique efficient outcome and all equilibria select this outcome. However, when players do not share their most preferred alternative, not every Pareto efficient outcome can be sustained in equilibrium. Hence, from now on, we focus on the case where players do not agree on what the best alternative is. This section characterizes equilibrium outcomes in this scenario by providing the maximum and minimum payoffs that are possible in equilibrium.

To show that a given strategy profile is an equilibrium we need to prove that (1) both players are playing their best consensual response, and that (2) they do not gain by deviating to their best non-consensual response. We now study how the best consensual and non-consensual responses of a player behave as the opponent changes her strategy.

For each player i, Theorem 2 allows us to focus on her set of sincere strategies $\mathscr{S}_i \subset S$. For each $\alpha_i \in [0,1]$ we associate the sincere strategy $\{u_i, \alpha_i\}$. And for each $\alpha_j \in [0,1]$, we let $\mathrm{CU}_i(\alpha_j)$ denote player i's utility value to the outcome induced by her best sincere consensual response to $\{u_j, \alpha_j\}$. Instead of working with the analogous expression for player i's best sincere non-consensual response (which, as argued before, might not exist), we introduce the following function defined on $\mathscr{U} \times (0,1) \times \mathscr{U} \times (0,1)$,

$$f_i(s_i, s_j) := \frac{\lambda(s_i)}{\lambda(s_i) + \lambda(s_j)} u_i(b(s_i)) + \frac{\lambda(s_j)}{\lambda(s_i) + \lambda(s_j)} u_i(b(s_j)).$$

Let $\mathrm{NU}_i(\alpha_j)$ denote player i's maximal value of $f_i(\cdot,\{u_j,\alpha_j\})$. The value of this function coincides with $u_i(b(s_1 \cup s_2))$ whenever $s_1 \cap s_2 = \varnothing$. Using the same arguments as in the proof of Lemma 1, one can prove that the maximizer s'_i of $f_i(\cdot,\{u_j,\alpha_j\})$ exists and satisfies $s'_i = \{u_i, f_i(s'_i, s_j)\}$. This fact means that the inequality $\mathrm{CU}_i(\alpha_j) < \mathrm{NU}_i(\alpha_j)$ is equivalent to saying that the best response to $\{u_j,\alpha_j\}$ is non-consensual and, furthermore, equal to s'_i. Similarly, the inequality $\mathrm{CU}_i(\alpha_j) \geq \mathrm{NU}_i(\alpha_j)$ is equivalent to saying that the best response to $\{u_j,\alpha_j\}$ is consensual.

The next Lemma describes, for each player i, her best response to any sincere strategy of her opponent. Given utility functions (u_i,u_j), there is a real number η^i such that if player j plays some sincere strategy $\{u_j,\alpha_j\}$ with $\alpha_j \leq \eta^i$ then player i's best response is consensual (i.e. $\mathrm{CU}_i(\alpha_j) \geq \mathrm{NU}_i(\alpha_j)$). On the other hand, if player j plays some sincere strategy $\{u_j,\alpha_j\}$ with $\alpha_j > \eta^i$ then player i's best response is non-consensual (i.e. $\mathrm{CU}_i(\alpha_j) < \mathrm{NU}_i(\alpha_j)$).

Lemma 2. *If players do not have a common best alternative, for each player i, there exists a unique $\eta^i \in (0,1)$ such that:*

$$\mathrm{CU}_i(\alpha_j) \geq \mathrm{NU}_i(\alpha_j) \text{ if and only if } \alpha_j \leq \eta^i.$$

We can now complete the full characterization of equilibria of the approval bargaining game.

Theorem 3. *Suppose that players do not have a common best alternative and let (α_1,α_2) be a utility profile derived from some Pareto efficient lottery. The profile $(\{u_1,\alpha_1\},\{u_2,\alpha_2\})$ is an equilibrium if and only if $\alpha_j \leq \eta^i$ for both $i = 1,2$.*

Proof. Let $p \in \Delta$ be Pareto efficient and let $\alpha_i = u_i(p)$ for $i = 1,2$. Consider the strategy profile $(\{u_1,\alpha_1\},\{u_2,\alpha_2\})$. The lottery p is approved by both $\{u_1,\alpha_1\}$ and $\{u_2,\alpha_2\}$ and, because p is Pareto efficient, such an intersection has an empty interior relative to Δ. Thus, no player has an incentive to deviate to a different consensual strategy. Furthermore, since $\alpha_1 \leq \eta^2$ and $\alpha_2 \leq \eta^1$, the previous lemma implies that no player has an incentive to deviate to a non-consensual strategy.

On the other hand, let (α_1,α_2) be an equilibrium payoff. From Theorem 2 we know that players are playing $(\{u_1,\alpha_1\},\{u_2,\alpha_2\})$ which, by Proposition 1, is a consensual strategy profile. Because players do not have an incentive to deviate to a non-consensual strategy we have $\alpha_1 \leq \eta^2$ and $\alpha_2 \leq \eta^1$. □

Thus, a player's set of equilibrium payoffs (as well as the set of equilibrium strategies) is a closed interval. Another implication of this theorem is that a player's upper contour set of an equilibrium outcome cannot be "too small" relative to that of the other player. To formalize this idea we introduce the following definition:

Definition 3. Let $C_i(p)$ be player i's upper contour set of the lottery p. We say that p is a *balanced* lottery if $u_i(p) \geq u_i(b(C_1(p) \cup C_2(p)))$ for $i = 1,2$.

That is, a lottery p is balanced if both players prefer that lottery to the expected outcome of the uniform measure over the set of lotteries that at least one agent prefers to p. Theorem 3 implies that, if players do not have a common best alternative, a payoff vector is an equilibrium payoff vector if and only if it is generated by an efficient and balanced outcome. Of course, almost trivially, if players have a common best alternative then the equilibrium outcome is also efficient and balanced. We have also assumed from the outset that a player's best and worst alternative are associated to different utility levels. If this is not the case for player i then partial honesty implies that player i can only play $s_i = \Delta$ or $s_i = \emptyset$. In either case, it follows that the resulting equilibrium outcome is again efficient and balanced. Hence, we obtain the following characterization of equilibrium payoffs.

Corollary 1. *A payoff vector is an equilibrium payoff vector if and only if it is generated by an efficient and balanced lottery.*

The last result of this section specifies a lower bound on the utility level that a player can obtain in equilibrium.

Proposition 2. *If $u_1 \neq -u_2$, each player i's equilibrium payoff is strictly larger than $u_i(b(\Delta))$. If $u_1 = -u_2$, the unique equilibrium payoff is $u_i(b(\Delta))$.*

Proof. The sincere strategy $\hat{s}_i = \{u_i, u_i(b(\Delta))\}$ guarantees player i a payoff of at least $u_i(b(\Delta))$ regardless of player j's strategy. Any equilibrium outcome is, then, at least as good as $b(\Delta)$ for both players. This fact readily implies that if $u_1 = -u_2$ then the unique equilibrium outcome is $b(\Delta)$. Note that if $u_1 \neq -u_2$ then an efficient lottery p satisfying $u_i(p) = u_i(b(\Delta))$ cannot be balanced. Indeed, player i strictly prefers $b(\hat{s}_i \cup \{u_j, u_j(p)\})$ to p. Thus, in equilibrium, each player's payoff is strictly larger than the one associated to the barycenter of the simplex. □

Theorem 3 and Proposition 2 jointly imply that, unless $u_1 = -u_2$, the set of equilibria of the approval bargaining game consist of a continuum such that both players obtain strictly more than their utility to $b(\Delta)$. We conclude this section with an example that illustrates this result. The example also shows that we can easily use the equilibrium characterization in Theorem 3 to compute the relevant bounds of the set of equilibria of any given game.

Example 2. Consider a bargaining situation with set of alternatives $X = \{x, y, z\}$. Players 1 and 2 have utility functions $u_1 = (1, \delta, 0)$ and $u_2 = (0, \mu, 1)$. We consider the family of games where $\delta = \mu$. If $\delta = .5$, then $b(\Delta)$ is the unique equilibrium outcome. If $\delta < .5$, then the set of Pareto efficient lotteries only includes lotteries that assign zero weight to the inefficient alternative y. This implies that if player i obtains the equilibrium payoff α_i^* then player j obtains the equilibrium payoff $\alpha_j^* = 1 - \alpha_i^*$. If $\delta > .5$, then lotteries that assign strictly positive probability to both x and z are

TABLE 1. Player 1's uniform benchmark ($u_1(b(\Delta))$), minimum ($\underline{\alpha}_1$), and maximum ($\bar{\alpha}_1$) equilibrium payoffs for utility profiles $u_1 = (1, \delta, 0)$ and $u_2 = (0, \mu, 1)$. Values are rounded up to three decimal places.

$\delta = \mu$	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1
$u_1(b(\Delta))$	.333	.367	.4	.433	.467	.5	.533	.567	.6	.633	.667
$\underline{\alpha}_1$	.442	.447	.454	.462	.476	.5	.543	.600	.671	.764	1
$\bar{\alpha}_1$	.558	.553	.447	.538	.524	.5	.638	.743	.832	.915	1

Pareto dominated. In this case, if player i obtains the equilibrium payoff α_i^*, then player j obtains the equilibrium payoff

$$\alpha_j^* = \begin{cases} \alpha_i^* + (\delta - \alpha_i^*)/\delta & \text{if } \alpha_i^* \le \delta, \text{ or} \\ (1-\alpha_i^*)\delta/(1-\delta) & \text{if } \alpha_i^* \ge \delta. \end{cases}$$

Table 1 shows player i's minimum $\underline{\alpha}_i$ and maximum $\bar{\alpha}_i$ equilibrium payoffs for any given $\delta = \mu \in \{0, .1, .2, \dots, 1\}$ together with her utility to the barycenter of the simplex. Of course, in an equilibrium where player i's payoff is at its maximum, player j's equilibrium payoff must be at its minimum and vice versa.

5. Conclusion

This paper develops a simultaneous bargaining mechanism between two players. It analyzes its equilibrium properties assuming that players are partially honest. The approval bargaining game triggers an agreement between the players in every equilibrium. In equilibrium, the players' best responses are sincere which, in turn, implies that any equilibrium outcome is Pareto efficient. Moreover, partial honesty allows us to derive a characterization of equilibrium payoffs: a payoff profile is an equilibrium one if and only if it is induced by a Pareto efficient and balanced lottery.

A natural research question is whether this mechanism can be extended to many players. The answer to this question seems far from obvious. With two players, they either agree or they do not. However, this duality is lost with three or more players. The main problem seems to be what the rules of the game should specify to determine the outcome when some but not all players agree on some subset of lotteries. While one might think of several possible extensions, none of them seems to conveniently extend the properties of the current two-player approval bargaining game.

Appendix A. Proofs of the lemmata

Proof of Lemma 1. For some full dimensional $s_j \in S$, let $s_i' = \{\omega_i, \beta_i\}$ be a non-consensual best response to s_j for player i. Define $\alpha_i := u_i(b(s_i' \cup s_j))$ and $s_i := \{u_i, \alpha_i\}$. Assume that $s_i' \neq s_i$ so that at least one of the following two open sets is nonempty:

$$A = \{p : \omega_i(p) > \beta_i \text{ and } u_i(p) < \alpha_i\},$$
$$B = \{p : u_i(p) > \alpha_i \text{ and } \omega_i(p) < \beta_i\}.$$

The set A contains lotteries that belong to s_i' but not to s_i and B contains lotteries that belong to s_i but not to s_i'. Also, since s_i' is a best response to s_j, the intersection $s_i \cap s_j$ has zero λ-measure because s_j cannot contain lotteries p that satisfy $u_i(p) > \alpha_i$. We can now estimate

$$\begin{aligned} u_i(b(s_i \cup s_j)) &= \frac{\int_{s_i'} u_i(p)d\lambda - \int_A u_i(p)d\lambda + \int_B u_i(p)d\lambda + \int_{s_j} u_i(p)d\lambda}{\lambda(s_i') - \lambda(A) + \lambda(B) + \lambda(s_j)} \\ &= \frac{\left[\lambda(s_i') + \lambda(s_j)\right]\alpha_i - \int_A u_i(p)d\lambda + \int_B u_i(p)d\lambda}{\lambda(s_i') - \lambda(A) + \lambda(B) + \lambda(s_j)} \\ &> \frac{\left[\lambda(s_i') + \lambda(s_j)\right]\alpha_i - \lambda(A)\alpha_i + \lambda(B)\alpha_i}{\lambda(s_i') - \lambda(A) + \lambda(B) + \lambda(s_j)} = \alpha_i = u_i(b(s_i' \cup s_j)). \end{aligned}$$

If $s_i \cap s_j = \varnothing$ then the previous inequality implies that player i can do better than s_i' by deviating to the non-consensual strategy s_i. On the other hand, if $s_i \cap s_j \neq \varnothing$ then s_i is a consensual best response to s_j that satisfies $u_i(b(s_i \cap s_j)) = \alpha_i$. But the previous equality implies that, for ϵ small enough, the sincere strategy $s_i^\varepsilon = \{u_i, \alpha_i - \varepsilon\}$ is a non-consensual response to s_j that still satisfies $u_i(b(s_i^\varepsilon \cup s_j)) > \alpha_i$. In either case, s_i' is not a best response to s_j. We have reached a contradiction. □

Proof of Lemma 2. We start by noting that CU_i and NU_i are continuous functions on $(0,1)$. Furthermore, CU_i is non-increasing in α_j because, as α_j increases, the set of lotteries approved by the sincere strategy $\{u_j, \alpha_j\}$ decreases in size.

If α_j is close enough to 0, player i can obtain a payoff close to 1 by playing a consensual best response, while she can only get a payoff close to $u_i(\Delta)$ by playing a non-consensual response, hence, $\text{CU}_i(\alpha_j) \geq \text{NU}_i(\alpha_j)$. In turn, if α_j is close enough to 1 then player i's best response to $\{u_j, \alpha_j\}$ is non-consensual. Indeed, player i can obtain a utility close to 1 by playing a non-consensual strategy whereas she can only get, at most, a utility close to the one corresponding to her second most preferred alternative if she plays a consensual best response (because players do not have a common best alternative). Thus, $\text{CU}_i(\alpha_j) < \text{NU}_i(\alpha_j)$. The continuity of CU_i and NU_i as functions of α_j implies the existence of some $\eta^i \in (0,1)$ for which $\text{CU}_i(\eta^i) = \text{NU}_i(\eta^i)$.

To prove uniqueness, suppose that $\text{CU}_i(\alpha_j) = \text{NU}_i(\alpha_j)$ for some $\alpha_j > \eta^i$. Since the lotteries approved by $\{u_j, \alpha_j\}$ are a subset of the lotteries approved by $\{u_j, \eta^i\}$ and the difference between the two sets has positive measure and only contains lotteries that give player i utility less than η^i we have $\text{NU}_i(\alpha_j) > \text{NU}_i(\eta^i)$. Moreover CU_i is nonincreasing in α_j so that $\text{CU}_i(\eta^i) \geq \text{CU}_i(\alpha_j)$. Hence, for any $\alpha_j > \eta^i$, we have $\text{NU}_i(\alpha_j) > \text{CU}_i(\alpha_j)$. □

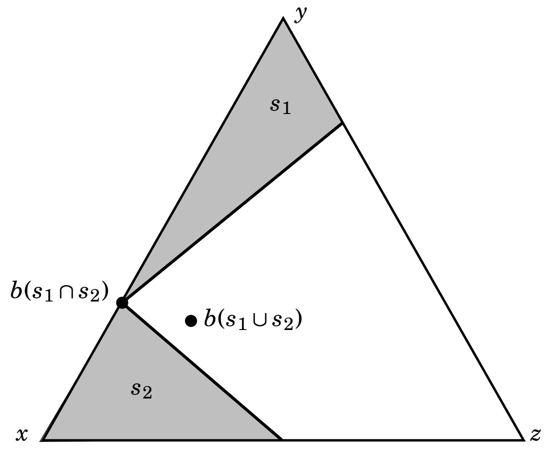


FIGURE 2

APPENDIX B. THE APPROVAL BARGAINING GAME WITHOUT PARTIAL HONESTY

In this appendix we investigate the consequences of dropping the partial honesty assumption so that the players' preference ordering over strategy profiles is the one induced by their preference over the outcomes that they produce. In this version of the game, it is still true that every game has an equilibrium in sincere strategies (the proof of Theorem 1 does not use partial honesty). It is, in general, also true that every equilibrium is consensual.[7] Nonetheless, without partial honesty, players do not necessarily play sincere strategies if there is a strategy that is not sincere and does just as well. This can cause players to coordinate on inefficient outcomes in equilibrium as illustrated in the next example.

Example 3. Figure 2 represents a bargaining game with alternatives x, y, and z. Player 1 has preferences $x \sim_1 y \succ_1 z$ and player 2 has preferences $x \succ_2 z \succ_2 y$. That is, the unique efficient outcome is x.

But the consensual strategy profile (s_1, s_2) indicated in the figure is an equilibrium that generates an inefficient outcome. To see that it is an equilibrium note that player 1 cannot do better because she already obtains her highest possible payoff. Player 2 is clearly playing her best consensual response and, moreover, the non-sincere strategy s_1 has been chosen so that player 2 prefers the outcome $b(s_1 \cap s_2)$ to $b(s_1 \cup s_2)$. Thus, player 2 does not have a profitable deviation either.

To construct an example with three alternatives with an inefficient equilibrium one player must be indifferent between her two top alternatives. Otherwise, the next Proposition establishes that every equilibrium is efficient. Unfortunately, this result does not extend to approval bargaining games with at least four alternatives.

[7] For example, if both players rank x and y first and they are both indifferent between x and y, then there is a non-consensual equilibrium where one player approves $\{x\}$ and the other player approves $\{y\}$. This is the only type of equilibrium ruled out by partial honesty in the proof of Proposition 1.

Proposition 3. *Suppose that the set of alternatives is $X = \{x, y, z\}$ and that players have a strict ordering over alternatives. Every equilibrium outcome is efficient.*

Proof. Consider first the case where both players rank x first. Assume there is an equilibrium (s_1, s_2) with $\theta(s_1, s_2) \neq \{x\}$. If alternative x is approved by s_i, player j can induce x by approving only x, a contradiction. Similarly, if s_i is not full dimensional, then player j does not have a best response (see Example 1) so that (s_1, s_2) is not an equilibrium. Therefore, for each i, $x \notin s_i$ and s_i is full dimensional. It follows that (s_1, s_2) is a consensual strategy profile, repeating the arguments in the proof of Proposition 1.

Furthermore, since players have strict preferences, $s_1 \cap s_2$ cannot be two dimensional. It cannot be one dimensional either. If this was the case, given that player i only approves her most preferred lotteries among those in s_j, players would be playing sincere strategies and have utility functions satisfying $u_1 = -u_2$. Hence, $s_1 \cap s_2$ consists of a single lottery p.

Since utilities are linear, p is in the boundary of the simplex. For each player i we have $u_i(x) > u_i(p) \geq u_i(b(\Delta))$ (see proof of Proposition 2). Therefore, letting $b_{yz} := \frac{1}{2}y + \frac{1}{2}z$, for each player i there is $\gamma_i \in [\frac{1}{3}, 1)$ such that $u_i(\gamma_i x + (1-\gamma_i)b_{yz}) = u_i(p)$. For each $i = 1, 2$, construct the sincere strategy $s_i^* := \{u_i, u_i(p)\}$ and write the equilibrium strategy as $s_i = \{w_i, \alpha_i\}$ where, of course, $\alpha_i = w_i(p)$. Strategy s_i can only intersect with s_j^* on the boundary of s_j^* because otherwise player j would deviate to s_j^*. Hence, either there is a $0 < \bar{\delta}_i \leq \gamma_j$ such that every $0 \leq \delta_i \leq \bar{\delta}_i$ satisfies $\delta_i x + (1-\delta_i)b_{zy} \in s_i$ or every lottery q on the line connecting x and b_{yz} satisfies $w_i(q) < \alpha_i$. In either case, $s_1 \cap s_2$ has points other than p. This provides a contradiction and implies that the unique equilibrium outcome of the game is x.

Suppose now that players do not have a common best alternative and take an equilibrium (s_1, s_2). The proof of Proposition 1 implies that (s_1, s_2) is consensual. As before, $s_1 \cap s_2$ cannot be two dimensional. If $s_1 \cap s_2$ is one dimensional then players are indifferent between any two points in this one dimensional set. Hence, $u_1 = -u_2$ and the equilibrium outcome is efficient as implied by Proposition 2.

Finally, suppose that $s_1 \cap s_2$ consists only of lottery p. Since utilities are linear, p belongs to the boundary of the simplex. And since utilities are strict, if p is a player's best lottery then p must be efficient. So suppose otherwise and consider player i's lower contour set of lottery p, denoted $L_i(p) := \{q \in \Delta : u_i(q) \leq u_i(p)\}$. We have $p \in L_i(p)$ and $b(\Delta) \in L_i(p)$ (see proof of Proposition 2). Since utilities are linear, the segment $\ell(p, b(\Delta)) := \{p - \beta(p - \Delta) : \beta > 0\} \cap \Delta$ satisfies $\ell(p, b(\Delta)) \subset L_i(p)$. Moreover, $\ell(p, b(\Delta))$ divides the simplex in two. Call Δ_i the closure of the half of the simplex that contains s_j. We have $\Delta_i \subset L_i(p)$ because $s_j \subset L_i(p)$ and, therefore, any lottery q such that $u_i(q) > u_i(p)$ belongs to $\Delta \setminus \Delta_i = \text{int}(\Delta_j)$ (where the interior is relative to Δ). But this implies $u_j(q) < u_j(p)$. That is, lottery p is efficient. □

Appendix C. Existence of equilibrium

We build a sequence of finite games that suitably approximate our game Φ. Each game in this sequence is an approval voting game with two players. This class of games is analyzed by Núñez and Laslier (2015). Each player selects a subset of the finite set of alternatives that she approves. If the intersection of these two subsets is nonempty then the outcome is determined by a uniform lottery over the intersection. If the intersection of the two subsets is empty then the outcome is decided by the uniform lottery over the union. We need the following properties proved in Núñez and Laslier (2015).

(1) Every two-player approval voting game has an equilibrium in sincere strategies. That is, an equilibrium where if a player approves some alternative then she also approves every alternative that she prefers to it.
(2) If an equilibrium outcome is non-consensual then each player approves *every* alternative that she prefers to the equilibrium outcome.
(3) In every sincere equilibrium, each player *only* approves alternatives that she prefers to the equilibrium outcome.

As we construct the sequence of finite two-player approval games we also construct a sequence of measures to approximate outcomes in Φ with sequences of outcomes of the approval games.

We embed the $(K-1)$-dimensional simplex Δ in $\mathbb{R}^{K-1}$ and consider the smallest hypercube $I \subset \mathbb{R}^{K-1}$ containing Δ. We construct a sequence of probability measures $\{\lambda^t\}$ on I iteratively. We first set $I^0 := I$ and let c be the barycenter of I^0 and $C^0 := \{c\}$. The probability measure λ^0 gives probability 1 to $c \in I^0$. For each $t > 0$, let I^t be the set of hypercubes that one obtains by dividing each hypercube in I^{t-1} into 2^{K-1} equally sized hypercubes. Each one of the 2^{K-1} hypercubes $h \in I^t$ has a barycenter $c(h)$. Let $C^t := \{c(h) : h \in I^t\}$. The probability measure λ^t gives probability $1/\#C^t$ to each $c(h)$ such that $h \in I^t$. Furthermore, the game Γ^t is defined as the approval voting game with 2-players and set of alternatives $X^t := C^t \cap \Delta$. Player's utilities over elements in X^t are computed by extending linearly their Bernoulli utility function over the original set of alternatives X.

The next lemma is used to approximate outcomes in the game Φ with a sequence of outcomes of the finite approval games constructed above. The proof consists in showing that the sequence of probability measures $\{\lambda^t\}$ converges weakly to the uniform measure $\lambda(\cdot)/\lambda(I)$ over the hypercube I. There are several equivalent definitions of weak convergence but for our purposes we only need two.[8] Given the hypercube I (with its Borel σ-algebra) the bounded sequence of positive finite measures $\{\lambda^t\}$ on I converges weakly to the finite positive measure $\lambda(\cdot)/\lambda(I)$ if any of the following equivalent conditions is true:

[8] See Theorem 25.8 in Billingsley (1986) for equivalent definitions of weak convergence.

- $\lim \lambda^t(E) = \lambda(E)/\lambda(I)$ for every set E whose boundary ∂E satisfies $\lambda(\partial E) = 0$.
- $\lim \int_I f d\lambda^t = \frac{1}{\lambda(I)} \int_I f d\lambda$ for every bounded and uniformly continuous function f.

Lemma 3. *Let $E \subset \Delta$ satisfy $\lambda(E) > 0$ and $\lambda(\partial E) = 0$, and define $E^t := X^t \cap E$. Then*

$$\lim_{t\to\infty} \frac{\sum_{e\in E^t} e}{\#E^t} = \frac{\int_E p d\lambda}{\lambda(E)}.$$

Proof. As we announced previously, we actually prove that the sequence of probability measures $\{\lambda^t\}$ converges weakly to the uniform measure $\lambda(\cdot)/\lambda(I)$ over I. A consequence is that conditional probabilities induced by members of $\{\lambda^t\}$ on subsets $E \subset I$ whose boundary has zero Lebesgue measure also converge to the corresponding uniform probability measures over those subsets (and, hence, also their means).

Take some hypercube $h \in I^t$ and note that, if $c(h)$ is its barycenter, $\lambda^t(c(h)) = 1/\#C^t = \lambda(h)/\lambda(I)$. That is, the probability of $c(h)$ coincides with the volume of h normalized by the volume of I. For any bounded, uniformly continuous function $f : I \to \mathbb{R}$,

$$\int_I f d\lambda^t = \frac{1}{\lambda(I)} \sum_{h\in I^t} f(c(h))\lambda(h) \stackrel{t\to\infty}{\longrightarrow} \frac{1}{\lambda(I)} \int_I f d\lambda,$$

which means that $\{\lambda^t\}$ converges weakly to the measure $\lambda(\cdot)/\lambda(I)$. $\square$

Now we can prove:

Theorem 1. *Every game Φ has an equilibrium in sincere strategies.*

Proof. Given Property 1 we can take a sequence $\{(s_1^t, s_2^t)\}_{t=1}^\infty$ of pairs of finite subsets of Δ such that (s_1^t, s_2^t) is a sincere equilibrium of Γ^t for every t. For $i = 1,2$ and for every t define $\alpha_i^t := \min_{p\in s_i^t} u_i(p)$. The utility to player i from *every* lottery in s_i^t is at least α_i^t. The sequence $\{(\alpha_1^t, \alpha_2^t)\}_{t=1}^\infty$ is contained in a compact set, therefore, it has a subsequence that converges to some (α_1^*, α_2^*). For each $i = 1,2$ define the sincere strategy $s_i^* := \{u_i, \alpha_i^*\}$. We claim that (s_1^*, s_2^*) is an equilibrium of Φ. We proceed in three steps.

Step 1: (s_1^*, s_2^*) induces a consensual outcome.

We prove this step by contradiction. Suppose that $s_1^* \cap s_2^* = \emptyset$. Since $\lim(\alpha_1^t, \alpha_2^t) = (\alpha_1^*, \alpha_2^*)$ continuity of the utility functions on Δ implies that, passing to a subsequence if necessary, for every t high enough we also have $s_1^t \cap s_2^t = \emptyset$. Because (s_1^t, s_2^t) is a non-consensual equilibrium of Γ^t, Property 2 above implies that the strategy s_i^t contains every lottery that player i prefers to $b(s_1^t \cup b_2^t)$. For $i = 1,2$, let $q_i^t := \arg\min_{p_i^t\in s_i^t} \|p_i^t, b(s_1^t \cup s_2^t)\|$ be the lottery approved by player i in the strategy s_i^t that is closest to the outcome $b(s_1^t \cup b_2^t)$. Clearly, for $i = 1,2$, the sequence $\|q_i^t, b(s_1^t \cup s_2^t)\|_{t=1}^\infty$ converges to zero. The triangular inequality implies that the sequence $\|q_1^t, q_2^t\|_{t=0}^\infty$ also converges to zero. This contradicts $s_1^* \cap s_2^* = \emptyset$ proving that (s_1^*, s_2^*) induces a consensual outcome.

Step 2: (s_1^*, s_2^*) *generates expected payoffs* (α_1^*, α_2^*).

To the contrary and without loss of generality, assume that player 1 gets a payoff strictly higher than α_1^* under the strategy profile (s_1^*, s_2^*) so that $u_1(b(s_1^* \cap s_2^*)) > \alpha_1^*$. There is a $\hat{p} \in s_2^*$ such that $u_1(\hat{p}) > \alpha_1^*$. Such an inequality also holds for every point in some closed neighborhood $\mathscr{P}$ of $\hat{p}$. Thus, for t high enough, we can choose a $\hat{p}^t \in S^t \cap \mathscr{P}$ such that $u_1(\hat{p}^t) > \alpha_1^*$ and $\hat{p}^t \in \text{int}(s_2^*)$ (i.e. $u_2(\hat{p}^t) > \alpha_2^*$). This means that $\hat{p}^t \in s_2^t$ for sufficiently high t. Therefore, $u_1(\theta(s_1^t, s_2^t)) \geq u_1(\hat{p}^t)$ for any sincere equilibrium (s_1^t, s_2^t) of Γ^t. But then, also for every sufficiently high t,

$$v_1^t \geq u_1(\theta(s_1^t, s_2^t)) \geq u_1(\hat{p}^t) > \alpha_1^*, \tag{A.1}$$

where the first inequality follows from Property 3. But this is impossible because α_1^* is the limit point of the sequence $\{\alpha_1^t\}_{t=1}^{\infty}$. This provides a contradiction so we can conclude that (s_1^*, s_2^*) generates expected payoffs (α_1^*, α_2^*).

Step 3: (s_1^*, s_2^*) *is an equilibrium.*

Suppose again, on the contrary, that (s_1^*, s_2^*) is not an equilibrium of Φ. Without loss of generality, let there be an $\hat{s}_1$ such that $u_1(\theta(\hat{s}_1, s_2^*)) > v_1^*$. The fact that (s_1^*, s_2^*) induces the consensual outcome $\theta(s_1^*, s_2^*)$ that generates the vector of utility levels (α_1^*, α_2^*), implies that player 1's deviation to $\hat{s}_1$ induces a non-consensual outcome $b(\hat{s}_1 \cup s_2^*)$. For each t, consider the strategy $\hat{s}_1^t$ that approves every lottery available in Γ^t that belongs to $\hat{s}_1$. By construction, the outcome $\theta(\hat{s}_1^t, s_2^t)$ is non-consensual and Lemma 3 guarantees that $\lim b(\hat{s}_1^t \cup s_2^t) = b(\hat{s}_1 \cup s_2^*)$. Hence, for every t high enough and some $\varepsilon > 0$ we obtain

$$u_1(b(\hat{s}_1^t \cup s_2^t)) > \alpha_1^* + \varepsilon. \tag{A.2}$$

Since each member of the sequence $\{(s_1^t, s_2^t)\}_{t=0}^{\infty}$ is an equilibrium of the corresponding game Γ^t, Property 3 implies that $u_1(s_1^t \cup s_2^t) \leq \alpha_1^t$ for every t. It follows that $\lim u_1(b(s_1^t \cup s_2^t)) \leq \alpha_1^*$ and, for every t high enough, $u_1(s_1^t \cup s_2^t) \leq \alpha_1^* + \varepsilon$. But this last inequality combined with (A.2) implies that (s_1^t, s_2^t) is not an equilibrium of Γ^t. This is a contradiction so (s_1^*, s_2^*) is an equilibrium of Φ. □

References

Y. Babichenko and L.J. Schulman. One-shot bargaining mechanisms. *arXiv preprint arXiv:1502.05238*, 2015.

M. Bagnoli and B. Lipman. Provision of Public Goods: Fully Implementing the Core through Private Contributions. *Review of Economic Studies*, 56:583–601, 1989.

P. Billingsley. *Probability and Measure*. Wiley, 1986.

S.J. Brams. *Mathematics and Democracy: Designing Better Voting and Fair-Division Procedures*. Princeton, NJ: Princeton University Press., 2008.

S.J. Brams and P.C. Fishburn. *Approval Voting*. Birkhauser, Boston, 1983.

J. Busetto and R. Colognato. Reconsidering Two-Agent Nash Implementation. *Social Choice and Welfare,*, 32:171–179, 2009.

B. Dutta and A. Sen. A Necessary and Sufficient Condition for Two-person Nash Implementation. *Review of Economic Studies*, 58:121–128, 1991.

B. Dutta and A. Sen. Nash Implementation with Partially Honest Individuals. *Games and Economic Behavior*, 74:154–169, 2012.

M. O. Jackson. Implementation in undominated strategies: A look at bounded mechanisms. *The Review of Economic Studies*, 59(4):757–775, 1992.

N. Kartik and O. Tercieux. Implementation with Evidence. *Theoretical Economics*, 7:323–355, 2012.

N. Kartik, O. Tercieux, and R. Holden. Simple mechanisms and preferences for honesty. *Games and Economic Behavior*, 83(0):284 – 290, 2014.

J.F. Laslier and R. Sanver. *Handbook on Approval Voting*. Heildelberg: Springer-Verlag, 2010.

E. Maskin. Nash Equilibrium and Welfare Optimality. *Review of Economic Studies*, 66: 23–38, 1999.

H. Matsushima. Role of honesty in full implementation. *Journal of Economic Theory*, 139: 353–359, 2008a.

H. Matsushima. Behavioral aspects of implementation theory. *Economics Letters*, 100(1): 161 – 164, 2008b.

S. Merill and J. Nagel. The Effect of Approval Balloting on Strategic Voting Under Alternative Decision Rules. *American Political Science Review*, 81:509–524, 1987.

J. Moore and R. Repullo. Nash Implementation: A Full Characterization. *Econometrica,*, 58:1083–1099, 1990.

J. Nash. Two-person cooperative games. *Econometrica*, 21:128–140, 1953.

M. Núñez. The Strategic Sincerity of Approval Voting. *Economic Theory*, 56(1):157–189, 2014.

M. Núñez and J.-F. Laslier. Bargaining through Approval. *Journal of Mathematical Economics*, 60:63–73, 2015.

M. Núñez. Threshold voting leads to type-revelation. *Economics Letters*, 136:211–213, 2015.

J. Ortner. Direct implementation with minimally honest individuals. *Games and Economic Behavior,*, 90:1–16, 2015.

M. Osborne and A. Rubinstein. *Bargaining and Markets*. Academic Press, 1990.

R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

A. Rubinstein. Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50:97–109, 1982.

R. Serrano. Bargaining. In S. Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics*. McMillan: London, 2008.