UNSW Business School Research Paper No. 2015 ECON 8

Reaching Consensus Through Simultaneous Bargaining

Jean-François Laslier
Matías Núñez
Carlos Pimienta

REACHING CONSENSUS THROUGH SIMULTANEOUS BARGAINING*

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

ABSTRACT. We propose a two-player bargaining game where each player simultaneously proposes a set of lotteries on a finite set of alternatives. If the two sets have elements in common the outcome is selected by the uniform probability measure over the intersection. If otherwise the sets do not intersect the outcome is selected by the uniform probability measure over the union. We show that this game always has an equilibrium in sincere strategies (i.e. such that players truthfully reveal their preferences). We also prove that every equilibrium is individually rational and consensual. If furthermore players are partially honest then every equilibrium is efficient and sincere. We use this result to fully characterize the set of equilibria of the game under partial honesty.

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

JEL CLASSIFICATION. C70, C72.

1. INTRODUCTION

We consider a bargaining situation in which two agents have to jointly make a selection out of a finite set of alternatives. Agents have complete information about the preferences of the other party and transfers are not possible. Osborne and Rubinstein (1990) argue that the outcome of this interaction should be, at least, Pareto optimal and individually rational: there should be no other

* We thank Don Keenan, Marcus Pivato and Dimitrios Xefteris for useful comments. Matías thanks financial support from the LABEX MME-DII (ANR-11-LBX-0023-01). Carlos thanks financial support from the Australian Research Council's Discovery Projects funding scheme DP140102426. The usual disclaimer applies.

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

[b] CNRS & THEMA, UNIVERSITY OF CERGY-PONTOISE, FRANCE.

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

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

Date: April 17, 2015.

outcome that they both prefer to the equilibrium outcome and the equilibrium outcome should not be worse than disagreement.[1,2]

More broadly and from a mechanism design perspective, Dutta and Sen (1991) show that if there are two players and a finite number of alternatives one can implement in Nash equilibrium the set of Pareto efficient and individually rational lotteries using an integer game under a suitable domain restriction.[3] The unappealing features of integer games has stimulated researchers to investigate the implementation problem using different approaches. A recent one explores the scope for implementation when players are *partially honest* (for a very incomplete list, see Matsushima, 2008a,b; Dutta and Sen, 2012; Kartik and Tercieux, 2012; Kartik et al., 2014). Under partial honesty, a player prefers a truthful message when it does not lead to a strictly worst outcome than what she would obtain otherwise. However, existing results do not necessarily 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).

In contrast, Núñez and Laslier (2014) show that *approval voting* (Brams and Fishburn, 1983; Laslier and Sanver, 2010) can be reinterpreted as a bargaining protocol when there are only two voters. It is a one-shot bargaining game where each agent approves a subset of the finite set of alternatives. If the two sets intersect then the final outcome is selected using the uniform lottery over the intersection. If the two sets do not intersect then the final outcome is selected using the uniform lottery over the union. Núñez and Laslier (2014) show that every equilibrium outcome of this game is individually rational and, if players are partially honest, not Pareto dominated by any pure alternative. The definition of sincerity in this setting is borrowed from the approval voting literature (Brams and Fishburn, 1983).[4] A strategy is sincere

[1] The precise outcome to be implemented typically depends on the particular bargaining protocol used. For instance, in Rubinstein (1982)'s seminal contribution, the equilibrium outcome depends on the players' discount factor given the dynamic setting of his model.

[2] The strategic bargaining literature is vast and we do not attempt here to give a full review. We refer the reader to Serrano (2008) for an excellent review

[3] Dutta and Sen (1991) derive a necessary and sufficient condition on the social choice correspondence for two-player implementation problems. They use this property to find the appropriate domain restriction. In particular, they assume that players have strict preferences and that there is no affine transformation of their utility functions u_1, u_2 that satisfies $u_1 = -u_2$.

[4] See Merill and Nagel (1987), Brams (2008), and Núñez (2014) for works dealing with sincerity under approval voting.

if whenever it approves one alternative it also approves every alternative that the agent prefers.

This paper departs from Núñez and Laslier (2014) and considers a one-shot game in which each player selects *a set of lotteries over the pure alternatives*. If the two sets intersect then the final outcome is selected using the uniform probability measure over the intersection. If the two sets do not intersect then the final outcome is selected using the uniform probability measure over the union. We show that this game retains every desirable property in Núñez and Laslier (2014) while also obtaining full efficiency when players are partially honest. Again, we borrow the definition of sincerity form the approval voting literature and say that a strategy is sincere if whenever it includes some lottery it also includes every lottery that is at least as good.

In some sense, the current game is similar to Nash (1953) demand game. In the demand game, two players make simultaneous demands and each one 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. For example, consider Figure 1. We consider a bargaining situation with three alternatives, each one represented by a degenerate lottery at each vertex of the simplex. The figure to the left corresponds to the strategy profile (s_1,s_2) while the figure to the left corresponds to a situation where players play the strategy profile (s'_1,s'_2). Under the strategy profile (s_1,s_2), Player 1 approves every lottery in the closed subset labeled s_1 and Player 2 approves every lottery in the closed subset labeled s_2. These two strategies do not intersect, thus we say that (s_1,s_2) is a *non-consensual* strategy profile. The outcome induced by this strategy profile is the uniform probability measure on $s_1 \cup s_2$ and the expected outcome of such a measure is the barycenter $b(s_1 \cup s_2)$ of the surface formed by the union of these two strategies. This figure suggests that, under a non-consensual strategy profile players have two joint incentives: (1) approving a large set so that the induced expected outcome is as close as possible to their approved sets, and (2) using the sincere strategy that approves every lottery in the upper-contour set of some indifference curve. Note this is the most effective way to obtain a more preferred outcome given the strategy of the opponent.

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. We prove that every equilibrium strategy profile must have a non-empty intersection in the same way as (s'_1,s'_2) in the right hand side of Figure 1. We say that (s'_1,s'_2) is a *consensual* strategy profile. In this particular example, the intersection is a singleton so that such a singleton is

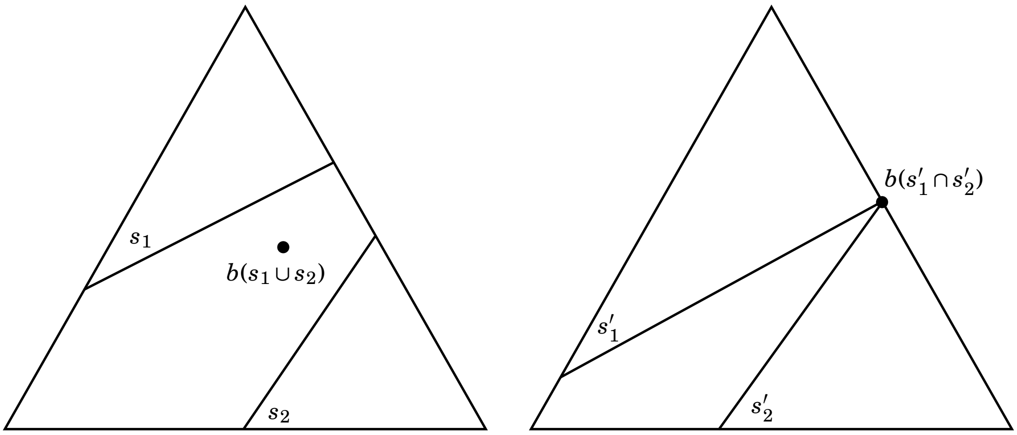

FIGURE 1. A non-consensual (left) and a consensual (right) strategy profile.

also the expected outcome induced by uniform probability measure over the intersection. Such an outcome is denoted $b(s'_1, s'_2)$. Note that players still have to approve a large set of lotteries as this will prevent the other player from deviating to a non-consensual strategy.

However, under a consensual strategy profile, a player may be playing a best response that is not necessarily a sincere strategy. The reason is that, under a consensual strategy profile, only the intersection matters for the outcome and many different sets have the same intersection. Nonetheless, it is not difficult to prove that the set of consensual best responses always includes a sincere strategy. Thus, a players that is partially honest always plays a sincere strategy in equilibrium.

This implies that players have a somewhat natural way of playing this game. A player cannot do better than choosing an approved set of the form $\{p \in \Delta \mid U(p) \geq v\}$ where U denotes the player's *true* expected utility function and v represents the minimal level of utility of the approved lotteries. When playing this strategy, the player fully reveals her utility function by announcing one indifference curve and approving every lottery to the side of the indifference curve where her utility increases.

Building on the previous remarks, we prove that this approval bargaining game has the following features.

(1) *Existence of Equilibrium*: Every game admits an equilibrium in *sincere* strategies. This game has an unusually complex strategy space and, moreover, it is discontinuous (the outcome changes discontinuously when the strategy profile moves from consensual to non-consensual). Hence, standard equilibrium existence results do not apply.

(2) *Sincere Best Responses*: A player has a sincere best response for any strategy of his opponent. Hence, if she is partially honest she always plays sincerely.
(3) *Individual Rationality*: In every equilibrium, a player obtains at least the same utility as from the uniform lottery over the set of alternatives.
(4) *Consensual Equilibria*: Every equilibrium must be consensual, that is, both players agree on some subset of lotteries. Note that this consensus occurs in equilibrium and depends on the players' beliefs on the consequences of not reaching the agreement. As pointed out by Baron and Ferejohn (1989), "bilateral exchange requires unanimous consent for an outcome, and this requirement gives each party veto power that is reflected in the equilibrium outcomes". This veto power is absent from our model; see Banks and Duggan (2000) for related bargaining models with complete information.
(5) *Welfare and Partial Honesty*: Every sincere equilibrium is efficient. Thus, if players are partially honest, every equilibrium is efficient. Without partial honesty players may coordinate in insincere strategies and induce an inefficient outcome. Moreover, also under partial honesty, we can fully characterize the set of equilibria—it is homeomorphic to a closed interval whose endpoints are the maximum and minimum utility levels that one of the players can obtain in equilibrium.

The rest of the paper is structured as follows. Section 2 presents the model and Section 3 describes the players' best responses. Section 4 discusses the game under the special case that both players agree on what the best alternative is. The general situation is analyzed in Sections 5 and 6. The latter section focuses on efficiency and partial honesty. The proof of existence of equilibria is contained in the Appendix.

2. The Game

Consider two *players* indexed by $i = 1,2$ and a finite set of *alternatives* $X \equiv \{x_1, x_2, ..., x_K\}$ with at least two elements. Each Player i is endowed with a Bernoulli utility function $u_i \in \mathbb{R}^X$. To only consider interesting cases we assume that a Player's best and a worst alternative are associated to different utility levels. Let $\Delta \equiv \{p \in \mathbb{R}_+^K \mid \sum p_i = 1\}$ denote the probability simplex over X. Furthermore, we identify an alternative $x \in X$ with the degenerate lottery that assigns probability one to x. Let $U_i : \Delta \to \mathbb{R}$ be Player i's corresponding expected utility function.

As mentioned in the Introduction, a strategy for Player i is a subset of lotteries in Δ that the player approves. If the strategies played by the two players have a nonempty intersection then the outcome of the game is decided by the uniform probability measure over the intersection. If otherwise the strategies do not intersect then the outcome is the realization the uniform probability measure over the union. Therefore, we cannot allow players to play "exotic" subsets of Δ where the uniform probability measure cannot be defined.[5]

We let S be the collection of all sets that can be written as the finite union of (not necessarily disjoint) convex and closed (thus compact and Lebesgue measurable) subsets of Δ. The collection of sets S is closed under finite union and finite intersection.[6] Lemma 1 below shows that if S is the strategy space of both players then the game is well-defined.[7]

We give two examples of strategies $s_i \in S$.

Example 1 (Approving alternatives)**.** Player i can choose a strategy $s_i \in S$ that approves a subset of the set of alternatives, that is, some $s_i \subseteq X$. Any such set s_i can be expressed as a finite union of singletons and, therefore, it is compact and convex. Note that these strategies coincide with those allowed under standard Approval voting.

Example 2 (Approving a half space)**.** Player i can choose a strategy $s_i \in S$ that contains every lottery that, for some utility function $\hat{u}_i$, gives at least some level of expected utility. For example, if $s_i = \{p \in \Delta \mid \hat{u}_i \cdot p \geq v\}$ for some $v \in \mathbb{R}$ and $\hat{u}_i$ coincides with Player i's true utility function, then she approves every lottery in the corresponding upper contour set associated with the utility level v.

A particular case of the strategies given in Example 2 is the collection of *sincere strategies*. Following the literature on approval voting, see Brams and Fishburn (1983), we say that a strategy of Player i that approves *every* lottery that gives him a utility above some certain threshold is sincere.

Definition 1 (Sincerity)**.** A strategy $s_i \in S$ is sincere for Player i if

$$p \in s_i \text{ and } U_i(q) \geq U_i(p) \text{ implies } q \in s_i.$$

[5] Not every compact metric space admits a uniform probability measure (see Dembski, 1990).

[6] Let $A, B \in S$. If $A \cap B \neq \emptyset$ then this intersection can also be written as the finite union of closed and convex subsets of Δ because the intersection of two closed and convex sets is also closed and convex. If $A \cap B = \emptyset$ the same is also true because the empty set is already closed and convex.

[7] For the model to be well-defined we need that the pairwise union and intersection of any two strategies admit a uniform distribution. Our restriction of the strategy space is sufficient, but it is not the largest collection of subsets of Δ satisfying this property.

Given a convex subset $A \subset \Delta$, 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. The dimension of a finite union of convex sets $\bigcup_{z\in Z} A_z$ is equal to $\max_{z\in Z} \dim(A_z)$ (see Rockafellar, 1997). Let λ_n be the Lebesgue measure in $\mathbb{R}^n$. For any n-dimensional set $A \in S$, the uniform measure with support A is given by $\mu(\cdot \mid A) = \lambda_n(\cdot)/\lambda_n(A)$. Hence, the barycenter $b(A)$ of A is

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

Since we work in the probability simplex over X, we will often refer to λ_{K-1}. For simplicity, we simply write λ instead of λ_{K-1}.

Given a strategy profile $s = (s_1, s_2) \in S$, the winning set, to be denoted $s_1 \otimes s_2$, is equal to:

$$s_1 \otimes s_2 \equiv \begin{cases} s_1 \cap s_2 & \text{if} \quad s_1 \cap s_2 \neq \emptyset, \\ s_1 \cup s_2 & \text{otherwise.} \end{cases}$$

If $s_1 \cap s_2 \neq \emptyset$ then the strategy profile s is *consensual*. If otherwise $s_1 \cap s_2 = \emptyset$ then the strategy profile s is *non-consensual*. Ties are broken randomly so that, given the strategy profile $s = (s_1, s_2)$, the expected outcome is $b(s_1 \otimes s_2)$.

The rules described above define the simultaneous game $\Phi = (S, S, u_1, u_2)$. With abuse of notation, for any $A \in S$, we write $U_i(A)$ instead of $U_i(b(A))$. The following lemma implies that the game Φ is well defined.

Lemma 1. *For any* $(s_1, s_2) \in S$*, the point* $b(s_1 \otimes s_2)$ *always exists and belongs to* Δ*.*

Proof. We already argued that S is closed under finite union and finite intersection. Furthermore, any $A \in S$ has a well defined dimension so that, for any strategy profile (s_1, s_2), the measure $\mu(\cdot \mid s_1 \otimes s_2)$ is well defined. Finally, since the convex hull of the support of $\mu(\cdot \mid s_1 \otimes s_2)$ is always a subset of Δ we have $b(s_1 \otimes s_2) \in \Delta$. □

Definition 2 (Equilibrium)**.** A strategy profile $s = (s_1, s_2)$ is an equilibrium if, for every Player i and every $s_i' \in S$, we have $U_i(s_i \otimes s_{-i}) \geq U_i(s_i' \otimes s_{-i})$.

3. Best Response Analysis

Player i's set of best responses against strategy $s_j \in S$ is

$$\mathrm{BR}_i(s_j) \equiv \arg\max_{s_i \in S} U_i(s_i \otimes s_j).$$[8]

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

Given the rules of the game, a best-response $s_i \in \mathrm{BR}_i(s_j)$ can either be consensual (if $s_i \cap s_j \neq \emptyset$) or non-consensual (if $s_i \cap s_j = \emptyset$). We begin analyzing consensual best responses to s_j. That is, those $s_i \in \mathrm{BR}_i(s_j)$ that satisfy $s_i \cap s_j \neq \emptyset$. These strategies can be thought of as "accepting" a subset of lotteries offered in s_j. Hence, in a consensual best response, Player i should "accept" only her most preferred lotteries in s_j. This implies that every accepted lottery must lead to the same utility level and that, therefore, the set of accepted lotteries has zero λ-measure.[9]

For any strategy $s_j \in S$, we let $T_i(s_j) \equiv \arg\max_{p \in s_j} U_i(p)$ denote the set of most preferred lotteries by Player i in s_j.

Lemma 2. *Let $s_i \in S$ be a consensual best-response to strategy $s_j \in S$. Then*

$$\mu(s_i \cap T_i(s_j) \mid s_i \cap s_j) = 1.$$

Proof. Assume by contradiction that there is some consensual best-response s_i to s_j with $\mu(s_i \cap T_i(s_j) \mid s_i \cap s_j) < 1$. Note that any $p \in s_i \cap T_i(s_j)$ satisfies $U_i(p) = \bar{V}_i$ whereas $U_i(p) < \bar{V}_i$ for any $p \in s_i \setminus T_i(s_j)$. Then,

$$\begin{aligned} U_i(s_i \cap s_j) &= \int_{s_i \cap s_j} U_i(p) d\mu(p \mid s_i \cap s_j) \\ &= \int_{s_i \cap T_i(s_j)} U_i(p) d\mu(p \mid s_i \cap s_j) + \int_{s_i \cap (s_j \setminus T_i(s_j))} U_i(p) d\mu(p \mid s_i \cap s_j) \\ &= \bar{V}_i \mu(s_i \cap T_i(s_j) \mid s_i \cap s_j) + \int_{s_i \cap (s_j \setminus T_i(s_j))} U_i(p) d\mu(p \mid s_i \cap s_j). \end{aligned}$$

Since $\mu(s_i \cap T_i(s_j) \mid s_i \cap s_j) < 1$ and $U_i(p) < \bar{V}_i$ for any $p \in s_i \setminus T_i(s_j)$, it follows that $U_i(s_i \cap s_j) < \bar{V}_i = U_i(T_i(s_j) \cap s_j)$. Therefore, s_i is not a consensual best response to s_j which provides the desired contradiction. □

Even if there is no best response that is consensual, there always is a *best consensual response*. Indeed, no other consensual response to s_j does better than the consensual response $T_i(s_j)$. The same property does not hold for non-consensual responses. The next example shows a situation where not only does Player 1 not have a *best non-consensual response* but also she does not have a best response overall.

Example 3. Let players 1 and 2 have strict preferences and let x_1 be Players 1's most preferred alternative. Take Player 2's strategy to be $s_2 = \{x_2\}$ so that $\lambda(s_2) = 0$.

Any consensual best response to s_2 by Player 1 includes x_2 and, hence, generates utility level $u_1(x_2)$. As far as non-consensual responses are concerned,

[9] Recall that we assumed that each player has a worst and a best alternatives so that indifference curves are lower-dimensional hyperplanes.

for any $\varepsilon > 0$ small enough, the sincere strategy $s_1^\varepsilon = \{p \in \Delta \mid U_1(p) \geq u_1(x_1) - \varepsilon\}$ generates expected utility

$$U_1(s_1^\varepsilon \otimes s_2) = U_1(s_1^\varepsilon \cup s_2) = U_1(s_1^\varepsilon),$$

where the last equality follows from $\lambda(s_2) = 0$. Hence, $U_1(s_1^\varepsilon)$ gets arbitrarily close to $u_1(x_1)$ as ε decreases. When $\varepsilon = 0$, the strategy s_1^ε collapses to $\{x_1\}$ so that $U_1(s_1^0 \otimes s_2) = \frac{1}{2}u_1(x_1) + \frac{1}{2}u_1(x_2) < U_1(s_1^\varepsilon \otimes s_2)$ for any $\varepsilon > 0$ small enough. Therefore, Player 1 has no best response to s_2.

In the example, it is critical that Player 2 is playing a lower-dimensional strategy. We will later prove that if s_j is a full-dimensional strategy then Player i always has a best response against s_j. In the meantime, we simply show that if s_j is full-dimensional and s_i happens to be a non-consensual best response against s_j then s_i approves every lottery that Player i prefers to the expected outcome of the strategy profile (s_i, s_j). For any pair of strategies $s_i \in S$ and $s_j \in S$, let $R_i(s_i, s_j) \equiv \{p \in \Delta \mid U_i(p) \geq U_i(s_i \cup s_j)\}$ be the set of lotteries Player i prefers to $b(s_i \cup s_j)$.

Lemma 3. *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*

$$R_i(s_i, s_j) \subseteq s_i \text{ and } \mu(R_i(s_i, s_j) \mid s_i) = 1.$$

Proof. We first prove that if s_i is a non-consensual best response to s_j then $R_i(s_i, s_j)$ is a subset of s_i. The set $R_i(s_i, s_j)$ coincides with the closure of its interior and s_i is a closed set, so it is enough to prove that every point $p \in \operatorname{int}(R_i(s_i, s_j))$ belongs to s_i. Assume to the contrary that $p \notin s_i$. In that case, there is a closed ball B centred at p such that $B \subset \operatorname{int}(R_i(s_i, s_j))$ and $B \cap s_i = \varnothing$. Note that $U_i(B) > U_i(s_i \cup s_j)$ and that, consequently, $B \cap s_j = \varnothing$ because otherwise B would be a better response to s_j than s_i.

Now consider the expected utility of $s_i \cup B$ against strategy s_j which is equal to:

$$\begin{aligned}
U_i(s_i \cup B, s_j) &= \frac{1}{\lambda(s_i \cup B \cup s_j)} \left[\int_{s_i} U_i(p) d\lambda + \int_B U_i(p) d\lambda + \int_{s_j} U_i(p) d\lambda \right] \\
&= \frac{\lambda(s_i \cup s_j)}{\lambda(s_i \cup B \cup s_j)} U_i(s_i \cup s_j) + \frac{1}{\lambda(s_i \cup B \cup s_j)} \int_B U_i(p) d\lambda \\
&> \frac{\lambda(s_i \cup s_j)}{\lambda(s_i \cup B \cup s_j)} U_i(s_i \cup s_j) + \frac{\lambda(B)}{\lambda(s_i \cup B \cup s_j)} U_i(s_i \cup s_j) \\
&= U_i(s_i \cup s_j),
\end{aligned}$$

where the strict inequality follows from $U_i(B) > U_i(s_i \cup s_j)$. Therefore, s_i is not a best response to s_j providing the desired contradiction.

We now prove that $\mu(R_i(s_i,s_j)\mid s_i)=1$. Suppose otherwise that the set $A\equiv s_i\backslash R_i(s_i,s_j)$ has positive measure. Note that the definition of $R_i(s_i,s_j)$ implies that $U_i(A)<U_i(s_i\cup s_j)$. Let $s_i'\equiv R_i(s_i,s_j)$. Then,

$$\begin{aligned}
U_i(s_i\cup s_j)=U_i(s_i'\cup A,s_j)&=\frac{\lambda(s_i'\cup s_j)}{\lambda(s_i'\cup A\cup s_j)}U_i(s_i'\cup s_j)+\frac{1}{\lambda(s_i'\cup A\cup s_j)}\int_A U_i(p)d\lambda\\
&<\frac{\lambda(s_i'\cup s_j)}{\lambda(s_i'\cup A\cup s_j)}U_i(s_i'\cup s_j)+\frac{\lambda(A)}{\lambda(s_i'\cup A\cup s_j)}U_i(s_i'\cup s_j)\\
&=U_i(s_i'\cup s_j).
\end{aligned}$$

Thus, s_i is not a best response against s_j, which provides the desired contradiction and concludes the proof. □

A consequence of the description of best responses given in Lemmas 2 and 3 is that players have a weak incentive to use sincere strategies, that is, to approve the set of lotteries that give her at least some utility level (see Definition 1).

Corollary 1. *If the set of best responses is non-empty then it always includes a sincere strategy.*

Proof. If Player i's has a consensual best response to s_j then, by Lemma 2 the strategy $\{p\in\Delta\mid U_i(p)\geq U_i(q)\text{ for any }q\in T_i(s_j)\}$ is a sincere best response to s_j. On the other hand, if Player i has a non-consensual best response to s_j then by Lemma 3 the strategy s_i that satisfies $s_i=R_i(s_i,s_j)$ is also a sincere best response to s_j. □

A second consequence of the description of best responses is the following.

Corollary 2. *The set of best responses cannot include both consensual and non-consensual strategies.*

Proof. Assume that Player i's set of best responses to s_j contains both consensual and non-consensual strategies. Due to the same argument as in the proof of the previous corollary, the strategy $\{p\in\Delta\mid U_i(p)\geq U_i(q)\text{ for any }q\in T_i(s_j)\}$ and the strategy s_i that satisfies $s_i=\{p\in\Delta\mid U_i(p)\geq U_i(s_i\cup s_j)\}$ are also best responses to s_j. However, both of them must lead to the same utility level so that $U_i(s_i\cup s_j)=U_i(q)$ for any $q\in T_i(s_j)$. In other words, they are both the same strategy. Such a strategy either intersects with s_j or it does not. In the first case, the set of best responses against s_j contains only consensual responses to s_j and, in the second case, it contains only non-consensual responses. □

We conclude this section by proving that players always have a best response against a full-dimensional strategy. To facilitate the analysis, for every strategy profile (s_i, s_j) and each Player i we define the function

$$V_i(s_i, s_j) = \frac{\lambda(s_i)}{\lambda(s_i) + \lambda(s_j)} U_i(s_i) + \frac{\lambda(s_j)}{\lambda(s_i) + \lambda(s_j)} U_i(s_j). \tag{3.1}$$

In particular, $V_i(s_i, s_j) = U_i(s_i \otimes s_j) = U_i(s_i \cup s_j)$ whenever $s_i \cap s_j = \emptyset$. A similar argument to the one used in the proof of Lemma 3 shows that, for every full-dimensional strategy s_j, the *unique* sincere strategy that maximizes $V_i(\cdot, s_j)$ is the strategy s_i that satisfies:

$$s_i = \{p \in \Delta \mid U_i(p) \geq V_i(s_i \cup s_j)\}.$$

The next lemma describes the conditions under which the best responses to a full-dimensional strategy are either consensual or non-consensual.

Lemma 4. *Let $s_j \in S$ be a full-dimensional strategy. Let $s_i \in S$ be the unique sincere strategy that maximizes $V_i(\cdot, s_j)$.*

(1) *If $s_i \cap s_j \neq \emptyset$ then the Player i's best response to s_j is consensual.*
(2) *If otherwise $s_i \cap s_j = \emptyset$ then Player i's best response to s_j is non-consensual and, moreover, s_i is a best response to s_j.*

Proof. (1) For every non-consensual response s_i' to s_j we have $V_i(s_i, s_j) \geq V_i(s_i', s_j) = U_i(s_i' \cup s_j)$. By definition, s_i approves every lottery that gives Player i a utility larger than $V_i(s_i, s_j)$. Since $s_i \cap s_j \neq \emptyset$, the strategy s_i includes $T_i(s_j)$. But then, $U_i(T_i(s_j) \cap s_j) = U_i(T_i(s_j)) \geq V_i(s_i, s_j)$. Thus, for every non-consensual strategy s_i' we find that the consensual strategy $T_i(s_j)$ satisfies $U_i(T_i(s_j) \cap s_j) \geq U_i(s_i' \cup s_j)$. We conclude that the best response to s_j is consensual.

(2) Since $s_i \cap s_j = \emptyset$ we have $U_i(s_i \cup s_j) = V_i(s_i, s_j)$. Furthermore, the fact that s_i maximizes $V_i(s_i, s_j)$ implies that for every non-consensual reply s_i' to s_j we obtain $U_i(s_i \cup s_j) \geq U_i(s_i' \cup s_j)$. Note that s_i approves every lottery that Player i prefers to $b(s_i \cup s_j)$. Therefore, $U_i(p) \leq U_i(s_i \cup s_j)$ for every $p \in T_i(s_j)$. This implies that for every consensual response s_i'' to s_j we have $U_i(s_i \cup s_j) \geq U_i(s_i'' \cap s_j)$. We conclude that s_i is a best response to s_j. □

As a corollary of the lemma 4, we obtain the following result.

Theorem 1. *If $s_j \in S$ is a full-dimensional strategy then* $\mathrm{BR}_i(s_j)$ *is nonempty.*

Note, however, that players may not play a full-dimensional strategy. Nonetheless, as long as players do not agree on what the best alternative is, there is some incentive to do so. In other words, if Player i's strategy s_i is of lower dimension than Player j's strategy and the strategy profile $s = (s_i, s_j)$ is nonconsensual then Player i's strategy has zero measure with respect to the uniform probability measure on $s_i \cup s_j$. This implies that s_i is absent when computing the outcome induced by the strategy profile so that $b(s_i \otimes s_j) = b(s_j)$.

4. Unanimous Best Alternative

We start by considering the simple case where one alternative is ranked as the best one for both players. From a normative viewpoint, the game should be able to facilitate the agreement on that alternative. We establish this result in the next proposition.

For each Player i, the set $B_i \equiv \{p \in \Delta \mid U_i(p) \geq U_i(q) \text{ for any } q \in \Delta\}$ is the set of Player i's most preferred lotteries. We say that players have a unanimous best alternative if $B_i \cap B_j \neq \emptyset$. In the particular case that preferences are strict then $B_i \cap B_j = \{x\}$ for some $x \in X$.

Proposition 1. *If players have a unanimous best alternative then*

(1) *the game has an undominated equilibrium, and*
(2) *both players obtain their maximum utilities in any undominated equilibrium of the game.*

Proof. To prove (1) we note that the strategy profile (B_1, B_2) is an equilibrium because both players obtain their maximum possible payoff in the game. Furthermore, B_i is undominated for Player i because it does strictly better than any alternative strategy s_i against any full dimensional strategy that either intersects with B_i but not with s_i or intersects with s_i but not with B_i.

To show (2) we first prove that any strategy s_i is (weakly) dominated by $s_i' \equiv s_i \cup B_i$ as long as $s_i \neq s_i'$. There are three cases to consider.

(1) If $s_i \cap s_j \neq \emptyset$ then $s_i' \cap s_j \neq \emptyset$ because $s_i \subset s_i'$. But $s_j \cap s_i'$ may contain lotteries that Player i prefers to any lottery in $s_j \cap s_i$, so $U_i(s_i' \otimes s_j) \geq U_i(s_i \otimes s_j)$.
(2) If $s_i \cap s_j = \emptyset$ and $s_i' \cap s_j = \emptyset$ then we also have $U_i(s_i' \otimes s_j) \geq U_i(s_i \otimes s_j)$ for the same reason as before.
(3) If $s_i \cap s_j = \emptyset$ and $s_i' \cap s_j \neq \emptyset$ then $(s_i' \cap s_j) \subset (s_i \cup s_j)$. Furthermore, any lottery $q \in s_i \cup s_j$ such that $q \notin s_i' \cap s_j$ satisfies $U_i(q) < U_i(p)$ for any $p \in B_i$. Hence $U_i(s_i' \otimes s_j) \geq U_i(s_i \otimes s_j)$.

It follows that every undominated strategy of Player i contains every lottery in B_i.

Consider an undominated equilibrium (s_1, s_2). We have $s_1 \cap s_2 \subset B_1 \cap B_2 \neq \emptyset$. If there is some positive measure set $A \subset s_1 \cap s_2$ with $U_i(q) < U_i(p)$ for any $q \in A$ and $p \in B_i$ then Player i is not playing a best response. Therefore, each player obtains his maximum utility in the game. □

The only remaining task is to analyze equilibrium behavior when the two players disagree on what the best alternative is. Henceforth, for the rest of the paper we make the following assumption:

Assumption 1. There is no alternative that is at least as good as any other alternative by both Players. That is $B_1 \cap B_2 = \emptyset$.

5. Equilibrium Properties

The first property is that this game admits an equilibrium in sincere strategies. This result does not follow from standard existence results because of the complexity of the strategy space—it is not finitely dimensional. Additionally, utility functions are not continuous.[10] Indeed, the outcome of the game (i.e. $b(s_1 \otimes s_2)$) "jumps" discontinuously whenever the limit of a sequence of non-consensual strategy profile is a consensual strategy profile. The proof of existence consists of approximating the game Φ using a sequence of finite two-player approval games whose set of alternatives contains the set of pure alternatives X and larger and larger (finite) subsets of $\Delta(X)$. Thus, each game in this sequence is a standard Approval voting game but with two players and a richer strategy space. Each such game admits an equilibrium in pure and sincere strategies as proved by Núñez and Laslier (2014). The limit of such a sequence of sincere equilibrium strategies, appropriately extended, is an equilibrium of the game Φ. The details of the proof of this result can be found in the Appendix.

Theorem 2. *Every game Φ has an equilibrium in sincere strategies.*

We turn to describing the equilibrium properties of the game.

Theorem 3. *Players play full-dimensional strategies in equilibrium.*

Proof. Let $s = (s_i, s_j)$ be an equilibrium and let $\bar{v}_i = \max_{p \in B_i} U_i(p)$ with $i = 1,2$. Proceeding by contradiction, assume first that $m \equiv \max\{\dim(s_1), \dim(s_2)\} < K-1$. Given Assumption 1 there is a Player i such that $U_i(s_i \otimes s_j) < \bar{v}_i$. Let s_i^ε denote the sincere strategy $s_i^\varepsilon = \{p \in \Delta \mid U_i(p) \geq \bar{v}_i - \varepsilon\}$. Note that s_i^ε is a full-dimensional strategy. Moreover, when ε is small enough, $s_i^\varepsilon \cap s_j = \emptyset$ because $U_i(s_i \otimes s_j) < \bar{v}_i$. Therefore, as ε decreases $U_i(s_i^\varepsilon \otimes s_j)$ becomes arbitrarily close to $\bar{v}_i$. This implies that Player i has a profitable deviation, proving that (s_i, s_j) is not an equilibrium. Therefore $m = K-1$.

Analogously, assume now $l \equiv \min\{\dim(s_1), \dim(s_2)\} < K-1$. Let $\dim(s_j) < K-1$. If $U_i(s_i \otimes s_j) < \bar{v}_i$ then, using the same definition for s_i^ε as before, Player i

[10] We have not specified a topology on the strategy space. However, the informal argument that follows should be sufficiently clear.

can make $U_i(s_i^\epsilon \otimes s_j)$ be arbitrarily close to $\bar{v}_i$, proving that she does not have a best response and contradicting that (s_i, s_j) be an equilibrium. If $U_i(s_i \otimes s_j) = \bar{v}_i$ then, in turn, Player j is not playing a best response to s_i. Indeed, playing a non-consensual strategy which contains all lotteries p with $U_j(p) > U_j(s_i \otimes s_j)$ strictly increases her utility. Therefore $l = K - 1$ as we wanted. □

In an intuitive sense, this property is related to the next equilibrium property which specifies that every equilibrium of the game is consensual. Each player plays a full-dimensional strategy in equilibrium so that her opponent does not find it profitable to deviate to a non-consensual strategy. Put differently, the outcome of any potential deviation by Player j to a non-consensual strategy is less harmful to Player i the "larger" the strategy that she plays is. Thus, for any equilibrium strategy (s_1, s_2), the equilibrium outcome is $b(s_1 \cap s_2)$ while the threat point sustaining such an equilibrium is $b(s_1 \cup s_2)$.

Theorem 4. *Every equilibrium is consensual.*

Proof. Suppose by contradiction that there is a non-consensual equilibrium (s_1, s_2). By Theorem 3, players play full dimensional strategies. Thus, we can use Lemma 3 to obtain both $b(s_1 \cup s_2) \in s_1$ and $b(s_1 \cup s_2) \in s_2$. But this implies $s_1 \cap s_2 \neq \emptyset$. Hence, any equilibrium must be consensual. □

The next property deals with the minimal utility level that a player can obtain from an ex ante viewpoint. This minimal level of utility corresponds to the utility level a player obtains from the barycenter $b(\Delta)$ of the simplex.

Theorem 5. *Each Player i gets at least $U_i(\Delta)$ in equilibrium.*

Proof. The sincere strategy $s_i^* \equiv \{p \in \Delta : U_i(p) \geq U_i(\Delta)\}$ guarantees a payoff of at least $U_i(\Delta)$ to Player i regardless of the strategy s_j played by Player j. This is clear if $s_i^* \cap s_j \neq \emptyset$. In turn, if Player j plays a non-consensual response to s_i^* then she plays a closed subset of $\Delta \setminus s_i^*$, that is, a (strict) subset of the set of lotteries that are less preferred than $b(\Delta)$ by Player i. Hence, $U_i(\Delta) = U_i(s_i^* \cup (\Delta \setminus s_i^*)) > U_i(s_i^* \cup s_j)$ for any strategy s_j that satisfies $s_j \subset \Delta \setminus s_i^*$. □

If the Bernoulli utility functions of the players are such that $u_1 = -u_2$ up to some affine transformation of utilities then we say that the Players have *opposing preferences*. In this case, the game has a unique equilibrium outcome.

Corollary 3. *If players have opposing preferences then the unique equilibrium outcome is the barycenter of the simplex $b(\Delta)$.*

Corollary 3 implies that the lower-bound on equilibrium payoffs given in Theorem 5 is the highest payoff that the game can guarantee players for any utility profile. In fact, when we consider a utility profile where players have opposing preferences we can see that the same statement is true for any mechanism whose set of possible outcomes is Δ.

Nonetheless, typically, the game has a continuum of Nash equilibrium outcomes. At the end of Section 6, after we characterize the set of equilibria in sincere strategies, we present an example that illustrates this fact.

6. Efficiency and Partial honesty

We turn to the efficiency properties of equilibria of the game Φ.

Definition 3 (Efficiency)**.** A lottery $p \in \Delta$ is (ex-ante) efficient if there is no $q \in \Delta$ such that $U_i(q) \geq U_i(p)$ for $i = 1,2$ with $U_i(q) > U_i(p)$ for at least some i.

If a lottery is Pareto efficient then it only gives positive probability to Pareto efficient alternatives. 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 that satisfies:

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

This shows that any lottery that assigns positive probability to inefficient alternatives is inefficient. If an efficient lottery is the equilibrium outcome of the game then, ex-post, players would never have a common incentive to renegotiate once the equilibrium outcome has realized into some alternative in X. In other words, the support of Pareto efficient lotteries only contains Pareto efficient alternatives.

Theorem 2 guarantees that the game has at least one sincere equilibrium. We now show that such an equilibrium is necessarily efficient.

Proposition 2. *Every sincere equilibrium outcome is efficient.*

Proof. Let (s_1, s_2) be a sincere equilibrium strategy. Every equilibrium is consensual (Theorem 4) so $s_1 \cap s_2 \neq \emptyset$. Lemma 2 implies that Players i's utility level associated with the sincere strategy s_i is $v_i \equiv \max_{p \in s_{-i}} U_i(p)$ and that, moreover, for every $p \in s_1 \cap s_2$ we have $U_i(p) = v_i$.

Suppose there is a $q \in \Delta$ such that $U_i(q) \geq v_i$ for $i = 1,2$, with strict inequality for at least one player. Then q is both in s_1 and s_2 because they are sincere strategies. But this contradicts our definition of v_i for at least one $i = 1,2$. Thus, every lottery in the winning set of a sincere every equilibrium is Pareto efficient. □

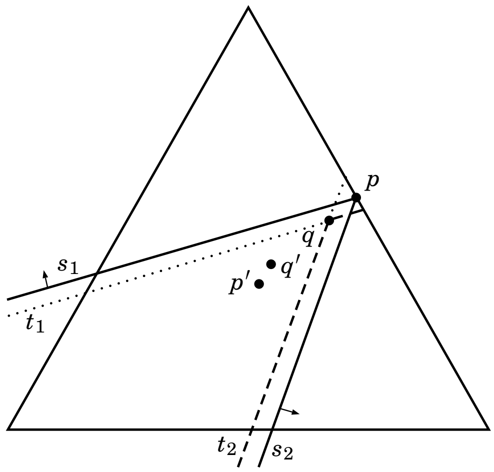

FIGURE 2. An inefficient equilibrium.

However, not every equilibrium of the game is efficient. This is illustrated in the next example.

Example 4. In Figure 2 we represent a bargaining game with three alternatives and a sincere equilibrium (s_1, s_2). The intersection of the equilibrium strategies s_1 and s_2 consists of only one point p and the strategies are defined by $s_i = \{r \in \Delta : U_i(r) \geq U_i(p)\}$ for $i = 1,2$. The lottery p' is defined by $p' = b(s_1 \cup s_2)$ and it is to be considered as the threat point of the equilibrium (s_1, s_2). Note that either player can induce an outcome as close as they wish to the lottery p' by deviating to a sincere non-consensual strategy. But both players prefer p to p', thus confirming that (s_1, s_2) is a equilibrium. Such an equilibrium is clearly efficient.

We now construct an inefficient equilibrium by first considering indifference curves associated with a slightly lower utility levels for both players. These new indifference curves cross in the lottery q in the interior of the simplex. We obtain the strategy profile (t_1, t_2) inducing the consensual outcome q by bending the indifference curves at q to obtain t_1 as the area to the north-west of the dotted line and t_2 as the area to the south-east of the dashed line. Note that no player can profitably deviate to a different consensual strategy. The new threat point $q' = b(t_1 \cup t_2)$ is close-by to the old threat point due to a continuity argument and, therefore, no player can profitably deviate to a non-consensual strategy either. Hence (t_1, t_2) is an equilibrium inducing the inefficient lottery q.

In the inefficient equilibrium of the previous example, both players are indifferent between playing their insincere equilibrium strategy and some sincere strategy. The inefficient outcome arises because players coordinate in their insincere strategies. However, if we slightly depart from rationality and assume that players always play a sincere strategy whenever they have one available in their set of best responses then this sort of equilibria disappears.

This assumption is equivalent to saying that players are *partially honest*, an assumption recently proposed in the implementation literature. We follow the formal definition of partial honesty given by Dutta and Sen (2012). Other definitions that are present in the literature (see among others the ones by Matsushima (2008a) or Kartik and Tercieux (2012)). While not being formally equivalent, they also share the common feature of triggering a lexicographic preference for sincerity. For this reason, our results do not depend on which definition of partial honesty we adopt.

Henceforth, the set of sincere strategies for Player i is denoted by $\mathscr{S}_i$. We denote by $\succeq_i$ Player i's ordering over the set of strategy profiles S when she is partially honest. Its asymmetric component is denoted by $\succ_i$.

Definition 4. Player i is partially honest if for any two (s_i, s_{-i}), $(s'_i, s_{-i}) \in S$.

(1) If $U_i(s_i \otimes s_{-i}) \geq U_i(s'_i \otimes s_{-i})$ and $s_i \in \mathscr{S}_i$, $s'_i \notin \mathscr{S}_i$, then $(s_i, s_{-i}) \succ_i (s'_i, s_{-i})$.
(2) In all other cases, $(s_i, s_{-i}) \succeq_i (s'_i, s_{-i})$ if and only if $U_i(s_i \otimes s_{-i}) \geq U_i(s'_i \otimes s_{-i})$.

The first part of the definition represents the individual's partial preference for honesty. She strictly prefers the strategy profile (s_i, s_{-i}) to (s'_i, s_{-i}) when s_i is a sincere strategy and s'_i is not, provided that the outcome corresponding to (s_i, s_{-i}) is at least as good as the one corresponding to (s_i, s_{-i}). The second part of the definition implies that in every other case, the player's preference ordering over the corresponding strategy profiles is not altered.

The preference profile $(\succeq_1, \succeq_2)$ now defines a modified normal form game. We omit formal definitions for the sake of brevity. The next proposition is a trivial and important implication of Corollary 1 and Proposition 2.

Proposition 3. *In the game with partially honest players, a player's best response is sincere and every equilibrium sincere and Pareto efficient.*

Assuming partial honesty allows us to focus, for each Player i, on her set of sincere strategies $\mathscr{S}_i \subset S$. Such a subset of strategies has a simple characterization. For each Player i let $\bar{v}_i = \max_{x \in X} u_i(x)$ and $\underline{v}_i = \min_{x \in X} u_i(x)$. To each utility value $v_i \in [\underline{v}_i, \bar{v}_i]$ we associate the sincere strategy $s_i(v_i) \equiv \{p \in \Delta : U_i(p) \geq v_i\}$.

We turn to characterizing the set of equilibria under partial honesty. Given a sincere strategy of a player, the other player's best response is either consensual or non-consensual. Since every equilibrium is consensual, to show that a given strategy profile is an equilibrium we need to prove (1) that both players are playing their best consensual response, and that (2) they do not gain by deviating to a 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 $v_j \in (\underline{v}_j, \bar{v}_j)$, we let $\mathrm{CU}_i(v_j)$ denote Player i's utility value from the best sincere consensual response to s_j. Instead of working with the analogous expression for Player i's best sincere non-consensual response (that, as we argued before, might not exist) we let $\mathrm{NU}_i(v_j)$ denote the utility value from the unique sincere strategy that maximizes $V_i(\cdot, v_j)$ (see Equation (3.1)). Recall that if the sincere strategy $s_i(v_i)$ maximizes $V_i(\cdot, v_j)$ and $s_i(v_i) \cap s_j(v_j) = \emptyset$ then $s_i(v_i)$ is the best response to $s_j(v_j)$ and, therefore, also the best non-consensual response to $s_j(v_j)$.

Note that CU_i and NU_i are continuous functions on $(\underline{v}_j, \bar{v}_j)$. Furthermore, CU_i is nonincreasing in v_j (because $s_j(v_j) \subset s_j(v'_j)$ whenever $v'_j > v_j$).

Proposition 4. *In the game with partially honest players, for each Player i there exists a unique $\eta^i \in (\underline{v}_j, \bar{v}_j)$ such that:*

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

Proof. We already argued (proof of Lemma 4) that if $v_j \in (\underline{v}_j, \bar{v}_j)$ and the best response to $s_j(s_j)$ is consensual then $\mathrm{CU}(v_j) \geq \mathrm{NU}(v_j)$ and that if $v_j \in (\underline{v}_j, \bar{v}_j)$ and the best response to $s_j(v_j)$ is non-consensual then $\mathrm{NU}(v_j) \geq \mathrm{CU}(v_j)$. If v_j is close enough to $\underline{v}_j$ then Player i's best response to $s_j(v_j)$ is consensual because she can obtain a payoff close to $\bar{v}_i$ by playing a consensual best response while she can only get a payoff close to $U_i(\Delta)$ by playing a non-consensual response. In turn, if v_j is close enough to $\bar{v}_j$ then Player i's best response to $s_j(v_j)$ is non-consensual because Player i can obtain a utility close to $\bar{v}_i$ by playing a non-consensual strategy (in a similar vein as in Example 3) whereas she can only get, at most, a utility close to her second most preferred alternative if she plays a consensual best response (due to Assumption 1). The continuity of CU_i and NU_i as functions of v_j implies the existence of some $\eta^i \in (\underline{v}_j, \bar{v}_j)$ for which $\mathrm{CU}_i(\eta^i) = \mathrm{NU}_i(\eta^i)$.

To prove uniqueness, suppose that $\mathrm{CU}_i(v_j) = \mathrm{NU}_i(v_j)$ for some $v_j > \eta^i$. Since $s_j(v_j) \subset s_j(\eta^i)$ and $s_j(\eta^i) \setminus s_j(v_j)$ is a set with positive measure that only contains lotteries that give Player i utility less than η^i we have $\mathrm{NU}_i(v_j) > \mathrm{NU}_i(\eta^i)$. Moreover CU_i is nonincreasing on v_j so that $\mathrm{CU}_i(\eta^i) \geq \mathrm{CU}_i(v_j)$. Hence, for any $v_j > \eta^i$, we have $\mathrm{NU}_i(v_j) > \mathrm{CU}_i(v_j)$. □

We can now complete the full characterization of the set of equilibria in the bargaining game when players are partially honest. on of the set of equilibria in the bargaining game when players are partially honest.

TABLE 1. Maximum and minimum equilibrium payoffs of Φ for utility profiles $u_1 = (10, u, 0)$ and $u_2 = (0, v, 10)$ when players are partially honest. Values are rounded up to two decimal places.

$u = v$	1	2	3	4	5	6	7	8	9	10
$\underline{u}_1$	8.91	8.26	7.01	5.35	5	5.43	6.00	6.71	7.64	10
$\overline{u}_1$	9.78	9.25	8.35	6.98	5	6.38	7.43	8.32	9.15	10

Proposition 5. *In the game with partially honest players, let (v_1, v_2) be a utility profile derived from some Pareto efficient lottery. The profile $(s_1(v_1), s_2(v_2))$ is an equilibrium payoff if and only if $v_j \leq \eta^i$ for both $i = 1, 2$.*

Proof. Let $p \in \Delta$ be Pareto efficient and let $v_i = U_i(p)$ for $i = 1, 2$. Consider the strategy profile $(s_1(v_1), s_2(v_2))$. We have $p \in s_1(v_1) \cap s_2(v_2)$ and, because p is Pareto efficient, such an intersection has an empty interior. Thus, no player has an incentive to deviate to a different consensual strategy. Furthermore, since $v_1 \leq \eta^2$ and $v_2 \leq \eta^1$, the previous proposition implies that no player has an incentive to deviate to a non-consensual strategy.

On the other hand, let (v_1, v_2) be an equilibrium payoff. From Lemma 3 we know that players are playing $(s_1(v_1), s_2(v_2))$ which, by Theorem 4, is a consensual strategy profile. Because players do not have an incentive to deviate to a non-consensual strategy we have $v_1 \leq \eta^2$ and $v_2 \leq \eta^1$. □

Thus, the set of equilibria is homeomorphic to a closed interval. The previous result can be used to compute the set of Nash equilibria for any given utility profile. We do so in the next example.

Example 5. Consider a bargaining situation with set of alternatives $X = \{x_1, x_2, x_3\}$. Players 1 and 2 have Bernoulli utility functions $u_1 = (10, u, 0)$ and $u_2 = (0, v, 10)$. If $u + v = 10$ then there is a unique equilibrium outcome $b(\Delta)$ (Corollary 3).

Otherwise, the game does not have a unique equilibrium outcome. We consider the family of games where $u = v$. For each of these games, Proposition 4 gives the maximum equilibrium utilities for Player i (hence, it also gives the minimum equilibrium utility for Player j.) For instance, if $u = v = 1$, then there is a continuum of equilibrium outcomes in which Player 1 obtains some utility value $\hat{u} \in [8.91334, 9.77993]$ and Player 2 obtains utility $\frac{90-\hat{u}}{9}$.

The next table depicts the interval for the equilibrium payoffs for player 1 for any given $u = v \in \{1, \ldots, 10\}$.

In any such situation, if Player 1 obtains payoff $\hat{u}$ then Player 2 obtains payoff $p(\hat{u})$ with:

$$p(\hat{u}) = \begin{cases} \frac{100-(10+\hat{u})u}{10-u} & u \leq 5, \\ \frac{10u-\hat{u}u}{10-u} & u > 5. \end{cases}$$

7. Conclusion

This paper develops an intuitive mechanism to reach agreements between two players. Among the several appealing features, we have shown that agreement always occurs in every equilibrium and that, if players are partially honest, every equilibrium is sincere and efficient.

A natural research question that arises is whether this mechanism can be extended to many players. The answer to this question seems far from obvious. Two players either agree on some common lottery or they do not. However, this duality is lost in a multiplayer settings. The main problem seems to be what the rules of the game should specify to determine the outcome when when some but not all players agree on some set of lotteries. While one might think of several possible extensions, none of them seems to conveniently extend the properties of the current approval bargaining game.

Appendix: Existence of equilibrium

The proof of existence of equilibrium builds 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 (2014). Each player selects a subset of the finite set alternatives that she approves. If the intersection of these two subsets is non-empty 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 (2014).

(α) 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.

(β) If an equilibrium outcome is non-consensual then each player approves *every* alternative that she prefers to the equilibrium outcome.

(γ) 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 $I^0 \equiv I$ and let c be the barycenter of I^0 and $C^0 \equiv \{c\}$. The probability measure λ^0 gives probability 1 to $c \in I^0$. For each $t > 0$, let I^t be the collection 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 $C^t \equiv \{c(h) : h \in I^t\}$. 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 with set of alternatives $X^t \equiv C^t \cap \Delta$. 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 will be used to approximate outcomes in the game Φ with a sequence of outcomes of the finite approval games constructed above. The proof consists of 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.[11] 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:

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

Lemma 5. *Let $E \subset \Delta$ satisfy $\lambda(E) > 0$ and $\lambda(\partial E) = 0$, and define $E^t \equiv X^t \cap E$. 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

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

h normalized by the volume of I. For any bounded, uniformly continuous function $f: I \to \mathbb{R}$,

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

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

Now we can finally prove:

Theorem 2. *Every game Φ has an equilibrium in sincere strategies.*

Proof. Given property (α) 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 $v_i^t \equiv \min_{p \in s_i^t} U_i(p)$. The utility to Player i from *every* lottery in s_i^t is at least v_i^t. The sequence $\{(v_1^t, v_2^t)\}_{t=1}^{\infty}$ is contained in a compact set, therefore, it has a subsequence that converges to some (v_1^*, v_2^*). For each $i = 1,2$ define the sincere strategy $s_i^* \equiv \{p \in \Delta : U_i(p) \geq v_i^*\}$. 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 $\lim(v_1^t, v_2^t) = (v_1^*, v_2^*)$ 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 (β) 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 $q_i^t \equiv \arg\min_{p_i^t \in s_i^t} \|p_i^t, b(s_1^t \cup s_2^t)\|$ 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 (v_1^*, v_2^*).

To the contrary and without loss of generality, assume that Player 1 gets a payoff strictly higher than v_1^* under the strategy profile (s_1^*, s_2^*) so that $U_1(s_1^* \cap s_2^*) > v_1^*$. There must be a $\hat{p} \in s_2^*$ such that $U_1(\hat{p}) > v_1^*$. Such an inequality also holds for every point in some closed neighborhood P of $\hat{p}$. Thus, for t high enough, we can choose a $\hat{p}^t \in S^t \cap P$ such that $U_1(\hat{p}^t) > v_1^*$ and $\hat{p}^t \in \mathrm{int}(s_2^*)$ (i.e. $U_2(\hat{p}^t) > v_2^*$). This means that $\hat{p}^t \in s_2^t$ for sufficiently high t. Therefore, $U_1(s_1^t \otimes s_2^t) \geq U_1(\hat{p}^t)$ for any sincere equilibrium (s_1^t, s_2^t) of Γ^t. But then, also for every sufficiently high t,

$$v_1^t \geq U_1(s_1^t \otimes s_2^t) \geq U_1(\hat{p}^t) > v_1^*, \tag{A.1}$$

where the first inequality follows from (γ). But this is impossible because v_1^* is the limit point of the sequence $\{v_1^t\}_{t=1}^{\infty}$. This provides a contradiction so we can conclude that (s_1^*, s_2^*) generates expected payoffs (v_1^*, v_2^*).

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

Suppose again by contradiction that (s_1^*, s_2^*) is not an equilibrium of Φ. Without loss of generality, let there be an $\hat{s}_1$ such that $U_1(\hat{s}_1 \otimes s_2^*) > v_1^*$. The fact that (s_1^*, s_2^*) induces the consensual outcome $b(s_1^* \otimes s_2^*)$ that generates the vector of utility levels (v_1^*, v_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 $b(\hat{s}_1^t \otimes s_2^t)$ is non-consensual and Lemma 5 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

$$U_1(\hat{s}_1^t \cup s_2^t) > v_1^* + \varepsilon. \tag{A.2}$$

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 (γ) implies that $U_1(s_1^t \cup s_2^t) \leq v_1^t$ for every t. It follows that $\lim U_1(s_1^t \cup s_2^t) \leq v_1^*$ and, for every t high enough, $U_1(s_1^t \cup s_2^t) \leq v_1^* + \varepsilon$. 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^*) must be an equilibrium of Φ. $\square$

References

J.S. Banks and J. Duggan. A bargaining model of collective choice. *American Political Science Review*, 94:73–88, 2000.

D.P. Baron and J.A. Ferejohn. Bargaining in legislatures. *American Political Science Review*, 83:1181–1206, 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.

W.A. Dembski. Uniform probability. *Journal of Theoretical Probability*, 3(4): 611–626, 1990.

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.

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. ISSN 0899-8256. doi: http://dx.doi.org/10.1016/j.geb.2013.11.011. URL `http://www.sciencedirect.com/science/article/pii/S0899825613001644`.

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

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. ISSN 0165-1765. doi: http://dx.doi.org/10.1016/j.econlet.2007.12.008. URL `http://www.sciencedirect.com/science/article/pii/S016517650700434X`.

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. 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. mimeo, University of Cergy-Pontoise, 2014.

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

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

A. Rubinstein. Perfect Eequilibrium 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.