Rational Bargaining: Characterization and Implementation.[*]

Philipp Peitler[‡]

October 25, 2023

[Latest version here]

Abstract

The von Neumann-Morgenstern axioms are uncontroversial desiderata for individual decision making. We say that a bargaining solution is rational, if it can be interpreted as the most preferred alternative under these axioms. Yet, we find that neither the Nash nor the Kalai-Smorodinsky bargaining solution is rational in this sense. We formalize two consequences of rationality, namely that one can neither be strictly better off nor strictly worse off from randomizing over different actions. These two axioms, together with other standard axioms, characterize the relative utilitarian bargaining solution. We then implement this bargaining solution in sub-game perfect equilibrium.

Keywords: axiomatic bargaining, relative utilitarian, rationality, implementation.

1 Introduction

The foundations of individual decision making in economics are laid by the axioms of von Neumann and Morgenstern (1944). While there is plenty evidence that these axioms do not always accurately describe the behavior of economic agents, there is

[*]I gratefully acknowledge financial support from the *Jubiläumsfonds* of the Austrian National Bank (OeNB), project number 18719.

[‡]University of Vienna; philipp.peitler@univie.ac.at

little disagreement about their normative desirability. These axioms are widely accepted, and acting in accordance with them is often synonymous with rationality. We feel that rationality is equally desirable in the bargaining setting, where a bargaining solution prescribes what a group or arbitrator should choose. We define a bargaining solution as rational, if it can be interpreted as the most preferred alternative under some von Neumann-Morgenstern (vNM) preference relation. A consequence of the vNM axioms is that an agent would only randomize over different actions if she were indifferent between them. Consequently, there is no benefit of randomizing compared to choosing a deterministic action in the support. However, we find that the prominent bargaining solutions by Nash (1950) and Kalai and Smorodinsky (1975) violate this condition. Consider a scenario where two agents bargain over a single indivisible item and an arbitrator is appointed to choose an allocation. Both the Nash and Kalai-Smorodinsky (KS) bargaining solution would prescribe that the arbitrator flips a fair coin and allocates the item to the winner. However, neither solution selects the two deterministic allocations over which the coin randomizes. The prominent bargaining solutions demand that the arbitrator strictly prefers the coin-flip over the deterministic allocations. Clearly, this violates the vNM axioms.

In this paper, we aim to find a rational bargaining solution. To capture a central aspect of rationality, we propose the *no benefit of randomization* (NBR) axiom. This axiom relates non-convex bargaining sets to their convex hull. A non-convex bargaining set represents a situation, where randomization is either not feasible or not permitted. Allowing for randomization means the group or arbitrator can choose from the convex hull of this set.[1] The axiom then says that if an alternative is selected by the bargaining solution in the non-convex set, it must still be selected in the convex hull of the set. Another consequence of the vNM axioms, and the flip-side of NBR, is that the agent is never strictly worse off when randomizing. Hence, when randomization is possible (i.e. the bargaining set is convex) and two alternatives are selected by the bargaining solution, then any mixture (i.e. convex combination) of these two alternatives must be selected as well. We call this axiom *convexity* (CONV). We find that NBR and CONV, together with standard axioms, characterize the *relative utilitarian* (RU) bargaining

[1] We assume here that individuals satisfy the vNM axioms. This is a standard interpretation of the bargaining set, shared by both Nash (1950) and Kalai and Smorodinsky (1975).

solution. The RU bargaining solution selects the alternatives with the highest sum of normalized utilities. Utilities are normalized such that the disagreement point has utility 0 and the best alternative has utility 1. The other axioms that underlie our characterization are invariance to the utility scale, strong Pareto, weak symmetry and weak IIA, which is a relaxation of Nash's independence of irrelevant alternatives axiom. The proof for two agents is simple and resembles the one of Nash (1950). The proof easily generalizes to any number of agents.

Reliance on randomization can be problematic for reasons other than the one outlined above. First, randomization might not be technologically feasible in a given situation. Second, even if feasible, it might not be transparent and might therefore be rejected by the agents. Third, randomization raises issues of commitment. An agent might agree to a lottery over the alternatives, but reject the outcome of the lottery if it isn't in her favor. Finally, randomization raises issues of ex-post fairness, because not only the lottery, but also the outcome of the lottery should satisfy normatively desirable postulates. These problems have sparked a large literature on the fair allocation of indivisible goods, which tries to find fair allocations without the use of randomization.[2] Note that if an arbitrator were to follow either the Nash or KS bargaining solution, he would have a desire to randomize in secret, when explicit randomization is prohibited. Our axiom NBR ensures that an arbitrator would have no desire to do so. This allows us to deal with problems of fair allocation, such as the division of an inheritance among descendants or the division of a household after a divorce, without relying on randomization.

Besides a characterization, we also implement the RU bargaining solution in subgame perfect equilibrium. This means that we identify a bargaining protocol, which in equilibrium leads to a RU-optimal alternative. We show that full implementation is not possible and identify a game form that weakly implements our bargaining solution. However, we are quite close to full implementation, as any RU-optimal alternative that is strictly better than the disagreement point for every agent is an equilibrium outcome. In this game, agents simultaneously make a proposal, consisting of an alternative and

[2] For a recent survey of the literature see Amanatidis et al. (2022). Note that the typical axioms of this literature differ from the ones in the fair bargaining setting. Typical desiderata are no-envy, proportionality and their relaxations by Budish (2011).

a probability for each agent. In equilibrium, all agents propose the same RU-optimal alternative and this alternative is implemented immediately. If there is disagreement among the agents, the proposal with the highest sum of probabilities has to be sequentially approved by all agents. Agents can choose whether to accept the alternative or receive a utility equal to the probability that the proposal assigns to them. Hence, the probabilities can be understood as a claim regarding how good the alternative is for each agent. If the claim was inflated, the alternative is rejected by at least one agent. If an agent rejects, all other agents receive 0 utility. Our game is similar to the ones by Moulin (1984) and Moore and Repullo (1988). The only other implementations of the RU bargaining solutions are by Miyagawa (2002) and Hagiwara (2020). They however consider the case of only two agents and strictly convex bargaining sets. Note that under strictly convex bargaining sets, there is a unique RU-optimal alternative. Our game on the other hand typically has multiple equilibrium outcomes, each corresponding to one of the multiple RU-optimal alternatives.

Other axiomatizations of the RU bargaining solution are by Pivato (2009) and Baris (2018). Pivato (2009) considers preferences over bargaining solutions and then imposes axioms on these preferences. This differs from the standard approach, established by Nash (1950), where axioms are imposed on the bargaining solution directly. Baris (2018) adapts the characterization of the utilitarian bargaining solution by Myerson (1981) to a utility-scale invariant setting. Their central axiom can be interpreted as a dynamic consistency condition. When facing uncertainty over what the bargaining set will be, the arbitrator makes a plan, which specifies for each possible bargaining set a utility vector. Then the expected utility vector must be the solution in the expected bargaining set. Cao (1982) identifies necessary axioms for the RU bargaining solution but does not provide a characterization. Note that Cao (1982), Pivato (2009) and Baris (2018) assume the bargaining set to be convex, whereas we contribute to the literature on bargaining over non-convex sets (Kaneko, 1980; Zhou, 1997; Mariotti, 1998a,b; Conley and Wilkie, 1996; Denicolò and Mariotti, 2000; Ok and Zhou, 1999; Nagahisa and Tanaka, 2002; Xu and Yoshihara, 2006; Zambrano, 2016).

Related to the RU bargaining solution is a large literature on preference aggregation which characterizes a relative utilitarian rule (Karni, 1998; Dhillon and Mertens, 1999;

Segal, 2000; Börgers and Choo, 2017; Marchant, 2019; Sprumont, 2019; Brandl, 2021; Peitler and Schlag, 2023). Especially related is Peitler and Schlag (2023). In an application of their aggregation rule, the RU bargaining solution is derived from the most preferred element of a menu-dependent social preference, where the menu consists of the alternatives which are better for every agent than the disagreement point.

Finally, there is a literature that considers the rationalizability of bargaining rules (Peters and Wakker, 1991; Bossert, 1994; Sánchez, 2000; Xu and Yoshihara, 2013). A bargaining rule is rationalizable if it can be interpreted as the most preferred alternative under a single preference relation over utility vectors, which applies independent of the bargaining set. In this literature, rationality is understood as satisfying the weak axiom of revealed preference (or similar conditions). We take rationality to mean that the bargaining solution is consistent with the maximization of a vNM preference relation, but we do not insist that it is the same preference relation for every bargaining set.

We present the axioms and state the representation theorem in Section 2. The implementation is in Section 3. In Section 4 we consider other rational solutions. Section 5 concludes.

2 Axiomatization

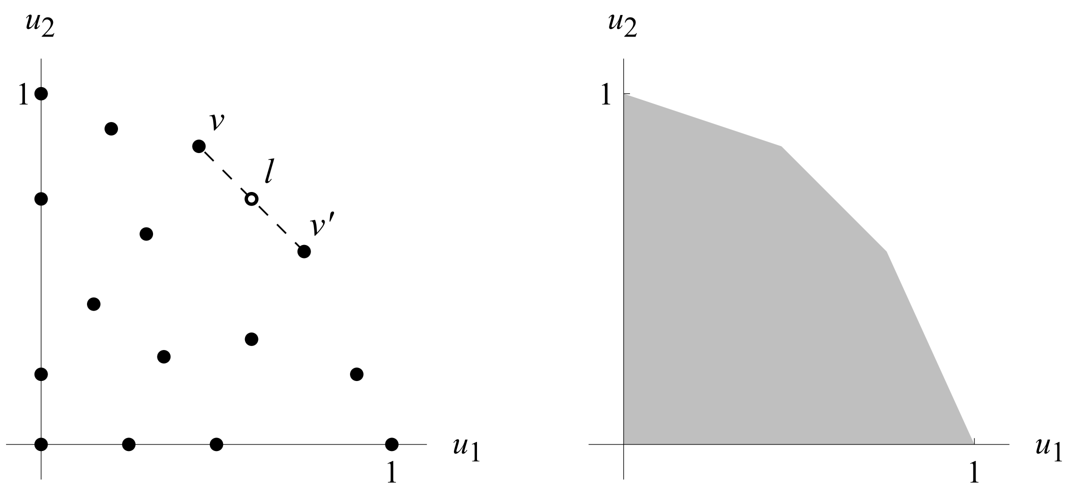
Let $N := \{1, ..., n\}$ be a set of agents where $n \in \mathbb{N}$ and $n \geq 2$. A bargaining problem (S, d) consists of a bargaining set $S \subseteq \mathbb{R}^n$ and a disagreement point $d \in S$. We simplify notation by normalizing the disagreement point to $d = (0, ..., 0)$ and write S instead of (S, d). Let $\mathbb{R}_+$ denote the positive real numbers including 0. We restrict attention to bargaining sets S where (i) $S \subseteq \mathbb{R}^n_+$ (ii) S is compact and (iii) for each $i \in N$ there exists a $u \in S$ such that $u_i > 0$. We denote the domain of bargaining sets that satisfy these properties by $\mathcal{S}$. A bargaining solution f is a correspondence that assigns to every $S \in \mathcal{S}$ a non-empty subset of S.

Our central axiom is *no benefit of randomization*. For any $R \subset \mathbb{R}^n$, let $\operatorname{conv} R$ denote the convex hull of R.

Axiom NBR (No Benefit of Randomization)**.** For every $S \in \mathcal{S}$,

$$f(S) \subseteq f(\operatorname{conv} S).$$

We illustrate the axiom with the help of following example. Consider a finite bargaining set S, arising from the allocation of finitely many indivisible items. Now consider a lottery l that realizes some allocation $v \in S$ with probability $\lambda \in (0, 1)$ and another allocation $v' \in S$ with probability $1 - \lambda$. Since the utilities in v and v' express individuals' vNM preferences, Agent i's utility of l is $\lambda v_i + (1 - \lambda)v'_i$. Hence, if l were feasible, it would be a point in the bargaining set at $\lambda v + (1 - \lambda)v'$. See the left hand side of Figure 1 for an illustration when $n = 2$. Consequently, if every lottery over the allocations in S were feasible, the bargaining set would be the convex hull of S. This case is depicted on the right hand side of Figure 1. NBR says that if an allocation is optimal in the non-convex bargaining set S, where randomization isn't feasible, then this allocation must still be optimal in the convex bargaining set $\operatorname{conv} S$, where randomization is feasible. Hence, randomization doesn't introduce a lottery, strictly better for the group than the best allocation, as this would contradict that the group acts in accordance with the vNM postulates. Note however that randomization can introduce new utility vectors that are equally optimal as an allocation, which is why NBR does not demand $f(S) = f(\operatorname{conv} S)$.



(a) S.　　(b) $\operatorname{conv} S$.

Figure 1: Convexification of the bargaining set.

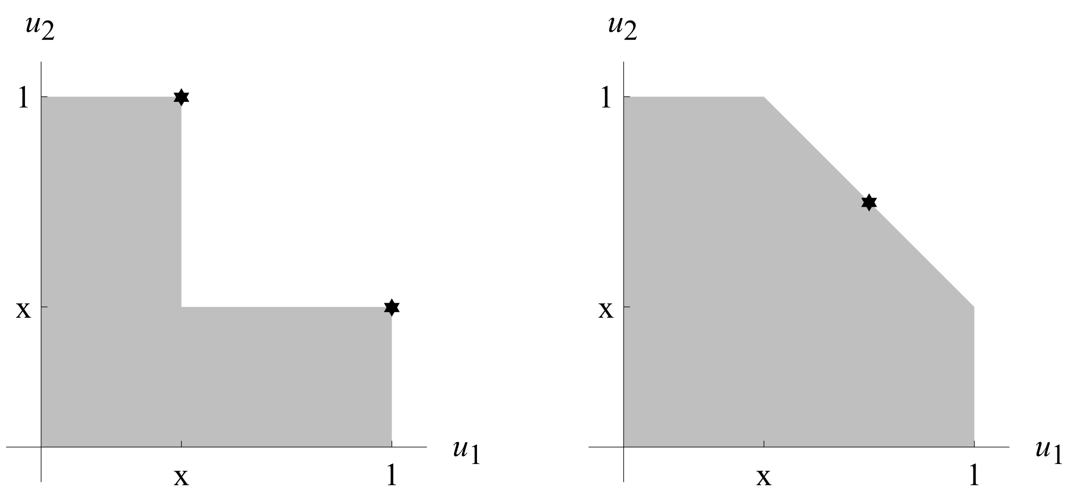
Note that NBR has no bite when the domain is restricted to convex bargaining sets, which is the classic setting of Nash (1950) and Kalai and Smorodinsky (1975). However, these popular bargaining solutions have been extended in various ways to domains that include non-convex sets. We find that these extensions violate NBR. In the following we demonstrate this for the extensions by Xu and Yoshihara (2006). Let f^{Nash} denote the Nash bargaining solution and f^{KS} denote the Kalai-Smorodinsky bargaining solution as in Xu and Yoshihara (2006).

Proposition 1. Both f^{Nash} and f^{KS} violate NBR.

Proof. Assume $n = 2$ and consider the barging set

$$S = \{(u_1, u_1) \in [0,1]^2 : u_1 \leq x \text{ or } u_2 \leq x\}$$

for some $x \in (0,1)$. Then $f^{\text{Nash}}(S) = \{(1,x),(x,1)\}$ and $f^{\text{KS}}(S) = \{(x,x)\}$. However, $f^{\text{Nash}}(\text{conv}\, S) = f^{\text{KS}}(\text{conv}\, S) = \{(\frac{1+x}{2}, \frac{1+x}{2})\}$. Hence, NBR is violated. Figure 2 illustrates this for the Nash solution. $\square$



(a) S.

(b) $\text{conv}\, S$.

Figure 2: Violation of NBR by f^{Nash}. Stars indicate the solution.

Another consequence of the vNM axioms is that the arbitrator cannot be strictly worse off under randomization. If two utility vectors are both optimal from the perspective of the arbitrator, then a lottery over these vectors, assuming it is feasible, must also be optimal. This is captured by the following axiom.

Axiom CONV (Convexity)**.** For every $S \in \mathcal{S}$, $f(S)$ is convex whenever S is convex.

Note that both the Nash and KS bargaining solution trivially satisfy this axiom, since these solutions are singletons whenever the bargaining set is convex.

Since the prominent bargaining solutions violate NBR and are therefore not rational, we are in need of an alternative solution. Besides the aforementioned axioms, this solution should satisfy agreed upon desiderata. Both Nash (1950) and Kalai and Smorodinsky (1975) agree that a solution should be Pareto efficient, invariant to the utility scale and that it should give equal treatment to symmetric agents. These axioms have been formulated under the assumption that the solution is single-valued. In the following, we translate these axioms to our setting, where the solution can be set-valued.

Axiom PO (Pareto Optimality)**.** For every $S \in \mathcal{S}$, if $u \in f(S)$ then there is no $v \in S$ such that $v \neq u$ and $v_i \geq u_i$ for all $i \in N$.

We say that α is a positive linear transformation if there exists $k_1, ..., k_n > 0$ such that for any $R \subseteq \mathbb{R}^n$, $\alpha(R) = \{u \in \mathbb{R}^n : (k_1 u_1, ..., k_n u_n) \in R\}$.

Axiom INV (Invariance)**.** For every $S \in \mathcal{S}$ and positive linear transformation α,

$$f(\alpha(S)) = \alpha(f(S)).$$

For any $R \subseteq \mathbb{R}^n$, we say that R is symmetric if for every $u \in R$, any permutation of u is in R as well.

Axiom SYM (Symmetry)**.** For every $S \in \mathcal{S}$, if S is symmetric and $u \in f(S)$ then $u = (x, ..., x)$ for some $x \in \mathbb{R}$.

Note that SYM violates PO in non-convex settings. To see this, consider $n = 2$ and $S = \{(0, 0), (1, 0.9), (0.9, 1)\}$. Then SYM would demand $f(S) = \{(0, 0)\}$, a violation of PO. Therefore, for domains that include non-convex bargaining sets, symmetry needs to be weakened.

Axiom WSYM (Weak Symmetry)**.** For every $S \in \mathcal{S}$, if S is symmetric then so is $f(S)$.

For the fourth and final axiom, Nash (1950) has independence of irrelevant alternatives (IIA) and Kalai and Smorodinsky (1975) have monotonicity. Note that IIA would not be compatible with the axioms we have imposed so far (NBR, CONV, PO, INV and WSYM) and there is no obvious extension of monotonicity to set-valued solutions. However, there is a weaker version of IIA that is satisfied by both the Nash and KS solution, which we call *weak IIA*.[3] Furthermore, this axiom replaces monotonicity in axiomatizations of the KS solution in non-convex settings (Nagahisa and Tanaka, 2002; Xu and Yoshihara, 2006). Since weak IIA is a common denominator of the Nash and KS solution, we will impose it as well. Before we introduce this axiom, let us first translate Nash's IIA to a setting with set-valued solutions.

Axiom IIA (Independence of Irrelevant Alternatives)**.** For every $S, S' \in \mathcal{S}$, if $S' \subset S$ and $f(S) \cap S'$ is non-empty then

$$f(S') = f(S) \cap S'.$$

Next we state weak IIA. For any $S \in \mathcal{S}$ and $i \in N$ let $m_i(S) := \max\{u_i : u \in S\}$. Furthermore, let $m(S) := (m_i(S))_{i \in N}$. Following Yu (1973), we call $m(S)$ the *utopian point* of S.

Axiom WIIA (Weak IIA)**.** For every $S, S' \in \mathcal{S}$, if $S' \subset S$, $m(S') = m(S)$ and $f(S) \cap S'$ is non-empty then

$$f(S') = f(S) \cap S'.$$

IIA says that removing points from the bargaining set does not change what is optimal from the perspective of the group. WIIA weakens this condition, by imposing the former demand only in cases where the removal of points does not change the utopian point. The utopian point $m(S)$ is the maximal possible utility of each agent under the bargaining set S. While the utopian point typically isn't feasible (i.e. $m(S) \notin S$), it is an anchor point that allows us to relate the utilities of different agents. Since vNM utilities are unique only up to a positive affine transformation, comparisons of absolute utility levels across different individuals are meaningless. However, by INV we can normalize every bargaining set such that $m(S) = (1, ..., 1)$. Then all utilities express how well-off each agent is relative to their best possible outcome. In contrast to absolute utility levels, this measure can be meaningfully compared across individuals.

[3] Weak IIA and variations thereof have already appeared in Yu (1973), Roth (1977), Cao (1982), Imai (1983), Dubra (2001), Nagahisa and Tanaka (2002), Xu and Yoshihara (2006) and Rachmilevitch (2019).

Removing a point $u \in S$ from S such that $m_i(S \setminus \{u\}) < m_i(S)$ for some $i \in N$ would require us to re-normalize the bargaining set, which would change how points other than u are perceived by the group. Unlike IIA, WIIA allows this change of perspective to influence what is collectively optimal.

We show that the above axioms characterize the *relative utilitarian* bargaining solution. Define f^{RU} such that for every $S \in \mathcal{S}$,

$$f^{\mathrm{RU}}(S) = \arg\max_{u \in S} \sum_{i \in N} \frac{u_i}{m_i(S)}.$$

Theorem 1. f satisfies NBR, CONV, PO, INV, WSYM and WIIA if and only if $f \equiv f^{\mathrm{RU}}$.

For $n = 2$ the proof is short, simple and resembles the one of Nash (1950). For this reason, we present it here in the main section. A proof for a general number of agents is in Appendix A.

Proof. First, consider the bargaining set $S_x = \{(0,0),(1,0),(0,1),(1,x),(x,1)\}$ for some $x \in [0,1]$, as illustrated in Figure 3. Note that $P_x := \{(1,x),(x,1)\}$ is the set of

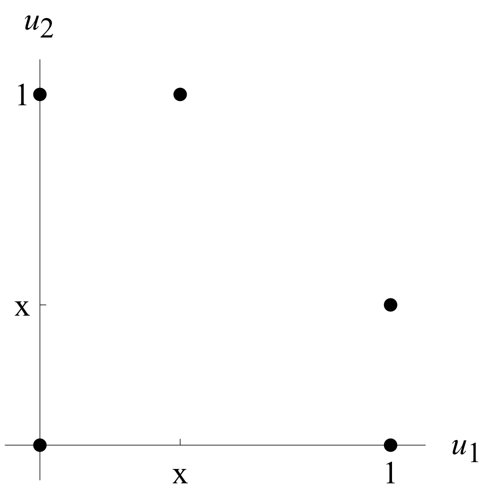


Figure 3: S_x.

Pareto optimal points in S_x and by PO, $f(S_x) \subseteq P_x$. Furthermore, S_x is symmetric. So by WSYM,

$$f(S_x) = P_x = \{(1,x),(x,1)\}. \tag{1}$$

Second, consider the bargaining set $\operatorname{conv} S_x$. By (1) and NBR, $P_x \subseteq f(\operatorname{conv} S_x)$. As $\operatorname{conv} S_x$ is convex, $\operatorname{conv} P_x \subseteq f(\operatorname{conv} S_x)$ by CONV. Note that $\operatorname{conv} P_x = \{u \in$

$[0,1]^2 : u_1 + u_2 = 1 + x\}$ and $\operatorname{conv} S_x = \{u \in [0,1]^2 : u_1 + u_2 \leq 1 + x\}$. Hence, $\operatorname{conv} P_x$ is the set of Pareto-optimal points in $\operatorname{conv} S_x$. Hence, by PO,

$$f(\operatorname{conv} S_x) = \left\{u \in [0,1]^2 : u_1 + u_2 = 1 + x\right\}. \tag{2}$$

See Figure 4 for an illustration. The dashed line indicates the solution.

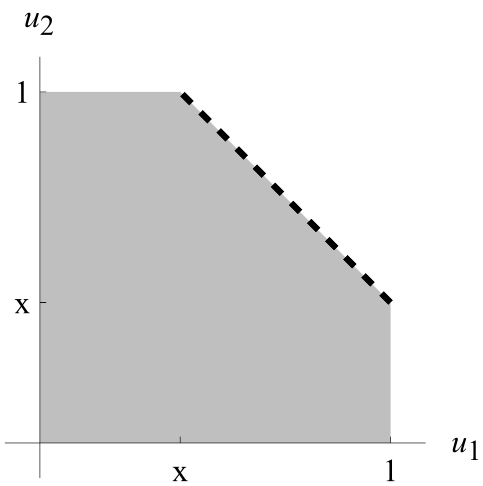


Figure 4: $\operatorname{conv} S_x$.

Third, consider any bargaining set $S \in \mathcal{S}$ where $m(S) = (1,1)$. Let $x^* := \max_{u \in S}(u_1 + u_2) - 1$, which must exist due to our assumption of compactness. Note that $S \subseteq \operatorname{conv} S_{x^*}$ and that $S \cap f(\operatorname{conv} S_{x^*})$ is non-empty. Hence, (2) and WIIA imply

$$f(S) = S \cap f(\operatorname{conv} S_{x^*}) = \arg\max_{u \in S}(u_1 + u_2). \tag{3}$$

Fourth, note that for any $S \in \mathcal{S}$ there exists a bargaining set $S' \in \mathcal{S}$ with $m(S') = (1,1)$ and a positive linear transformation α such that $\alpha(S) = S'$. Then by (3) and INV,

$$f(S) = \arg\max_{u \in S} \left(\frac{u_1}{m_1(S)} + \frac{u_2}{m_2(S)}\right). \tag{4}$$

$\square$

Above we have identified axioms that characterize the RU bargaining solution. Next we show that these axioms are independent. We drop each of the axioms in Theorem 1 and show that there is a solution, other than the RU solution, that satisfies the remaining axioms.

(1) The Nash solution satisfies all axioms but NBR.

(2) Consider the solution which selects all maximal elements of the leximax preorder whenever $m(S) = (1, ..., 1)$. The solution for $m(S) \neq (1, ..., 1)$ is given by INV. This solution satisfies all axioms but CONV.

(3) $f(S) = \{(0, ..., 0)\}$ for all $S \in \mathcal{S}$ satisfies all axioms but PO.

(4) The utilitarian solution (Section 4.2) satisfies all axioms but INV.

(5) The asymmetric relative utilitarian solution (Section 4.1) satisfies all axioms but WSYM.

(6) Let $\omega_i(S) = 1 + \max\{u_i : u \in S \text{ and } u_{i+1} = m_{i+1}(S)\}m_i(S)^{-1}$ if $i < n$ and $\omega_n(S) = 1 + \max\{u_n : u \in S \text{ and } u_1 = m_1(S)\}m_n(S)^{-1}$. Then the solution $f(S) = \arg\max_{u \in S} \sum_{i \in N} \omega_i(S)m_i(S)^{-1}u_i$ satisfies all axioms but WIIA.

This proves that the axioms are independent.

3 Implementation

In the previous section we have provided a normative theory on how a group should come to a solution in a bargain problem. However, how a group will arrive at a solution depends on the game form that describes the bargaining process. In the spirit of the Nash program (Nash, 1953), we implement the RU bargaining solution in subgame perfect equilibrium (SPE), meaning we identify a game form where the the RU bargaining solution arises as the SPE outcome. In line with the existing literature on implementation of the Nash bargaining solution (Nash, 1953; Rubinstein, 1982; Binmore et al., 1986; Herrero, 1989; Howard, 1992; Chae and Yang, 1994; Krishna and Serrano, 1996; Trockel, 2000; Miyagawa, 2002; Trockel, 2002; Güth et al., 2004; Gómez, 2006; Britz et al., 2010; Okada, 2010; Anbarci and Sun, 2013; Britz et al., 2014; Abreu and Pearce, 2015; Qin et al., 2019; Hagiwara, 2020; Hu and Rocheteau, 2020; Harstad, 2023) and the Kalai-Smorodinsky bargaining solution (Moulin, 1984; Trockel, 1999; Miyagawa, 2002; Haake, 2009; Anbarci and Boyd, 2011; Hagiwara, 2020), we assume that the players have full information, i.e. know the utility functions of the other players.

The bargaining protocol we propose can be used by an arbitrator, who wants to bring about a desirable outcome, but who does not know the utility functions of the agents. For applicability in actual bargaining situations, it is important that the game is simple and intuitive. For instance, we would not be satisfied with a game where individuals have to report the state of the world, i.e. the entire utility function of every player, as in Moore and Repullo (1988). Note that, as in the previous section, we allow for non-convex bargaining problems, such that multiple alternatives can be optimal. This complicates the game somewhat, as players have to coordinate on one of multiple equilibria.

In the next section we outline the basic setting.

3.1 Preliminaries

Let A be a set of alternatives. We designate one alternative in A as the disagreement alternative and denote it by a_{dis}. Let Θ be a set of states. For every $\theta \in \Theta$ and $i \in N$, let $u_i^\theta : A \to [0,1]$ denote Player i's vNM utility function over A, normalized such that $u_i^\theta(a_{\text{dis}}) = 0$ and $\max_{a \in A} u_i^\theta(a) = 1$. For every $\theta \in \Theta$, let S_θ denote the bargaining set associated with θ, i.e. $S_\theta := \{(u_1^\theta(a), ..., u_n^\theta(a)) : a \in A\}$. We assume that for each $\theta \in \Theta$, $S_\theta \in \mathcal{S}$. For any $\theta \in \Theta$, we say that $a \in A$ is *RU-optimal* under θ if $(u_1^\theta(a), ..., u_n^\theta(a)) \in f^{\text{RU}}(S_\theta)$. For a game form g and any $\theta \in \Theta$, we denote by (g, θ) the game with the game form g and players' preferences according to u_1^θ to u_n^θ. We say that a game form g *fully implements* the RU bargaining solution in SPE if for every $\theta \in \Theta$ and $a \in A$, a is RU-optimal under θ if and only if a is an SPE outcome of (g, θ). Unfortunately, full implementation is not possible, which we discuss in Section 3.3. Instead, we fully implement the subset of RU-optimal alternatives that are strictly better than a_{dis} for every player. Note that this *weakly implements* the RU bargaining solution, meaning every SPE outcome is RU-optimal. We make two assumptions on Θ.

Assumption A1. For every $\theta \in \Theta$, there exists an $a \in A$ such that a is strictly RU-optimal under θ.

Assumption A2. For every $\theta \in \Theta$ and $i \in N$, there exists a $b_i^\theta \in A$ such that $u_i^\theta(b_i^\theta) = 1$ and $u_j^\theta(b_i^\theta) = 0$ for all $j \in N \setminus \{i\}$.

Next, we present the game form.

3.2 The Game

We now define the game form that fully implements the set of strictly RU-optimal outcomes. We denote this game form by g^*. The game form has two stages, an initial stage and, depending on the actions in the initial stage, an approval stage.

Initial Stage: Each player simultaneously makes a proposal (a, p), consisting of an alternative $a \in A$ and a list of n strictly positive probabilities $p \in (0, 1]^n$. We distinguish three cases:

(1) All players make the same proposal (a, p). Then a is implemented.

(2) There are exactly two distinct proposals (a, p) and (a', p').

  (2a) If $\sum_{i\in N} p_i = \sum_{i\in N} p'_i$, then a_{dis} is implemented.

  (2b) If $\sum_{i\in N} p_i \neq \sum_{i\in N} p'_i$, then go to the *approval stage*.

(3) There are three or more distinct proposals. We implement the alternative of the proposal with the highest sum of probabilities that lies below n. In case of a tie, one of those proposals is chosen at random.

Approval Stage: Let (a, p) and (a', p') denote the two distinct proposals of the initial stage and assume without loss of generality that $\sum_{i\in N} p_i > \sum_{i\in N} p'_i$. Each player sequentially decides between *accept* and *reject*. If all players accept, then a is implemented. If Player i rejects, then with probability p_i, Player i can choose an alternative and with probability $1 - p_i$, a_{dis} is implemented. Players who proposed (a', p') get to choose first, otherwise the order is determined randomly. Figure 5 illustrates the approval stage for three players, where Player 1 and 3 proposed (a, p) and Players 2 proposed (a', p').

Now that we have described the game g^*, we can state the following theorem.

Theorem 2. If A1 and A2 are satisfied, then for every $\theta \in \Theta$ and $a \in A$, a is a SPE outcome of (g^*, θ) if and only if a is strictly RU-optimal under θ.

We sketch the proof here and provide a formal proof of the theorem in Appendix B. We fix some $\theta \in \Theta$ and omit it from now on. First, consider the approval stage and let (a, p) denote the proposal with the higher sum of probabilities. Note that Player

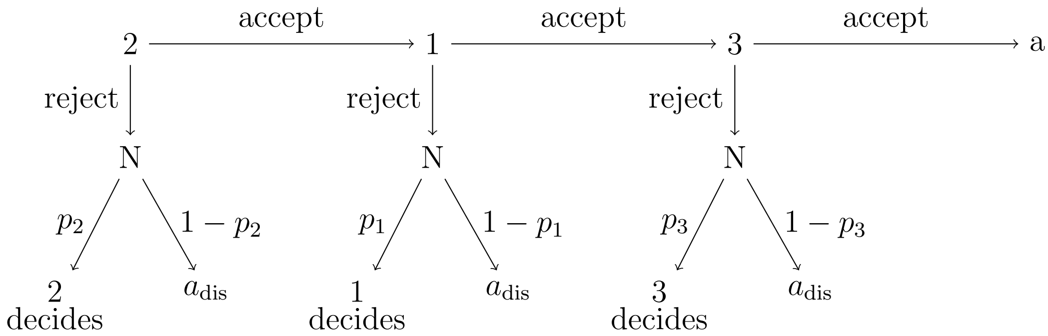


Figure 5: Example of approval stage.

i's expected utility of rejecting is $p_i \max_{b \in A} u_i(b) + (1 - p_i)u_i(a_{\text{dis}}) = p_i$. Hence, if $p_i > u_i(a)$ for all $i \in N$, then the unique SPE is that all players accept and a is implemented. However, if $p_i < u_i(a)$ for some $i \in N$ then some player will reject. Hence, one can interpret the vector p as a *report* about the utility of a for each player and the approval stage as a test of whether this report was truthful. We say that the report p of a proposal (a, p) is *inflated* if $\sum_{i \in N} p_i > \sum_{i \in N} u_i(a)$ and we say that it is *truthful* if $p = u(a)$.

Next, consider the initial stage and assume that all players propose $(a, u(a))$ for some strictly RU-optimal alternative a. We show that no deviation (a', p') is profitable for Player j. There are three options for Player j, to report a higher sum, an equal sum or a lower sum of probabilities. Reporting an equal sum would lead to a_{dis}. Reporting a higher sum would mean that a' is put to the test in the approval stage and that j will decide last. However, because there is no alternative with a higher sum of utilities than a, p' is necessarily inflated and a' will be rejected by some player before j. Reporting a lower sum would mean that a is put to the test in the approval stage and that j will decide first. However, since the report of the proposal $(a, u(a))$ was truthful, there is an equilibrium where all accept and a is implemented anyways. In conclusion, there is no profitable deviation for Player j.

Conversely, consider the case where all players propose (a, p), but a isn't strictly RU-optimal. If p is inflated, then there must be one Player j for whom $p_j > u(a)$. This player can then put (a, p) to the test by deviating to a report with a lower sum. Player j is allowed to decide first and rejects. If p is not inflated, then there is some player who would prefer a strictly RU-optimal alternative a' to a. This player can propose

a' and report a probability for each player that is slightly below the true utility of a'. Then a' is put to the test in the approval stage and every player accepts.

Above we have shown that a non-optimal alternative cannot arise in equilibrium through unanimous proposal, i.e. Case (1). However, we have yet to show that such an alternative cannot arise in equilibrium via Cases (2b) or (3). Note that we have designed Case (3) similar to the integer game in Moore and Repullo (1988), such that it cannot arise in equilibrium. A player can always "outbid" the others by choosing an even higher sum below n. Furthermore, with three or more players, one can deviate from Case (2b) to induce Case (3) and implement their most preferred alternative. Showing that no sub-optimal equilibrium exists for two players is more involved and we leave this to the formal proof in the appendix.

Next, we return to the impossibility of full implementation.

3.3 Full Implementation

It is not possible to fully implement the RU bargaining solution with a domain as general as $\mathcal{S}$.

Proposition 2. Let $n = 2$ and assume that for every $S \in \mathcal{S}$ there exists a $\theta \in \Theta$ such that $S_\theta = S$. Then there doesn't exist an extensive game form g that fully implements the RU bargaining solution.

A formal proof is in Appendix C. The intuition for this result goes as follows. Consider a two agent bargaining problem $S \in \mathcal{S}$ where both $(1, 0)$ and $(0, 1)$ are in $f^{\mathrm{RU}}(S)$. An example of such a problem is the division of a dollar among risk neutral agents, where any efficient division is RU-optimal, including allocating the entire dollar to one of the agents. Now consider a state θ such that $S_\theta = S$ and a game form g that implements the solution. Then there must exist two strategy profiles s^+, s^- that are SPE of (g, θ) and where $u^\theta(s^+) := (u_1^\theta(s^+), u_2^\theta(s^+)) = (1, 0)$ and $u^\theta(s^-) = (0, 1)$. Since by assumption, 0 is the minimal utility for both players, $u_1^\theta(s_1, s_2^-) = 0$ for all s_1 and $u_2^\theta(s_1^+, s_2) = 0$ for all s_2. This in turn implies that (s_1^+, s_2^-) is a Nash equilibrium with pay-offs $(0, 0)$. In the formal proof we then use (s_1^+, s_2^-) to construct an SPE with pay-offs $(0, 0)$. Since, $(0, 0)$ is not in $f^{\mathrm{RU}}(S)$, the RU bargaining solution cannot be fully implemented.

We feel that the division of a dollar among two risk neutral agents is a canonical problem that the implementation should be able to address. Hence, we did not want to rule out such problems, for instance by assuming that the bargaining set is strictly convex as in Miyagawa (2002) and Hagiwara (2020). Instead, we decided to exclude solutions that assign a pay-off of 0 to one of the players.

4 Other Rational Solutions

4.1 Asymmetric Relative Utilitarian Solution

In Section 2 we have imposed a symmetry axiom as a normative requirement to treat all agents of the group fairly. In applications however, we might want to take into account that individuals have different bargaining power. This can lead, in otherwise symmetric situations, to asymmetric outcomes. It is therefore of interest to generalize a given bargaining solution to a $n-1$ parameter family of solutions, where each parameter is a weight on an individual's utility, representing their bargaining power. Asymmetric generalizations have been provided for the Nash solution (Harsanyi and Selten, 1972; Kalai, 1977) and the Kalai-Smorodisnky solution (Dubra, 2001). In the following we provide the first generalization of the relative utilitarian solution. We say that f is an *asymmetric relative utilitarian solution* if there exists $(\mu_1, ..., \mu_n) \in (0,1)^n$ with $\sum_{i\in N} \mu_i = 1$ such that for every $S \in \mathcal{S}$,

$$f(S) = \arg\max_{u\in S} \sum_{i\in N} \mu_i \frac{u_i}{m_i(S)}.$$

We denote this solution by f^{ARU}. Unfortunately, for the characterization of f^{ARU} it doesn't suffice to merely drop the symmetry axiom.

Proposition 3. There exists an f such that $f \neq f^{\text{ARU}}$ and f satisfies NBR, CONV, PO, INV and WIIA.

We prove the proposition by providing two solutions as counter-examples. Each of these solutions can be ruled out by an additional axiom. These two additional axioms, together with NBR, CONV, PO, INV and WIIA then characterize f^{ARU}. For ease of exposition, we provide these counter-examples for $n = 2$.

The first counter-example is the solution f', which selects among the (symmetric) relative utilitarian outcomes the one that most benefits Agent 1, formally

$$f'(S) = \arg\max_{u \in f^{\mathrm{RU}}(S)} u_1.$$

First, we show that indeed the axioms, as stated in Proposition 3, are satisfied. It is easy to see that PO, INV and WIIA are satisfied. Since the solution is a singleton, CONV is trivially satisfied. While convexification can increase the set of relative utilitarian outcomes, it cannot change the best relative utilitarian outcome of each agent. Hence, NBR is satisfied as well. Next, we identify an axiom that rules out such a solution. Note that the solution is discontinuous, in the sense of vNM continuity. To see this, consider the following example, illustrated by Figure 6. An arbitrator has

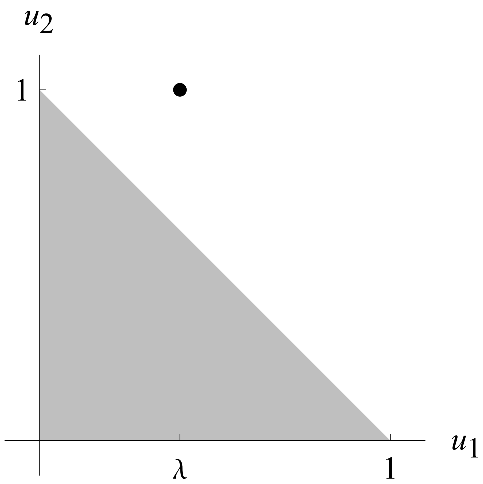


Figure 6: Division of a dollar, with the risky technology at $(\lambda, 1)$.

to divide a dollar among two risk neutral agents. In addition, the arbitrator could choose to invest the dollar into a risky technology, which gives one dollar to Agent 2 for sure and with probability λ one dollar to Agent 1. If $\lambda = 1$, then the technology is Pareto-dominant and selected by f'. If $\lambda = 0$, the technology is just as good as giving the dollar to Agent 2. Under f', the arbitrator strictly prefers to give the dollar to Agent 1. Continuity would require that for some λ, the arbitrator is indifferent between the technology and giving the dollar to Agent 1, such that both are selected by the solution. However, such a λ does not exist under f', as the solution uniquely selects the technology for every $\lambda > 0$. We impose the following axiom to ensure that a bargaining solution is continuous.

Axiom C (Continuity)**.** For every $S \in \mathcal{S}$ and $u \in S \setminus f(S)$, there exists a $\lambda \in [0,1]$ such that

$$f(S \cup \{\lambda m(S) + (1-\lambda)u\}) = f(S) \cup \{\lambda m(S) + (1-\lambda)u\}.$$

The axiom generalizes our previous example. Consider an arbitrary bargaining set S and some point u in S that is not selected by the solution. Add a convex combination $\lambda m(S) + (1-\lambda)u$ between u and the utopian point $m(S)$ to the original bargaining set. For $\lambda = 1$, any solution satisfying PO must select the convex combination, as it is equal to the utopian point. For $\lambda = 0$, the convex combination is equal to u and is therefore not selected by the solution. Axiom C then states that for some λ between 0 and 1, the convex combination must be in the solution, together with the solution of the original bargaining set S.

The second counter-example is a solution that can be described by linear, but non-parallel indifference curves. See Figure 7 for an illustration. The solution selects the

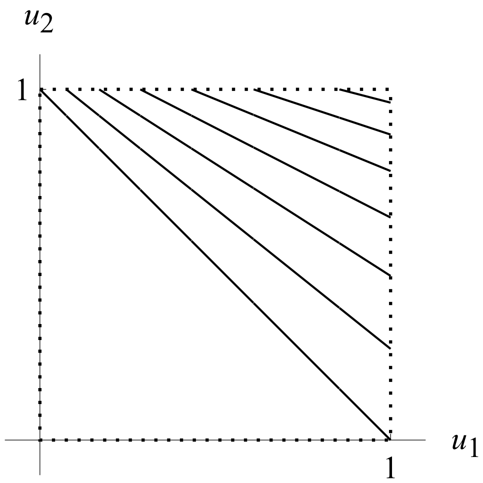


Figure 7: Non-parallel indifference curves.

utility vectors on the highest indifference curve. These indifference curves are the same across all $S \in \mathcal{S}$ with $m(S) = (1,1)$. Solutions for bargaining sets with $m(S) \neq (1,1)$ are derived from INV. We call this solution f''. INV is then satisfied by assumption. It is easy to see that PO and WIIA are satisfied. Finally, CONV and NBR are satisfied because indifference curves are linear. Note that the fanning-out of indifference curves is reminiscent of weighted expected utility (Chew and MacCrimmon, 1979; Chew, 1983). Such risk preferences violate the vNM independence axiom. Similarly, we can rule out non-parallel indifference curves through an independence axiom.

Axiom I (Independence)**.** For every $S \in \mathcal{S}$, if $u, v \in f(S)$ then for any $x \in [0, 1]$,

$$\{xm(S)+(1-x)u, xm(S)+(1-x)v\} \subseteq f(S \cup \{xm(S)+(1-x)u, xm(S)+(1-x)v\}).$$

The axiom captures vNM independence in the bargaining setting. Consider any bargaining set, where the solution contains at least two points u and v. It is as if the group was indifferent between these points. Now add two lotteries l_u and l_v that with probability x give the utopian point $m(S)$ and otherwise u, in case of l_u, or v, in case of l_v. The possibility of the irrelevant alternative $m(S)$ in both l_u and l_v does not change the relative desirability between the two. Hence, both l_u and l_v must be in the solution of the new bargaining set.

Imposing Axioms C and I in addition, leads to the asymmetric relative utilitarian bargaining solution.

Theorem 3. f satisfies NBR, CONV, PO, INV, WIIA, I and C if and only if $f \equiv f^{\text{ARU}}$.

We prove the theorem in Appendix D.

4.2 Utilitarian Solution

In the conventional understanding of the bargaining context, utilities derive from individuals' vNM preferences over the available alternatives. Since a utility representation of a vNM preference relation is unique only up to a positive affine transformation, the invariance axiom ensures that the solution is insensitive to the choice of the representation. However, in alternative scenarios, the utility scale might convey information. For instance, utilities might represent an individual's willingness to pay to bring about a given social alternative. Consider such a setting and think of the two agents who bargain over a single indivisible item. If Agent 1 values the item at \$100 and Agent 2 at \$50, it seems reasonable to award the item to Agent 1. Under the invariance axiom however, this bargaining problem should be treated identically to one where both value the item at \$50. In order to take the absolute scale of valuations into account, INV must be dropped. Previously we argued that IIA imposes too stringent demands on the solution, when utility scales lack significance. However, if utility scales are meaningful, IIA can be assumed in its full strength. Adopting these two modifications leads to the *utilitarian bargaining solution* (Myerson, 1981).

Theorem 4. f satisfies NBR, CONV, PO, WSYM and IIA if and only if for every $S \in \mathcal{S}$,

$$f(S) = \arg\max_{u \in S} \sum_{i \in N} u_i.$$

Proof. Follow the first three steps of the proof of Theorem 1 to find that $f(S) = \arg\max_{u \in S} \sum_{i \in N} u_i$ whenever $m(S) = (1, ..., 1)$. Note that an analogous argument can be made when $m(S) = (x, ..., x)$ for any $x > 0$. Then by IIA, $f(S) = \arg\max_{u \in S} \sum_{i \in N} u_i$ even if $m(S) \neq (x, ..., x)$. $\square$

Note that if utilities are indeed valuations, the utilitarian sum is identical to economic surplus, an ubiquitous measure of welfare. Hence, a group that bargains rationally will also bargain welfare-optimally in the typical sense. Conversely, our axiomatization provides justification for the use of economic surplus as a welfare measure.

5 Conclusion

This paper considers the implications of the vNM axioms on fair bargaining. We find that these axioms are inconsistent with the prominent bargaining solutions by Nash (1950) and Kalai and Smorodinsky (1975) and offer an alternative, the relative utilitarian bargaining solution. On first sight, this solution might not seem very fair. For instance, when dividing a dollar among risk neutral agents, giving the entire dollar to one of the agents is permitted by the solution. However, fairness can be understood as treating symmetric agents in the same way, giving them them equal consideration. This is captured by the weak symmetry axiom. If giving the entire dollar to Agent 1 is permitted, then it must also be permitted to give the entire dollar to Agent 2. However, we do not insist on equality. This allows us to make trade-offs. The relative utilitarian bargaining solution is willing to improve one agent's utility if it comes at a smaller cost to another agent.

Appendix A

We begin with a definition of S_x and P_x for a general $n \in \mathbb{N}$. For this purpose, the following notation is introduced. For $x \in [1, n]$, let $\lfloor x \rfloor := \max\{k \in \{1, ..., n\} : k \leq x\}$, meaning $\lfloor x \rfloor$ is x rounded down to the closest integer. For any $u \in \mathbb{R}^n$, let $\pi(u) \subset \mathbb{R}^n$ denote the set containing u and all its permutations. For any $k \in \{0, ..., n\}$, let $g(k)$ denote the sequence of length n where $g_i(k) = 1$ if $i \leq k$ and $g_i(k) = 0$ otherwise. Hence, $g(0) = (0, ..., 0)$, $g(1) = (1, 0, ..., 0)$ and so on. For any $x \in [1, n]$, let $h(x)$ denote the sequence of length n where $h_i(x) = 1$ if $i \leq x$, $h_i(x) = x - \lfloor x \rfloor$ if $i = \lfloor x \rfloor + 1$ and $h_i(k) = 0$ otherwise. For any $x \in [1, n]$, let

$$P_x := \pi(h(x)), \qquad\qquad S_x := P_x \cup \bigcup_{k=0}^{\lfloor x \rfloor} \pi(g(k)).$$

The following lemma identifies $\operatorname{conv} S_x$ and $\operatorname{conv} P_x$.

Lemma 1. For any $x \in [1, n]$, $\operatorname{conv} S_x = \{u \in [0,1]^n : \sum_{i \in N} u_i \leq x\} =: R_x$ and $\operatorname{conv} P_x = \{u \in [0,1]^n : \sum_{i \in N} u_i = x\} =: Q_x$.

Proof. An element $u \in S \subset \mathbb{R}^n$ is an *extreme point* of S if there doesn't exist $v, w \in S$ and $\lambda \in (0, 1)$ such that $v \neq w$ and $u = \lambda v + (1 - \lambda) w$. In the following we show that S_x (resp. P_x) contains all extreme points of R_x (resp. Q_x). Then, by the Krein–Milman theorem, the lemma follows from the fact that $S_x \subseteq R_x$ and $P_x \subseteq Q_x$.

First, we show that an extreme point u of R_x (resp. Q_x) can have at most one coordinate u_j such that $0 < u_j < 1$. Consider $u \in R_x$ (resp. Q_x) with $0 < u_j < 1$ and $0 < u_k < 1$ for some $j \neq k$. Then there exists $v, w \in R_x$ (resp. Q_x) and $\varepsilon > 0$ such that $v_i = w_i = u_i$ whenever $i \notin \{j, k\}$ and

$$\begin{aligned} v_j &= u_j + \varepsilon, & w_j &= u_j - \varepsilon, \\ v_k &= u_k - \varepsilon, & w_k &= u_k + \varepsilon. \end{aligned} \tag{5}$$

Since $\frac{1}{2}v + \frac{1}{2}w = u$, u is not an extreme point of R_x (resp. Q_x).

Second, we show that if there is an extreme point of R_x with exactly one coordinate u_j such that $0 < u_j < 1$ then $u \in \pi(h(x))$. Assume $u \notin \pi(h(x))$ and $0 < u_j < 1$. Then there exists $v, w \in R_x$ and $\varepsilon > 0$ such that $v_i = w_i = u_i$ whenever $i \neq j$ and

$$v_j = u_j + \varepsilon, \qquad\qquad w_j = u_j - \varepsilon.$$

Since $\frac{1}{2}v + \frac{1}{2}w = u$, u is not an extreme point of R_x.

By the above arguments, the only candidates for extreme points of R_x (resp. Q_x) are the points in S_x (resp. P_x).

□

We now present the proof of Theorem 1. The proof is nearly identical to the sketch in Section 2. Nevertheless, we state the proof for sake of completeness.

Proof of Theorem 1

First, consider a bargaining set S_x for some $x \in [1, n]$. Note that P_x is the set of Pareto optimal points in S_x. Hence, by PO, $f(S_x) \subseteq P_x$. Furthermore, note that S_x is symmetric. So by WSYM,

$$f(S_x) = P_x. \tag{6}$$

Second, consider the bargaining set $\operatorname{conv} S_x$. By (6) and NBR, $P_x \subseteq f(\operatorname{conv} S_x)$. As $\operatorname{conv} S_x$ is convex, $\operatorname{conv} P_x \subseteq f(\operatorname{conv} S_x)$ by CONV. From Lemma 1 we can see that $\operatorname{conv} P_x$ is the set of Pareto-optimal points in $\operatorname{conv} S_x$. Then by PO,

$$f(\operatorname{conv} S_x) = \left\{ u \in [0,1]^n : \sum_{i \in N} u_i = x \right\}. \tag{7}$$

Third, consider any bargaining set $S \in \mathcal{S}$ where $m(S) = (1, ..., 1)$. Let $x^* := \max_{u \in S} \sum_{i \in N} u_i$, which must exist due to our assumption of compactness. Note that $S \subseteq \operatorname{conv} S_{x^*}$ and that $S \cap f(\operatorname{conv} S_{x^*})$ is non-empty. Hence, (7) and WIIA imply

$$f(S) = S \cap f(\operatorname{conv} S_{x^*}) = \arg\max_{u \in S} \sum_{i \in N} u_i. \tag{8}$$

Fourth, note that for any $S \in \mathcal{S}$ there exists a bargaining set $S' \in \mathcal{S}$ with $m(S') = (1, ..., 1)$ and a positive linear transformation α such that $\alpha(S) = S'$. Then by (8) and INV,

$$f(S) = \arg\max_{u \in S} \left(\sum_{i \in N} \frac{u_i}{m_i(S)} \right). \tag{9}$$

Appendix B

Proof of Theorem 2

We have already shown in the Section 3.2 that for every strictly RU-optimal alternative there exists an SPE with this alternative is the outcome. Here, we prove the other direction, namely that every SPE outcome is strictly RU-optimal. We fix some $\theta \in \Theta$ and omit it from now on. Let a denote some alternative that isn't strictly RU-optimal.

First, consider Case (1) of the initial stage, where a is implemented through some unanimous proposal (a, p). If $\sum_{i\in N} p_i > \sum_{i\in N} u_i(a)$ then there exists a Player j for whom $p_j > u_j(a)$. This player can deviate to some proposal (a', p') with $\sum_{i\in N} p'_i < \sum_{i\in N} p_i$ and be the first to reject in the approval stage, which gives an expected utility of p_j. If $\sum_{i\in N} p_i \leq \sum_{i\in N} u_i(a)$, then there is at least one player who strictly prefers some strictly RU-optimal alternative a'. Then this player can deviate to $(a', (u_1(a') - \varepsilon, ..., u_n(a') - \varepsilon))$ for ε sufficiently small such that $\sum_{i\in N} p_i < \sum_{i\in N} (u_i(a') - \varepsilon)$. Then a' is put to the test in the approval stage and the unique SPE outcome is that all players accept and a' is implemented.

Second, consider Case (2a), where there are two distinct proposals (a', p') and (a'', p'') with $\sum_{i\in N} p'_i = \sum_{i\in N} p''_i$, leading to a_{dis}. If $n \geq 3$, then some Player i can deviate in the initial stage to bring about Case (3) and choose b_i. So assume $n = 2$. If $p'_1 + p'_2 = p''_1 + p''_2 > 1$, then a player would be better off to propose a lower sum, be the first to choose in the approval stage and then reject. If $p'_1 + p'_2 = p''_1 + p''_2 \leq 1$, then for some strictly RU-optimal alternative a''' a player can propose $(a''', (u_1(a''') - \varepsilon, u_2(a''') - \varepsilon))$ for ε sufficiently small, leading to a'.

Third, consider Case (2b), where a is implemented through acceptance of the proposal (a, p) by all players in the approval stage. Unless only a single Player j has made the other proposal (a', p'), any player can bring about Case (3) and choose a more preferred alternative. Hence, assume only single Player j has made the other proposal and a is among the best alternatives for all players other than j. There must be some strictly RU-optimal alternative a'' that is preferred to a by j. Since a is unanimously approved, $\sum_{i\in N} p_i \leq \sum_{i\in N} u_i(a)$. Player j can deviate to $(a'', (u_1(a'') - \varepsilon, ..., u_n(a'') - \varepsilon))$ for ε sufficiently small, leading to a''.

Fourth, consider Case (2b), where a is chosen after some player rejects in the approval stage. Let $\Sigma := \sum_{i\in N} u_i(b)$ for any RU-optimal alternative b. First, consider $n = 2$. Let (a', p') be Player 1's proposal and (a'', p'') be Player 2's proposal. Without loss of generality, assume that $p'_1 + p'_2 > p''_1 + p''_2$, such that Player 2 decides first in the approval stage. Player 2 rejects and with probability p_2' chooses a with $u(a) = (x, 1)$ for some $x \in [0, 1)$. This gives expected utility p_2' to Player 2 and $p_2'x$ to Player 1. Consider the case where $p_1''+p_2'' \geq \Sigma$. Player 1 has an incentive to deviate to a proposal with a lower sum, unless

$$p_1'' \leq p_2'x. \tag{10}$$

Furthermore, since $p_1'' + p_2'' \geq \Sigma$,

$$p_1'' + p_2'' \geq x + 1. \tag{11}$$

Then (10) and (11) imply $p_2'' = 1$ and either $x = 0$ or $p_2' = 1$. But $x = 0$ is not possible since this would imply $p_1'' = 0$ and this is not in the strategy-space of g^*. Hence $p_2' = 1$ and $p_1'' = x$. Since we consider the case $p_1''+p_2'' \geq \Sigma$ and have found $p'' = u(a)$, it must be that a is RU-optimal. Furthermore, since $p_2' = 1$, a is implemented with certainty. This contradicts the assumption that the SPE outcome is sub-optimal. If $p_1''+p_2'' < \Sigma$, then Player 1 can implement any strictly RU-optimal alternative b by deviating to the proposal $(b, (u_1(b)-\varepsilon, u_2(b)-\varepsilon))$ for ε sufficiently small. The only case in which Player 1 would have no incentive to do so, is if $p_2' = 1$ and if b is the best strictly RU-optimal alternative for Player 1, which contradicts the assumption that the SPE outcome is sub-optimal. This concludes the case $n = 2$. The argument extends to a general n.

Finally, consider Case (3) of the initial stage, where a is the alternative of the proposal with the highest sum strictly below n. Then at least one Player j prefers b_j to a. This player can deviate to a proposal with an even higher sum below n and the alternative b_j.

This concludes the proof of Theorem 2.

Appendix C

We follow the notation of Moore and Repullo (1988). Consider a two player extensive game form g. Let T denote the set of nodes. For any $t \in T$ we denote by $g(t)$ the sub-game starting at node t. For any $t \in T$ and $i \in \{1, 2\}$ we denote by $\sigma_i(t)$ the set of actions of Player i. We assume that $|\sigma_i(t)| \geq 1$ for all $t \in T$. If $|\sigma_i(t)| = 1$, then Player i has no decision at t. If both $|\sigma_1(t)| > 1$ and $|\sigma_2(t)| > 1$ then both players move simultaneously at t. Let $\sigma_i := \bigtimes_{t\in T} \sigma_i(t)$ denote the strategy space of Player i and let

$\sigma = \sigma_1 \times \sigma_2$ denote the set of strategy profiles. For any $s \in \sigma$ and $t \in T$ we denote by $s|t$ the part of s that specifies the strategy profile for the game $g(t)$. For any $s \in \sigma$ and $t \in T$ we denote by $s(t) \in \sigma_1(t) \times \sigma_2(t)$ the action pair that s prescribes for the node t. Terminal nodes are alternatives in A. For any $s \in \sigma$ and $\theta \in \Theta$, $u^\theta(s) = (u_1^\theta(s), u_2^\theta(s))$ denotes the utility vector of the terminal node that results from s. For any $s \in \sigma$ and $t \in T$ we write $u^\theta(s|t)$ to denote the utility vector of the terminal node that is reached when starting at t and playing according to s.

Proof of Proposition 2

Fix some $\theta \in \Theta$ such that $f^{\mathrm{RU}}(S_\theta)$ contains both $(1,0)$ and $(0,1)$. In the following we omit θ on the individual utility functions and the game. We have already established in the main section that there must be $s^+, s^- \in \sigma$ such that both are SPE of g with $u(s^+) = (1,0)$ and $u(s^-) = (0,1)$ and that $(s_1^+, s_2^-) =: s^0$ is a NE with $u(s^0) = (0,0)$. In the following we construct $s^* \in \sigma$ such that s^* is an SPE with $u(s^*) = (0,0)$.

Let $t_0, t_0, ..., t_k, t_{k+1}$ denote the nodes of the equilibrium path of s^0, where t_0 is the initial node of g and t_{k+1} is a terminal node of g associated with the pay-off $(0,0)$. Consider the second to last node t_k. Let $(x_k, y_k) \in \sigma_1(t_k) \times \sigma_2(t_k)$ denote the action pair that leads to t_{k+1}, formally $t_{k+1} = (t_k, (x_k, y_k))$. If there are only terminal nodes succeeding t_k, then $g(t_k)$ is a one-stage game and $s^0|t_k$ is not only a NE of $g(t_k)$ but also an SPE. Hence, $s^*|t_k = s^0|t_k$ ensures that $s^*|t_k$ is an SPE of $g(t_k)$ with outcome $(0,0)$. If there are non-terminal nodes succeeding t_k, then we construct $s^*|t_k$ as follows.

Choose $s^*(t_k) = (x_k, y_k)$. For any (x_k, y) with $y \in \sigma_2(t_k)$ choose $s^*|(t_k, (x_k, y)) = s^+|(t_k, (x_k, y))$. This ensures that $s^*|(t_k, (x_k, y))$ is an SPE of $g(t_k, (x_k, y))$ for all $y \in \sigma_2(t_k)$. Furthermore, it must be that $u_2(s^*|(t_k, (x_k, y))) = 0$ for all $y \in \sigma_2(t_k)$, because $(t_k, (x_k, y))$ can be reached by a unilateral deviation of Player 2 in the strategy profile s^+. Similarly, choose $s^*|(t_k, (x, y_k)) = s^-|(t_k, (x, y_k))$ for all $x \in \sigma_1(t_k)$. For $s^*|(t_k, (x, y))$ such that neither $x = x_k$ nor $y = y_k$ choose an arbitrary SPE. Note that $s^*|t_k$ has been constructed such that an SPE is played at all nodes succeeding t_k, such that the outcome is $(0,0)$ and such that no Player has an incentive to deviate at t_k. Therefore, $s^*|t_k$ is an SPE of $g(t_k)$ with outcome $(0,0)$.

Next consider t_l for any $l \in \{0, ..., k-1\}$. Let $(x_l, y_l) \in \sigma_1(t_l) \times \sigma_2(t_l)$ denote the decision that leads to t_{l+1}, formally $t_{l+1} = (t_l, (x_l, y_l))$. Assume $s^*|t_{l+1}$ is an SPE of $g(t_{l+1})$ with outcome $(0,0)$. Choose $s^*(t_l) = (x_l, y_l)$ and construct $s^*|(t_l, (x, y))$ for $(x, y) \neq (x_l, y_l)$ just as before. Then $s^*|t_l$ is an SPE of $g(t_l)$ with outcome $(0,0)$. By induction, s^* is an SPE of g with outcome $(0,0)$. This concludes the proof.

Appendix D

We prove the Theorem 3 for $n = 2$. The proof for general n follows similarly as that of Theorem 1. First, consider the bargaining set $S = \{(0,0), (1,0), (0,1)\}$. By PO, there are three possible cases.

Case 1: $f(S) = \{(1,0)\}$

Case 2: $f(S) = \{(0,1)\}$

Case 3: $f(S) = \{(1,0), (0,1)\}$

Assume Case 1 holds true. By NBR, $(1,0) \in f(\operatorname{conv} S)$ and by WIIA, $(0,1) \notin f(\operatorname{conv} S)$. By C, there exists a $\lambda \in [0,1]$ such that $f(\operatorname{conv} S \cup \{(\lambda,1)\}) = f(\operatorname{conv} S) \cup \{(\lambda,1)\}$. Note that $\lambda > 0$, as it would otherwise contradict the assumption that $(0,1) \notin f(\operatorname{conv} S)$. Furthermore, note that $\lambda < 1$ as it would otherwise contradict PO. By NBR, $\{(1,0), (\lambda,1)\} \subseteq f(\operatorname{conv}(S \cup \{(\lambda,1)\}))$ and by CONV,

$$f(\operatorname{conv}(S \cup \{(\lambda,1)\})) = \left\{u \in [0,1]^2 : u_1 + (1-\lambda)u_2 = 1\right\}. \tag{12}$$

Second, consider the bargaining set

$$S_x := \operatorname{conv}\left(S \cup \{(\lambda,1)\}\right) \cup \{(1,x), (x + (1-x)\lambda, 1)\}$$

for any $x \in [0,1]$. By I and (12), $\{(1,x), (x + (1-x)\lambda, 1)\} \subseteq f(S_x)$. By NBR, $\{(1,x), (x + (1-x)\lambda, 1)\} \subseteq f(\operatorname{conv} S_x)$ and by CONV and PO,

$$f(\operatorname{conv} S_x) = \left\{u \in [0,1]^2 : u_1 + (1-\lambda)u_2 = 1 + (1-\lambda)x\right\}. \tag{13}$$

Third, consider any bargaining set $S \in \mathcal{S}$ where $m(S) = (1,1)$. Let $x^* := (1-\lambda)^{-1}\left(\max_{u \in S}(u_1 + (1-\lambda)u_2) - 1\right)$, which must exist due to our assumption of compactness. Note that $S \subseteq \operatorname{conv} S_{x^*}$ and that $S \cap f(\operatorname{conv} S_{x^*})$ is non-empty. Hence, (13) and WIIA imply

$$f(S) = S \cap f(\operatorname{conv} S_{x^*}) = \arg\max_{u \in S}(u_1 + (1-\lambda)u_2). \tag{14}$$

Fourth, note that for any $S \in \mathcal{S}$ there exists a bargaining set $S' \in \mathcal{S}$ with $m(S') = (1,1)$ and a positive linear transformation α such that $\alpha(S) = S'$. Then by (14) and INV,

$$f(S) = \arg\max_{u \in S} \left(\frac{u_1}{m_1(S)} + (1-\lambda)\frac{u_2}{m_2(S)} \right). \tag{15}$$

This concludes Case 1. Note that by an analogous argument, we can find an analogous solution for Case 2 and 3. We can summarize the results for the different cases as follows. There exists $(\mu_1, \mu_2) \in (0,1)^2$ where $\mu_1 + \mu_2 = 1$ such that for any $S \in \mathcal{S}$,

$$f(S) = \arg\max_{u \in S} \left(\mu_1 \frac{u_1}{m_1(S)} + \mu_2 \frac{u_2}{m_2(S)} \right). \tag{16}$$

This concludes the proof.

References

Abreu, D. and Pearce, D. (2015). A dynamic reinterpretation of Nash bargaining with endogenous threats. *Econometrica*, 83(4):1641–1655.

Amanatidis, G., Birmpas, G., Filos-Ratsikas, A., and Voudouris, A. A. (2022). Fair division of indivisible goods: A survey. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 5385–5393, Vienna, Austria. International Joint Conferences on Artificial Intelligence Organization.

Anbarci, N. and Boyd, J. H. (2011). Nash demand game and the Kalai–Smorodinsky solution. *Games and Economic Behavior*, 71(1):14–22.

Anbarci, N. and Sun, C.-j. (2013). Asymmetric Nash bargaining solutions: A simple Nash program. *Economics Letters*, 120(2):211–214.

Baris, O. F. (2018). Timing effect in bargaining and ex ante efficiency of the relative utilitarian solution. *Theory and Decision*, 84(4):547–556.

Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 17(2):176–188.

Börgers, T. and Choo, Y.-M. (2017). Revealed relative utilitarianism. CESifo Working Paper Series No. 6613.

Bossert, W. (1994). Rational choice and two-person bargaining solutions. *Journal of Mathematical Economics*, 23(6):549–563.

Brandl, F. (2021). Belief-averaging and relative utilitarianism. *Journal of Economic Theory*, 198:105368.

Britz, V., Herings, P. J.-J., and Predtetchinski, A. (2010). Non-cooperative support for the asymmetric Nash bargaining solution. *Journal of Economic Theory*, 145(5):1951–1967.

Britz, V., Herings, P. J.-J., and Predtetchinski, A. (2014). On the convergence to the Nash bargaining solution for action-dependent bargaining protocols. *Games and Economic Behavior*, 86:178–183.

Budish, E. (2011). The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103.

Cao, X. (1982). Preference functions and bargaining solutions. In *1982 21st IEEE Conference on Decision and Control*, pages 164–171.

Chae, S. and Yang, J.-A. (1994). An n-person pure bargaining game. *Journal of Economic Theory*, 62(1):86–102.

Chew, S. H. (1983). A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the allais paradox. *Econometrica*, 51(4):1065–1092.

Chew, S. H. and MacCrimmon, K. R. (1979). Alpha-nu choice theory: A generalization of expected utility. University of British Columbia Working Paper.

Conley, J. P. and Wilkie, S. (1996). An extension of the Nash bargaining solution to nonconvex problems. *Games and Economic Behavior*, 13(1):26–38.

Denicolò, V. and Mariotti, M. (2000). Nash bargaining theory, nonconvex problems and social welfare orderings. *Theory and Decision*, 48(4):351–358.

Dhillon, A. and Mertens, J.-F. (1999). Relative utilitarianism. *Econometrica*, 67(3):471–498.

Dubra, J. (2001). An asymmetric Kalai–Smorodinsky solution. *Economics Letters.*

Gómez, J. C. (2006). Achieving efficiency with manipulative bargainers. *Games and Economic Behavior*, 57(2):254–263.

Güth, W., Ritzberger, K., and van Damme, E. (2004). On the Nash bargaining solution with noise. *European Economic Review*, 48(3):697–713.

Haake, C.-J. (2009). Two support results for the Kalai–Smorodinsky solution in small object division markets. *Mathematical Social Sciences*, 57(2):177–187.

Hagiwara, M. (2020). Subgame-perfect implementation of bargaining solutions: Comment. *Games and Economic Behavior*, 122:476–480.

Harsanyi, J. C. and Selten, R. (1972). A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information. *Management Science*, 18(5-part-2):80–106.

Harstad, B. (2023). Pledge-and-review bargaining. *Journal of Economic Theory*, 207:105574.

Herrero, M. J. (1989). The Nash program: Non-convex bargaining problems. *Journal of Economic Theory*, 49(2):266–277.

Howard, J. (1992). A social choice rule and its implementation in perfect equilibrium. *Journal of Economic Theory*, 56(1):142–159.

Hu, T.-W. and Rocheteau, G. (2020). Bargaining under liquidity constraints: Unified strategic foundations of the Nash and Kalai solutions. *Journal of Economic Theory*, 189:105098.

Imai, H. (1983). Individual monotonicity and lexicographic maxmin solution. *Econometrica*, 51(2):389.

Kalai, E. (1977). Nonsymmetric Nash solutions and replications of 2-person bargaining. *International Journal of Game Theory*, 6(3):129–133.

Kalai, E. and Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica*, 43(3):513–518.

Kaneko, M. (1980). An extension of the Nash bargaining problem and the Nash social welfare function. *Theory and Decision*, 12(2):135–148.

Karni, E. (1998). Impartiality: Definition and representation. *Econometrica*, 66(6):1405.

Krishna, V. and Serrano, R. (1996). Multilateral bargaining. *The Review of Economic Studies*, 63(1):61–80.

Marchant, T. (2019). Utilitarianism without individual utilities. *Social Choice and Welfare*, 53(1):1–19.

Mariotti, M. (1998a). Extending Nash's axioms to nonconvex problems. *Games and Economic Behavior*, 22(2):377–383.

Mariotti, M. (1998b). Nash bargaining theory when the number of alternatives can be finite. *Social Choice and Welfare*, 15(3):413–421.

Miyagawa, E. (2002). Subgame-perfect implementation of bargaining solutions. *Games and Economic Behavior*.

Moore, J. and Repullo, R. (1988). Subgame perfect implementation. *Econometrica*, 56(5):1191.

Moulin, H. (1984). Implementing the Kalai-Smorodinsky bargaining solution. *Journal of Economic Theory*, 33(1):32–45.

Myerson, R. B. (1981). Utilitarianism, egalitarianism, and the timing effect in social choice problems. *Econometrica*, 49(4):883.

Nagahisa, R.-i. and Tanaka, M. (2002). An axiomatization of the Kalai-Smorodinsky solution when the feasible sets can be finite. *Social Choice and Welfare*, 19(4):751–761.

Nash, J. (1953). Two-person cooperative games. *Econometrica*, 21(1):128–140.

Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2):155–162.

Ok, E. A. and Zhou, L. (1999). Revealed group preferences on non-convex choice problems. *Economic Theory*, 13(3):671–687.

Okada, A. (2010). The Nash bargaining solution in general n-person cooperative games. *Journal of Economic Theory*, 145(6):2356–2379.

Peitler, P. and Schlag, K. H. (2023). Putting Context into Preference Aggregation. Working Paper.

Peters, H. and Wakker, P. (1991). Independence of irrelevant alternatives and revealed group preferences. *Econometrica*, 59(6):1787–1801.

Pivato, M. (2009). Twofold optimality of the relative utilitarian bargaining solution. *Social Choice and Welfare*, 32(1):79–92.

Qin, C.-Z., Tan, G., and Wong, A. C. L. (2019). Implementation of Nash bargaining solutions with non-convexity. *Economics Letters*, 178:46–49.

Rachmilevitch, S. (2019). Egalitarianism, utilitarianism, and the Nash bargaining solution. *Social Choice and Welfare*, 52(4):741–751.

Roth, A. E. (1977). Independence of irrelevant alternatives, and solutions to Nash's bargaining problem. *Journal of Economic Theory*, 16(2):247–251.

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

Sánchez, M. (2000). Rationality of bargaining solutions. *Journal of Mathematical Economics*, 33(4):389–399.

Segal, U. (2000). Let's agree that all dictatorships are equally bad. *Journal of Political Economy*, 108(3):569–589.

Sprumont, Y. (2019). Relative utilitarianism under uncertainty. *Social Choice and Welfare*, 53(4):621–639.

Trockel, W. (1999). Unique Nash implementation for a class of bargaining solutions. *International Game Theory Review*, 01(03n04):267–272.

Trockel, W. (2000). Implementations of the Nash solution based on its Walrasian characterization. *Economic Theory*, 16(2):277–294.

Trockel, W. (2002). A universal meta bargaining implementation of the Nash solution. *Social Choice and Welfare*, 19(3):581–586.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ, USA: Princeton University Press.

Xu, Y. and Yoshihara, N. (2006). Alternative characterizations of three bargaining solutions for nonconvex problems. *Games and Economic Behavior*, 57(1):86–92.

Xu, Y. and Yoshihara, N. (2013). Rationality and solutions to nonconvex bargaining problems: Rationalizability and Nash solutions. *Mathematical Social Sciences*, 66(1):66–70.

Yu, P. L. (1973). A Class of Solutions for Group Decision Problems. *Management Science*, 19(8):936–946.

Zambrano, E. (2016). 'Vintage' Nash bargaining without convexity. *Economics Letters*, 141:32–34.

Zhou, L. (1997). The Nash bargaining theory with non-convex problems. *Econometrica*, 65(3):681.