Columbia University

Department of Economics
Discussion Paper Series

Subgame-Perfect Implementation of Bargaining Solutions

Eiichi Miyagawa

Discussion Paper #:0102-16

Department of Economics
Columbia University
New York, NY 10027

March 2002

Columbia University
Department of Economics Discussion Paper No. 0102-16
Subgame-Perfect Implementation of Bargaining Solutions*
Eiichi Miyagawa**
March 2002

Abstract:

February 1997
Final Version: October 15, 2001

This paper provides simple four-stage game forms that fully implement a large class of two-person bargaining solutions in subgame-perfect equilibrium. The solutions that can be implemented by our game forms are those that maximize a monotonic and quasi-concave function of utilities after normalizing each agent's utility function so that the maximum utility is 1 and the utility of the disagreement outcome is 0. This class of solutions includes the Nash, Kalai{Smorodinsky, and Relative Utilitarian solutions. The game forms have a structure of alternating offers and contain no integer device.

JEL: C72, C78, D70, D74, D82.

Key Words: Mechanism design, Nash program, alternating offers, complete information.

---

* This paper is a revised version of the paper previously entitled \Implementation of the Normalized Utilitarian Bargaining Solution." I would like to thank William Thomson for numerous helpful comments. I am grateful to two anonymous referees for detailed and helpful suggestions. I also thank John Duggan, Herve Moulin, and James Schummer for helpful comments. I also thank Anne Jones for her assistance. All errors are mine.
** Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027, USA. Email: em437@columbia.edu.

1 Introduction

This paper provides game forms that solve the implementation problem for two-person bargaining problems. We exhibit simple four-stage game forms that implement *any* bargaining solution that maximizes a monotonic and quasi-concave function of normalized utilities. The solutions that can be implemented by our game forms include the Nash, Kalai–Smorodinsky, and Relative Utilitarian solutions.

Consider a bargaining problem, where two agents have to choose an alternative from a certain set. They can agree on any alternative in the set, and if they disagree, they obtain a predetermined alternative. To avoid disagreements, they have decided to follow the recommendation of an arbitrator. His recommendation is described by a bargaining solution, which associates an alternative with each profile of preferences.

Unfortunately, agents' preferences are typically unknown to the arbitrator, and agents may not be willing to report their preferences truthfully if it is not in their interests. A way to solve the manipulation problem is to confront the agents with a well-chosen game form (or mechanism). A (sequential) game form specifies when an agent can take action, what actions are available for an agent at each move, and which alternative is selected for a given sequence of actions. Our hope is that there exists a game form whose equilibrium outcome is the solution outcome for the true preferences. When this is the case, the solution is called *implementable*. The idea is that the arbitrator may be able to exploit strategic interactions among the agents to extract their private information.

Whether or not a solution is implementable depends on the type of game forms as well as the equilibrium concept being considered. This paper considers a sequential game form (finite-horizon extensive game form with perfect information) together with subgame-perfect equilibrium.

Our game forms implement a large class of bargaining solutions. To describe this class, we first identify the unique von Neumann–Morgenstern utility function for each agent such that the maximum utility is 1 and the utility of the disagreement outcome is 0. Recall that the solutions proposed by Nash (1950) and Kalai and Smorodinsky (1975) are both defined by maximizing a function of normalized utilities. For the Nash solution, the function to be maximized is the Cobb–Douglas, and for the Kalai–Smorodinsky solution, it is the minimum function. Another interesting solution of this type is what is called the Relative (or Normalized) Utilitarian solution, which maximizes the sum of the agents' normalized utilities. This solution, introduced by Cao (1982), has received increasing attention recently (e.g., Dhillon and Mertens, 1999; Segal, 2000; Sobel, 2001).

We are able to implement *any* bargaining solution that maximizes a monotonic and quasi-concave function of normalized utilities. We show that for any monotonic and quasi-concave function, the associated bargaining solution can be implemented by one of our sequential game forms. All of our game forms are identical except for their second stage, in which the strategy space is given by an indifference curve of the function associated with the solution. The other stages are defined independently of the solution to be implemented.

The literature provides general necessary or sufficient conditions on solutions for them to be implementable, for many equilibrium concepts including subgame-perfect equilibrium.[1] A typical sufficiency proof, following Maskin (1999), provides an algorithm that produces a game form that implements a given solution satisfying certain conditions. The negative side of these general results is that the game forms produced in such proofs are typically complex and not intuitive.

Moore and Repullo (1988) provide a sufficient condition for subgame-perfect implementation. The usefulness of their condition in our context depends on the bargaining setting being considered. Their condition is satisfied by all of our solutions in a certain important setting, but in other settings, it can be violated by all of our solutions. When their condition is violated by our solutions, no other result in the literature proves whether or not our solutions are implementable.

Abreu and Sen (1990) provide a necessary condition for subgame-perfect implementation. It turns out that their condition has no bite in the context of bargaining, since it is satisfied by *any* bargaining solution whether or not it maximizes some function.[2]

We also show that in a canonical bargaining setting, none of our solutions satisfies Maskin's (1999) well-known necessary condition for Nash implementation.

We show not only that our solutions are implementable in subgame-perfect equilibrium, but also that they can be implemented by means of "simple" game forms. Some of the advantages of our game forms are: (i) They contain no "integer (modulo) device." Many game forms that have been constructed in the literature involve each agent announcing an integer as part of his strategy and, for certain strategy profiles, the agent who announces the highest integer gets his most preferred alternative. The use of integer devices makes many results in the literature somewhat unconvincing. As Jackson (1992) demonstrated, a solution may be implementable by means of a game form involving integer devices, but not by means of any "reasonable" game form. (ii) Unlike many game forms in the literature, our game forms do not ask agents to announce preference relations. Such information is hard to communicate in practice. In our game forms, agents announce numbers and alternatives.[3]

Furthermore, our game forms are intuitive. It is partly because the game forms have a structure of alternating offers. In the first part of our game forms, an agent announces a utility vector, which sets the minimum utility levels guaranteed in later stages. The other agent is then allowed to announce a counter-proposal that satisfies a certain condition. In the second part of the game forms, an agent is allowed to announce an alternative, and then the other agent either accepts it or chooses a certain predetermined alternative.

The implementation of bargaining solutions has been studied by various authors. Moulin (1984) and Howard (1992) provide simple game forms that implement the Kalai–Smorodinsky and Nash solutions, respectively, in subgame-perfect equilibrium. Binmore, Rubinstein, and Wolinsky (1986) showed that a version of the alternating-offer game form of Rubinstein (1982) approximately implements the Nash solution in subgame-perfect equilibrium. Conley and Wilkie (1995) provide a similar game form that approximately implements their extension of the Nash solution to non-convex problems (Conley and Wilkie, 1996).

Note that each of the papers cited in the previous paragraph achieves implementation for a *single* bargaining solution, while we achieve implementation for a family of solutions. While our game forms can implement the Nash and Kalai–Smorodinsky solutions, our game forms can also implement a number of solutions whose implementation has not been examined in the literature (e.g., the Relative Utilitarian solution).

A partial list of contributions providing simple game forms to implement solutions in other contexts is: Hurwicz (1979) and Walker (1981) for the Walrasian and Lindahl solutions, Thomson (1996) for the no-envy solution and several variants of it, and Jackson and Moulin (1992) for solutions to provide a public project. General results on implementation by simple game forms—for various definitions of simplicity—are obtained by Dutta, Sen, and Vohra (1995), Jackson, Palfrey, and Srivastava (1994), Saijo, Tatamitani, and Yamato (1996), and Sjöström (1994).

The remaining sections are organized as follows. In Section 2, we describe the model and define various concepts such as bargaining problems, solutions, and implementation. We describe our game forms in Section 3, and we prove the result in Section 4. In Appendix A.1, we prove that all of our solutions violate Maskin's necessary condition for Nash implementation. In Appendix A.2, we prove that any solution satisfies Abreu and Sen's necessary condition for subgame-perfect implementation. Finally, in Appendix A.3, we prove that Moore and Repullo's sufficient condition for subgame-perfect implementation is satisfied by all of our solutions in a certain bargaining setting, but violated by all of our solutions in other settings.

2 Preliminaries

2.1 The Model

Let X be a compact metric space of alternatives and $d \in X$ be a distinguished alternative. Let ΔX be the set of lotteries defined over X having finite support, and let it be endowed with the weak topology. We denote by $p \cdot x + (1-p) \cdot y$ the lottery that selects x with probability p and y with probability $1-p$. The lottery $p \cdot x + (1-p) \cdot d$ is denoted simply by $p \cdot x$. The degenerate lottery that always selects $x \in X$ is denoted by x. We reserve the term *alternative* for a degenerate lottery.

There are two agents, and we denote the set of agents by $N = \{1, 2\}$. Each agent $i \in N$ has a preference relation (weak order) R_i defined over ΔX. The associated strict preference and indifference relations are denoted by P_i and I_i, respectively. We assume that R_i is continuous and satisfies the independence axiom of von Neumann–Morgenstern. The set of preference relations satisfying these conditions is denoted by $\mathcal{R}$. Then, for each $R_i \in \mathcal{R}$, there exists a continuous function $u_i \colon X \to \mathbb{R}$, called *vN–M utility function*, such that (i) the expectation of u_i represents R_i and (ii) u_i is unique up to positive affine transformations (e.g., Kreps, 1988). Let $u(\cdot, R_i) \colon X \to [0, 1]$ be the unique continuous vN–M utility function whose expectation represents R_i with the normalization:

$$u(d, R_i) = 0$$
$$\max_{x \in X} u(x, R_i) = 1.$$

Abusing notation, we denote the utility level of a lottery $a \in \Delta X$ by

$$u(a, R_i) = \sum_{x \in \mathrm{supp}(a)} a(x) u(x, R_i),$$

where $a(x)$ denotes the probability that lottery a assigns to alternative x, and $\mathrm{supp}(a)$ denotes the support of a. We write $u(a, R) = (u(a, R_1), u(a, R_2))$.

We denote by $b(R_i)$ the set of most preferred lotteries for R_i. Our normalization implies that for any $a \in \Delta X$ and any $b \in b(R_i)$, we have $a \; I_i \; [u(a, R_i) \cdot b]$.

A preference profile is denoted by $R = (R_1, R_2)$. Let $S(R) = \{u(a, R) : a \in \Delta X\}$ be the image under $(u(\cdot, R_1), u(\cdot, R_2))$ of ΔX. Since X is compact and $u(., R_i)$ is continuous in x, $\{u(x, R) : x \in X\}$ is compact. Since $S(R)$ is the convex hull of $\{u(x, R) : x \in X\}$, $S(R)$ is compact and convex.

2.2 Bargaining Problems

A *bargaining problem* is a triple (X, d, R) where $R \in \mathcal{R}^N$ and the following conditions hold.[4]

A1. There exists a lottery that both agents strictly prefer to d.

A2. For each agent, d is a least preferred alternative.

A3. For any lottery $b \in b(R_i)$, $b\, I_j\, d$ for $j \neq i$.

A4. The utility possibility frontier is strictly convex; i.e., for all $\lambda \in (0,1)$ and all lotteries $a, a' \in \Delta X$, if it is not the case that $a\, I_i\, a'$ for all $i \in \{1,2\}$, then there exists a lottery $a'' \in \Delta X$ such that $a''\, P_i\, (\lambda \cdot a + (1-\lambda) \cdot a')$ for all $i \in \{1,2\}$.

A5. No two alternatives are Pareto indifferent; i.e., for all $x, x' \in X$,

$$[x\, I_i\, x' \text{ for all } i \in \{1,2\}] \Longrightarrow [x = x'].$$

A1 excludes trivial problems where Pareto improvement is not feasible. A2 and A3 are standard assumptions and imposed for simplicity. A4 and A5 are imposed to avoid ties.

A4 and A5 imply that all Pareto efficient lotteries are degenerate. A2, A3, and A5 imply that $b(R_i)$ is a singleton containing a degenerate lottery.[5]

Here is an example of a bargaining problem satisfying all of our assumptions.

Example 1. Two agents have to divide $I > 0$ units of a good that is perfectly divisible and freely disposable. Let $X = \{(x_1, x_2) \in [0, I]^2 : x_1 + x_2 \leq I\}$ and $d = (0,0)$. Agent i's preferences are represented by a vN–M utility function $u_i(x)$ such that $u_i(x)$ depends only on x_i (selfishness) and $u_i(\cdot, x_j)$ is a continuous, strictly increasing, and strictly concave function of x_i. Then assumptions A1 – A5 are all satisfied.

We fix X and d, and let $\mathcal{B} \subseteq \mathcal{R}^N$ be the set of preference profiles $R \in \mathcal{R}^N$ such that (X, d, R) is a bargaining problem. Note that $\mathcal{B}$ may not be a product set.

2.3 Bargaining Solutions

A *solution* is a function $f\colon \mathcal{B} \to \Delta X$ associating a lottery with each preference profile in $\mathcal{B}$. Note that with this formulation, a solution is a function of *preferences*, and hence its outcome is invariant with respect to the choice of a vN–M utility representation.

This property is called *scale invariance* in the literature and embedded in our definition of solutions. Many important solutions in the literature are defined in utility space and violate this property (e.g., the Egalitarian, Utilitarian, and Equal Loss solutions). Thus these solutions are not considered in this paper.

Two most important solutions considered in the literature are the Nash and Kalai–Smorodinsky solutions, which are both scale invariant. The *Nash solution* (Nash, 1950) is defined by $N(R) = \arg\max_{x \in X} u(x, R_1) \cdot u(x, R_2)$. The *Kalai–Smorodinsky solution* (Kalai and Smorodinsky, 1975) is defined by $KS(R) = \arg\max_{x \in X} \min\{u(x, R_1), u(x, R_2)\}$. Both solutions are defined by maximizing a function of the form $W(u(x, R_1), u(x, R_2))$. We are interested in implementing *every* solution defined in this fashion.

We define a solution f^W by

$$f^W(R) = \arg\max_{x \in X} W(u(x, R_1), u(x, R_2)), \tag{1}$$

where $W\colon [0,1]^2 \to \mathbb{R}$ is continuous, monotonic (i.e., for all $v, v' \in [0,1]^2$, $v' \gg v$ implies $W(v') > W(v)$), and quasi-concave. The set of functions W satisfying these conditions is denoted by $\mathcal{W}$. The function W may be interpreted as the objective function of the arbitrator.

Note that in (1), we maximize over X and not over ΔX. This is without loss of generality, since as mentioned above, assumptions A4 and A5 imply that any Pareto efficient lottery is degenerate. Furthermore, assumption A4 and A5 together with the quasi-concavity of W imply that the maximizer in (1) is unique and hence f^W is single-valued.

The Nash and Kalai–Smorodinsky solutions are defined in (1) with $W(u) = u_1 \cdot u_2$ and $W = \min\{u_1, u_2\}$, respectively. Another interesting example of f^W is the one with $W(u) = u_1 + u_2$. This solution, introduced by Cao (1982), is called the *Relative* (or *Normalized*) *Utilitarian solution*, which has received increasing attention recently (e.g., Dhillon and Mertens, 1999; Segal, 2000; Sobel, 2001).

2.4 Subgame-Perfect Implementation

We now introduce our definition of implementation. By a sequential game form, we mean a finite-horizon extensive game form with perfect information. A sequential game form Γ is said to *implement* a solution f *in subgame-perfect equilibrium* if for any preference profile $R \in \mathcal{B}$, $f(R)$ is the unique pure-strategy subgame-perfect equilibrium outcome of the game (Γ, R). Note that this condition is what is called *full* implementation, since *all* equilibria are required to induce the desirable outcome.

3 The Game Forms

We now present a sequential game form that implements solution f^W for a given function $W \in \mathcal{W}$. See Figure 1. In stage 1, agent 1 announces a vector $p \in [0,1]^2$ such that $p_1 + p_2 \geq 1$. Having observed p, agent 2 offers a "counter-proposal" $p' = (p'_1, p'_2) \in [0,1]^2$. It is required that $W(p'_1, p'_2) = W(p_1, p_2)$. The agent who moves in the next stage, called agent i, is then determined based on whether agent 2 agrees ($p' = p$) or disagrees ($p' \neq p$). If agent 2 agrees, then he moves next ($i = 2$). Otherwise, agent 1 moves next ($i = 1$). Agent i then chooses either "quit" or "stay," and then announces a lottery a_i. If he chooses to "quit," then the game ends with $p'_i \cdot a_i$ as the outcome. If agent i chooses to "stay," then agent $j \neq i$ either "accepts" a_i, in which case the outcome is a_i, or he selects another lottery a'_j, in which case the outcome is $p'_j \cdot a'_j$.

[Fig 1 about here]

To summarize, game form Γ^W is defined as follows:

Stage 1. Agent 1 announces a vector $p \in [0,1]^2$ that satisfies $p_1 + p_2 \geq 1$.

Stage 2. Agent 2 announces a vector $p' \in [0,1]^2$ that satisfies $W(p'_1, p'_2) = W(p_1, p_2)$.

If $p' = p$, let $(i,j) = (2,1)$. Otherwise let $(i,j) = (1,2)$.

Stage 3. Agent i announces $\lambda_i \in \{$"stay", "quit"$\}$ and a lottery $a_i \in \Delta X$.

If λ_i is "quit," the game ends with $p'_i \cdot a_i$ as the outcome. Otherwise, go to Stage 4.

Stage 4. Agent j chooses either "accept" or a lottery $a'_j \in \Delta X$, and then the game ends. If agent j chooses "accept," the outcome is a_i. If he chooses a lottery $a'_j \in \Delta X$, the outcome is $p'_j \cdot a'_j$.

The first two stages determine a vector p'. The value p'_i is going to be the utility level for agent i when he chooses to "quit" in stage 3, and p'_j is the utility level for agent j when he rejects agent i's proposal in stage 4. To determine p', agent 1 first offers a proposal p, and then agent 2 can either agree on it or offer a counter-proposal p'. Agreeing is appealing for agent 2 because it allows him to move in the next stage and enjoy a "first-mover's advantage."

Note that all game forms Γ^W are identical except for their second stage. The other stages are defined independently of the function W.

Our main result states that game form Γ^W implements f^W in subgame-perfect equilibrium. To get the intuition of the result, ignore the case $p' = p$ for a moment and suppose that agent 1 moves in stage 3. First note that in stage 3, agent 1 chooses “stay” only when there exists a proposal a_1 that gives each agent $k \in \{1,2\}$ a utility larger than or equal to p'_k. Indeed if $u(a_1, R_2) < p'_2$, agent 2 will reject agent 1’s proposal in stage 4 (which is as bad as d for agent 1), and if $u(a_1, R_1) < p'_1$, agent 1 is better off choosing “quit.” Since agent 2 in stage 2 does not want agent 1 to choose “quit,” agent 2 announces a point p' inside the normalized utility possibility set $S(R)$. Typically, agent 2’s best action is to choose the point in $S(R)$ that has the maximal second coordinate (i.e., point p' in Figure 2), and this point gives the utility vector of the equilibrium outcome. What agent 1 does in stage 1 is to choose an indifference curve of W, on which agent 2 has to set p'. Since agent 2 chooses p' in $S(R)$ to maximize p'_2, agent 1’s best action is to choose the indifference curve that is tangent to the frontier of $S(R)$. [Fig 2 about here]

Our game form allows agent 2 to move in stage 3 if agent 2 “agrees” ($p' = p$). Without this device, agent 1 can set $p = (1,1)$ and then choose “quit,” which induces his ideal alternative (since then agent 2 has no option but to announce $p' = p$). By allowing agent 2 to move in stage 3 if he agrees, the game form prevents agent 1 from choosing an indifference curve that lies outside $S(R)$. If agent 1 chooses an indifference curve that lies outside $S(R)$, then any $p' \neq p$ makes agent 1 to choose “quit” and thus agent 2’s best response is to announce $p' = p$ and then choose “quit,” which is as bad as d for agent 1.

Theorem 1. *For each $W \in \mathcal{W}$, game form Γ^W implements solution f^W in subgame-perfect equilibrium.*

A proof of the theorem is given in the next section.

4 Proof of Theorem 1

We fix $R \in \mathcal{B}$. For simplicity, agent i’s normalized utility function is denoted by $u_i(\cdot)$. We first assume the existence of equilibrium and characterize equilibrium outcomes. We will exhibit an equilibrium at the end of the proof. We start with the last stage.

4.1 Stage 4

It is straightforward to see that agent j’s best responses are:

$$\begin{cases} \text{“accept”} & \text{if } u_j(a_i) > p'_j, \\ \{\text{“accept”}, b(R_j)\} & \text{if } u_j(a_i) = p'_j, \\ b(R_j) & \text{if } u_j(a_i) < p'_j. \end{cases}$$

This implies that to avoid $b(R_j)$, agent i must propose a_i such that $u_j(a_i) \geq p'_j$. That is, p'_j is the minimum utility level that agent i must guarantee to agent j to avoid a worst outcome for agent i.

4.2 Stage 3

4.2.1 Case 1: $p' \in \operatorname{int} S(R)$

See Figure 3. Agent i can obtain a utility level of p'_i by choosing "quit." Suppose he chooses "stay" and proposes a_i. If $u_j(a_i) < p'_j$, agent j will select $b(R_j)$, and thus the outcome is $p'_j \cdot b(R_j)$, which is as bad as d for agent i. If $u_j(a_i) > p'_j$, agent j will accept the proposal. Since $p' \in \operatorname{int} S(R)$, there exists a lottery a such that $u(a) \gg p'$. Such a lottery is acceptable for agent j and gives agent i a utility level of more than p'_i. This means that "quit" is not optimal for agent i, and his best response is to "stay" and propose the alternative [Fig 3 about here]

$$\varphi(p') \equiv \arg \max_{\substack{x \in X \text{ s.t.} \\ u_j(x) = p'_j}} u_i(x).$$

Recall that ("stay", $\varphi(p')$) induces either $\varphi(p')$ or $p'_j \cdot b(R_j)$ depending on whether or not agent j accepts $\varphi(p')$. But agent i has a best response if and only if agent j accepts $\varphi(p')$. Thus $\varphi(p')$ is the unique equilibrium outcome of the subgame.

4.2.2 Case 2: $p' \notin S(R)$

Then there is no lottery a such that $u(a) \geqq p'$. This means that "stay" is not optimal for agent i and his response is ("quit", $b(R_i)$). The outcome is $p'_i \cdot b(R_i)$.

4.2.3 Case 3: p' is on the boundary of $S(R)$

Since agent i is indifferent between $\varphi(p')$ and $p'_i \cdot b(R_i)$, the subgame has two equilibria: (i) agent i chooses ("quit", $b(R_i)$) and the outcome is $p'_i \cdot b(R_i)$; (ii) agent i chooses ("stay", $\varphi(p')$) and the outcome is $\varphi(p')$.

4.3 Stage 2

Let

$$P'(p) = \{p' \in [0,1]^2 : W(p'_1, p'_2) = W(p_1, p_2)\}$$

be the indifference curve for W passing through p. This is the set of vectors that agent 2 is allowed to announce in this stage. We distinguish several cases (see

Figure 4). [Fig 4 about here]

4.3.1 Case 1: $P'(p) \cap S(R) = \emptyset$

This means that agent 2 cannot announce a p' inside $S(R)$. Thus for any p', the game ends with $p'_i \cdot b(R_i)$, where $i = 2$ if agent 2 "agrees" ($p' = p$), and $i = 1$ otherwise. Thus agent 2's best response is to "agree," i.e., to announce $p' = p$.[6]

4.3.2 Case 2: $|P'(p) \cap S(R)| > 1$

This means that the boundary of $S(R)$ intersects $P'(p)$ at two points. Choose the intersection whose second coordinate is higher than the other's, and denote it by $\hat{p}$. Let $\hat{x} \in X$ be the unique alternative such that $u(\hat{x}) = \hat{p}$. We distinguish three subcases.

Subcase i: $\hat{p}_2 > p_2$ Whatever action agent 2 chooses in this stage, the outcome is not better than $\hat{x}$ for agent 2 (since any p' such that $p'_2 > \hat{p}_2$ is outside of $S(R)$, and induces agent 1's "quit"). Furthermore, by choosing p' such that $p'_2 = \hat{p}_2 - \varepsilon$ for small $\varepsilon > 0$, agent 2 can make the outcome arbitrarily close to $\hat{x}$ in terms of welfare. Thus the unique equilibrium outcome of this subgame is $\hat{x}$.

Subcase ii: $p_2 = \hat{p}_2$ In this case too, the outcome is never better than $\hat{x}$ for agent 2. And if agent 2 chooses $p' = p = \hat{p}$, the outcome is either $\hat{x}$ or $p_2 \cdot b(R_2)$, which are equivalent for agent 2 in terms of welfare. Thus the equilibrium outcome is either $\hat{x}$ or $p_2 \cdot b(R_2)$.

Subcase iii: $p_2 > \hat{p}_2$ This means that p is outside of $S(R)$. Thus agent 2's best response is $p' = p$, and the equilibrium outcome is $p_2 \cdot b(R_2)$.

4.3.3 Case 3: $P'(p) \cap S(R) = \{u(f^W(R))\}$

Let $\hat{p} = u(f^W(R))$.

Subcase i: $p = \hat{p}$ If agent 2 announces $p' \neq p$, then since p' is outside of $S(R)$, the outcome is $p'_1 \cdot b(R_1)$, which is as bad as d for agent 2. Thus his best response is $p' = p$, which induces either $f^W(R)$ or $p_2 \cdot b(R_2)$.

Subcase ii: $p \neq \hat{p}$ Then $p' = \hat{p}$ induces either $f^W(R)$ or $\hat{p}_1 \cdot b(R_1)$. Note that p is outside of $S(R)$, and hence $p' = p$ induces $p_2 \cdot b(R_2)$. Any other vector p' induces $p'_1 \cdot b(R_1)$. Thus the equilibrium outcome is either $f^W(R)$ or $p_2 \cdot b(R_2)$, depending on whether the outcome associated with $p' = \hat{p}$ is better than $p_2 \cdot b(R_2)$ for agent 2.

4.4 Stage 1

We have distinguished several cases in the analysis of stage 2. Note that the only cases in which the equilibrium outcome is better than d for agent 1 are Case 2-i, Case 2-ii, and Case 3. In each of these cases, the outcome better than d for agent 1 is

$$\hat{x}(p) \equiv \arg \max_{\substack{x \in X \text{ s.t.} \\ u(x) \in P'(p)}} u_2(x),$$

which is well-defined in Cases 2 and 3. For agent 1, the best outcome among $\{\hat{x}(p)\}_p$ is $f^W(R)$. Furthermore, he can make the outcome arbitrarily close (in terms of welfare) to $f^W(R)$ by choosing $p = u(f^W(R)) - (\varepsilon, \varepsilon)$ for small $\varepsilon > 0$. Thus the unique equilibrium outcome of the game is $f^W(R)$, if an equilibrium exists.

4.5 Equilibrium Existence

Finally, we exhibit an equilibrium strategy profile. In this profile, all ties are broken in a way that avoids outcomes of the form $p \cdot b(R_i)$.

Stage 1: Agent 1 chooses $p = u(f^W(R))$.

Stage 2: Given p, agent 2 chooses

$$p' = \begin{cases} p & \text{if } P'(p) \cap S(R) = \emptyset, \\ p & \text{if } P'(p) \cap S(R) \neq \emptyset \text{ and } p_2 \geq u_2(\hat{x}(p)), \\ u(\hat{x}(p)) & \text{if } P'(p) \cap S(R) \neq \emptyset \text{ and } p_2 < u_2(\hat{x}(p)). \end{cases}$$

Stage 3: Given p and p', agent i chooses

$$\begin{cases} (\text{“stay”}, \varphi(p')) & \text{if } p' \in S(R), \\ (\text{“quit”}, b(R_i)) & \text{otherwise.} \end{cases}$$

Stage 4: Given p and p', and ("stay", a_i), agent $j \neq i$ chooses

$$\begin{cases} \text{“accept”} & \text{if } u_j(a_i) \geq p'_j, \\ b(R_j) & \text{otherwise.} \end{cases}$$

This completes the proof of Theorem 1.

A Appendix

A.1 All f^W violate Maskin's necessary condition for Nash implementation

Maskin (1999) provides a necessary condition for implementation in Nash equilibrium. We prove that his condition is violated by solutions f^W for all $W \in \mathcal{W}$ in the bargaining setting introduced in Example 1.

A solution (correspondence) $f\colon \mathcal{B} \twoheadrightarrow \Delta X$ is *Maskin monotonic* if for all $R \in \mathcal{B}$, all $a \in f(R)$, all $i \in N$, and all $R'_i \in \mathcal{R}$ with $(R'_i, R_j) \in \mathcal{B}$,

$$[\text{for all } a' \in \Delta X, a \; R_i \; a' \text{ implies } a \; R'_i \; a'] \Longrightarrow a \in f(R'_i, R_j).$$

In words, if a lottery a is selected for profile R, then it should also be selected for profile (R'_i, R_j) if the lower contour set of R'_i at a contains that of R_i at a. Maskin shows that this condition is necessary for Nash implementability.

Proposition 1. *In the bargaining setting of Example 1, none of the solutions f^W is Maskin monotonic.*

Proof. Without loss of generality, assume that the total amount to be divided is $I = 1$. Consider a preference profile R in which agent i has a differentiable, strictly increasing, and strictly concave vN–M utility function $g_i\colon [0,1] \to \mathbb{R}$, such that $g_i(0) = 0$ and $g_i(1) = 1$. Let $x = f^W(R)$.

Suppose now that agent i's vN–M utility function changes to a function $h_i\colon [0,1] \to \mathbb{R}$ that is differentiable, strictly increasing, and strictly concave, and satisfies

$$\begin{aligned} h_i(z) &= g_i(z) \quad \text{if } z \geq x_i, \\ h_i(z) &< g_i(z) \quad \text{if } z < x_i. \end{aligned}$$

Note that h_i is not normalized. Let R'_i be the associated preference relation. Then the upper contour set of R'_i at allocation x is contained in the corresponding set for

R_i. Indeed, for any lottery $a = \sum_k p^k \cdot y^k$, if $x \, R_i \, a$, then

$$\sum_k p^k h_i(y_i^k) \leq \sum_k p^k g_i(y_i^k) \leq g_i(x_i) = h_i(x_i),$$

which implies $x \, R_i' \, a$. We show that either $f^W(R_1', R_2) \neq x$ or $f^W(R_1, R_2') \neq x$, which is a desired violation of Maskin monotonicity.

We first normalize h_i to obtain

$$\hat{h}_i(z) = \frac{h_i(z) - h_i(0)}{1 - h_i(0)}.$$

If $f^W(R_1', R_2) = x$, then at point $(\hat{h}_1(x_1), g_2(x_2))$, the indifference curve of W is tangent to the frontier of $S(R_1', R_2)$. Thus, the indifference curve at this point has a slope equal to $g_2'(x_2)(1 - h_1(0))/g_1'(x_1)$, which is larger than $g_2'(x_2)/g_1'(x_1)$. This means that the indifference curve is steeper at $(\hat{h}_1(x_1), g_2(x_2))$ than at $(g_1(x_1), g_2(x_2))$. By a symmetric argument, if $f^W(R_1, R_2') = x$, then the indifference curve is flatter at $(g_1(x_1), \hat{h}_2(x_2))$ than at $(g_1(x_1), g_2(x_2))$. These slope conditions cannot be satisfied simultaneously because W is quasi-concave and $\hat{h}_i(x_i) > g_i(x_i)$. □

A.2 All solutions satisfy Abreu and Sen's necessary condition for subgame-perfect implementation

Abreu and Sen (1990) provide a necessary condition for subgame-perfect implementation. Their condition is satisfied by f^W for all W because our game forms do implement all f^W. More interestingly, it can be shown that Abreu and Sen's condition is satisfied by *any* solution for our bargaining problems. That is, their condition fails to identify a single solution that is not implementable.

A solution (correspondence) $f \colon \mathcal{B} \twoheadrightarrow \Delta X$ satisfies *Condition* α if there exists a set $B \subseteq \Delta X$ with range$(f) \subseteq B$ such that for all $R, R' \in \mathcal{B}$ and all $a \in f(R) \setminus f(R')$, there exist a sequence of agents $\{j(0), \ldots, j(l)\}$ and a sequence of lotteries $\{a = a_0, a_1, \ldots, a_{l+1}\}$ in B such that

(i) $a_0 \, R_{j(0)} \, a_1 \, R_{j(1)} \, a_2 \ldots a_l \, R_{j(l)} \, a_{l+1} \, P'_{j(l)} \, a_l$,

(ii) for all $k \in \{0, \ldots, l-1\}$, a_k is not maximal in B for $R'_{j(k)}$,

(iii) if a_{l+1} is maximal in B for R_i' for some $i \neq j(l)$, then either $l = 0$ or $j(l-1) \neq j(l)$.

Proposition 2. *Any solution satisfies Condition* α*.*

Proof. We show that any solution f satisfies Condition α with $B = \Delta X$. Let $R, R' \in \mathcal{B}$ be such that $f(R) \equiv \alpha \neq f(R')$. Since $R \neq R'$, there exist an agent i and lotteries β and γ such that either $\beta\, R_i\, \gamma\, P'_i\, \beta$ or $\beta\, R'_i\, \gamma\, P_i\, \beta$.

Case 1: $\beta\, R_i\, \gamma\, P'_i\, \beta$.

Since $b(R'_1) \neq b(R'_2)$, there exists an agent j such that $\alpha \neq b(R'_j)$. Let $k \neq j$ be the other agent, and let $p \in (0,1)$. Then since $p \cdot b(R_k)$ is as bad as d for agent j, we have

$$\alpha\, R_j\, (p \cdot b(R_k))\, R_k\, (p \cdot \beta)\, R_i\, (p \cdot \gamma)\, P'_i\, (p \cdot \beta),$$

and hence condition (i) holds. Condition (ii) holds as well, because $\alpha \neq b(R'_j)$ and $p < 1$. Moreover, (iii) holds because $p < 1$.

Case 2: $\beta\, R'_i\, \gamma\, P_i\, \beta$.

Subcase 1: $\gamma\, P'_i\, d$. Since $\gamma\, P_i\, \beta$, there exists $p \in (0,1)$ such that $(p \cdot \gamma)\, P_i\, \beta$. Note also that $\beta\, R'_i\, \gamma\, P'_i\, d$ implies $\beta\, P'_i\, (p \cdot \gamma)$. Thus we have

$$(p \cdot \gamma)\, P_i\, \beta\, P'_i\, (p \cdot \gamma),$$

and hence Case 1 applies.

Subcase 2: $\gamma\, I'_i\, d$. This means $b(R'_i)\, P'_i\, \gamma$. Since $\gamma\, P_i\, \beta$, there exists $p \in (0,1)$ such that $\gamma\, P_i\, [p \cdot b(R'_i) + (1-p) \cdot \beta]$. This together with $b(R'_i)\, P'_i\, \gamma$ and $\beta\, R'_i\, \gamma$ implies

$$\gamma\, P_i\, [p \cdot b(R'_i) + (1-p) \cdot \beta]\, P'_i\, \gamma,$$

and hence Case 1 applies. □

A.3 On Moore and Repullo's sufficient condition for subgame-perfect implementation

Moore and Repullo (1988) provide a sufficient condition for subgame-perfect implementation. We show that their condition is satisfied by solutions f^W for essentially all W in the bargaining setting of Example 1, while in other settings the condition can be violated by f^W for all W.

To define Moore and Repullo's condition, let $f\colon B \twoheadrightarrow \Delta X$ be a solution (correspondence) satisfying (i) of Condition α for $B = \Delta X$. Given $R, R' \in \mathcal{B}$ and $a \in f(R) \setminus f(R')$, let $\Sigma(R, R', a)$ be the set of sequences $\{a, a_1, \ldots, a_{l+1}\}$ satisfying condition (i). For a given selection σ from Σ, let

$$Q^\sigma = \bigcup \{a' \in \sigma(R, R', a) : R, R' \in \mathcal{B}, a \in f(R) \setminus f(R')\}.$$

For each set $B \subseteq \Delta X$, let $M(R_i, B)$ be the set of most preferred lotteries in B for R_i.

The following is Moore and Repullo's sufficient condition for subgame-perfect implementation for the two-person case. The solution f satisfies *Condition* C^{++} if there exist a selection σ from Σ, a set $B \subseteq \Delta X$ such that $B \supseteq Q^\sigma$, and a lottery z, such that the following conditions hold for all $R \in \mathcal{B}$:

1. $M(R_i, B)$ is non-empty for each i.
2. $M(R_1, B)$, $M(R_2, B)$, and Q^σ are pairwise disjoint.
3. Both agents strictly prefer any lottery in Q^σ to z.
4. Both agents weakly prefer any lottery in $M(R_1, B) \cup M(R_2, B)$ to z.

Proposition 3. *Consider the bargaining setting of Example 1. For any $W \in \mathcal{W}$, if $f^W(R) \notin \{(0, I), (I, 0)\}$ for all $R \in \mathcal{B}$, then f^W satisfies Condition C^{++}.*

Proof. It suffices to find a selection σ such that all lotteries in Q^σ give each agent a positive amount with a positive probability, since then we can take $B = \Delta X$ and $z = d$.

So, let $R, R' \in \mathcal{B}$ be such that $f^W(R') \neq f^W(R)$ and denote $x = f^W(R)$. Let $u_i, u_i' \colon [0, I] \to [0, 1]$ be agent i's normalized vN–M utility functions associated with R_i and R_i', respectively.

Since $f^W(R) \neq f^W(R')$, we have $R_i \neq R_i'$ for at least one of the agents and we assume that it holds for agent 1. Then, there exists $y \in [0, I]$ such that $u_1(y) \neq u_1'(y)$. Since the utility functions are continuous, we can assume that $y \notin \{0, I, x_1\}$. Take a small number $\varepsilon \in (0, 1)$ such that

$$\varepsilon < u_2(x_2), \tag{2}$$

$$1 - \varepsilon > \max\{u_1(y), u_1'(y)\}. \tag{3}$$

Case 1: $u_1(y) > u_1'(y)$. Take a number p such that

$$u_1'(y) < p < u_1(y). \tag{4}$$

Define three lotteries as follows:

$$\ell_1 = (1 - \varepsilon) \cdot (I, 0) + \varepsilon \cdot (0, I),$$
$$\ell_2 = (y, I - y),$$
$$\ell_3 = p \cdot (I, 0) + (1 - p) \cdot (0, I).$$

Then (i) in Condition α holds with $(a_0, a_1, a_2, a_3) = (x, \ell_1, \ell_2, \ell_3)$ and $(j(0), j(1), j(2)) = (2, 1, 1)$, because (2) implies $x \, P_2 \, \ell_1$, and (3) and (4) imply $\ell_1 \, P_1 \, \ell_2 \, P_1 \, \ell_3 \, P'_1 \, \ell_2$.

Case 2: $u_1(y) < u'_1(y)$. Take a number p such that

$$u_1(y) < p < u'_1(y) \tag{5}$$

and redefine lottery ℓ_3 with this p. Then (i) in Condition α holds with $(a_0, a_1, a_2, a_3) = (x, \ell_1, \ell_3, \ell_2)$ and $(j(0), j(1), j(2)) = (2, 1, 1)$, because (2) implies $x \, P_2 \, \ell_1$ as before, and (3) and (5) imply $\ell_1 \, P_1 \, \ell_3 \, P_1 \, \ell_2 \, P'_1 \, \ell_3$. □

In other bargaining settings, Condition C^{++} may be violated by solutions f^W for all W. For example, consider a setting where any alternative $x \in X \setminus \{d\}$ is some agent's most preferred alternative for *some* preference profile. In such a setting, Condition C^{++} is violated by f^W for all W. To see this, let σ and B be such that $Q^\sigma \subseteq B$. Since $f^W(R) \in X \setminus \{d\}$, there exists $x \in (X \setminus \{d\}) \cap Q^\sigma$. Let $R \in \mathcal{B}$ and $i \in N$ be such that x is a most preferred alternative for R_i. Then $x \in M(R_i, B) \cap Q^\sigma$, violating condition 2 in Condition C^{++}.

References

Abreu, D., and Sen, A. (1990). "Subgame perfect implementation: A necessary and almost sufficient condition," *J. Econ. Theory* 50, 285–299.

Binmore, K. G., Rubinstein, A., and Wolinsky, A. (1986). "The Nash bargaining solution in economic modeling," *Rand J. Econ.* 17, 176–188.

Cao, X. (1982). "Preference functions and bargaining solutions," *Proceedings of the 21st IEEE Conference on Decision and Control,* p. 3 vol. xxxi+1386, pages 164-171.

Conley, J. P., and Wilkie, S. (1995). "Implementing the Nash extension bargaining solution for non-convex problems," *Econ. Design* 1, 205–216.

——— (1996). "An extension of the Nash bargaining solution to nonconvex problems," *Games Econ. Behav.* 13, 26–38.

Dhillon, A., and Mertens, J.-F. (1999). "Relative Utilitarianism," *Econometrica* 67, 471–498.

Dutta, B., Sen, A., and Vohra, R. (1995). "Nash implementation through elementary mechanisms in economic environments," *Econ. Design* 1, 173–203.

Howard, J. V. (1992). "A social choice rule and its implementation in perfect equilibrium," *J. Econ. Theory* 56, 142–159.

Hurwicz, L. (1979). "Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points," *Rev. Econ. Stud.* 46, 217–225.

Jackson, M. (1992). "Implementation in undominated strategies: A look at bounded mechanisms," *Rev. Econ. Stud.* 59, 757–775.

Jackson, M., and Moulin, H. (1992). "Implementing a public project and distributing its cost," *J. Econ. Theory* 57, 125–140.

Jackson, M., Palfrey, T. R., and Srivastava, S. (1994). "Undominated Nash implementation in bounded mechanisms," *Games Econ. Behav.* 6, 474–501.

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

Kreps, D. M. (1988). *Notes on the Theory of Choice.* Boulder, Colo.: Westview Press.

Maskin, E. (1999). "Nash equilibrium and welfare optimality," *Rev. Econ. Stud.* 66, 23–38.

Moore, J. (1991). "Implementation, contracts and renegotiation in environments with complete information," in *Advances in Economic Theory: Proceedings of the Sixth World Congress of the Econometric Society* (J.-J. Laffont, Ed.). Cambridge: Cambridge University Press.

Moore, J., and Repullo, R. (1988). "Subgame perfect implementation," *Econometrica* 56, 1191–1220.

Moulin, H. (1984). "Implementing the Kalai-Smorodinsky bargaining solution," *J. Econ. Theory* 33, 32–45.

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

Osborne, M. J., and Rubinstein, A. (1994). *A Course in Game Theory.* Cambridge, MA: MIT Press.

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

Rubinstein, A., Safra, Z., and Thomson, W. (1992). "On the interpretation of the Nash bargaining solution and its extension to non-expected utility preferences," *Econometrica* 60, 1171–1186.

Saijo, T., Tatamitani, Y., and Yamato, T. (1996). "Toward natural implementation," *Int. Econ. Rev.* 37, 949–980.

Segal, U. (2000). "Let's agree that all dictatorships are equally bad," *J. Polit. Econ.* 108, 569–89.

Sjöström, T. (1994). "Implementation in undominated Nash equilibria without integer games," *Games Econ. Behav.* 6, 502–511.

Sobel, J. (2001). "Manipulation of preferences and relative utilitarianism," *Games Econ. Behav.* 37, 196–215.

Thomson, W. (1996). "Divide and permute," Mimeo, University of Rochester.

Walker, M. (1981). "A simple incentive compatible scheme for attaining Lindahl allocations," *Econometrica* 49, 65–71.

Notes

[1]For an excellent introduction to the literature, see Moore (1991). Moore and Repullo (1988) and Abreu and Sen (1990) obtained general results on subgame-perfect implementation.

[2]Abreu and Sen also provide a sufficient condition for the case of three or more agents. The condition does not apply to our problems because we consider two-person problems.

[3]We should admit that in our game forms, an agent may not have a best response when ties are broken in wrong ways in subsequent stages, and this is certainly not desirable. On the other hand, an equilibrium does exist in each subgame.

[4]We follow Rubinstein, Safra, and Thomson (1992). See also Osborne and Rubinstein (1994).

[5]A3 implies that all lotteries in $b(R_i)$ are Pareto indifferent, and by A2, all of them are lotteries over alternatives in $b(R_i)$. By A5, $b(R_i)$ contains only one alternative.

[6]Note that $p_2 > 0$ because $S(R)$ is convex and W is quasi-concave.

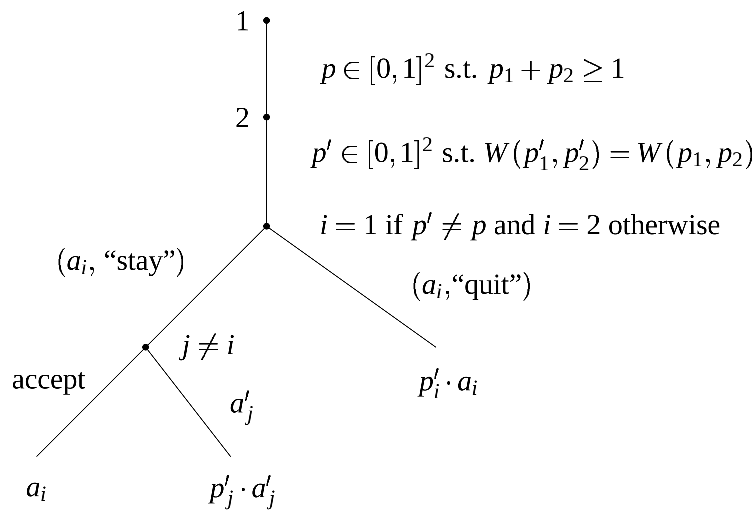
Figure Legends

Figure 1: Game Form Γ^W.

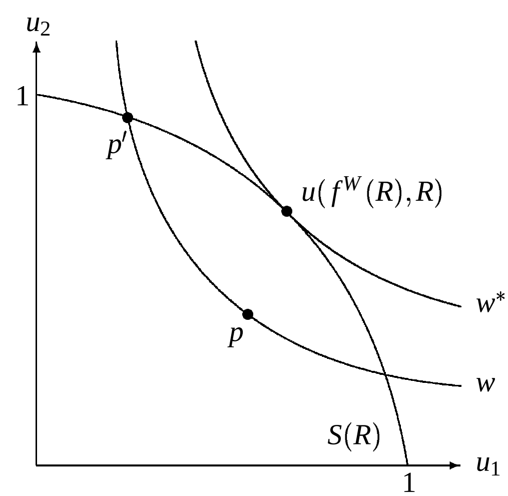
Figure 2: If agent 1 chooses p, the outcome is p' in utility space.

Figure 3: Stage 3: In each case, the arrows point at the normalized utility vectors of equilibrium outcomes.

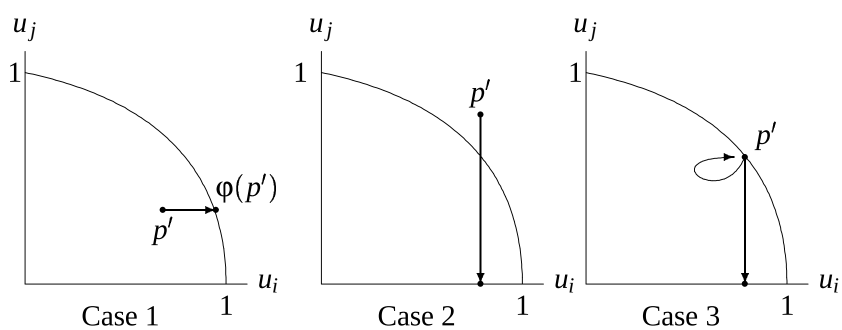
Figure 4: Stage 2: In each case, the arrows point at the normalized utility vectors of equilibrium outcomes.

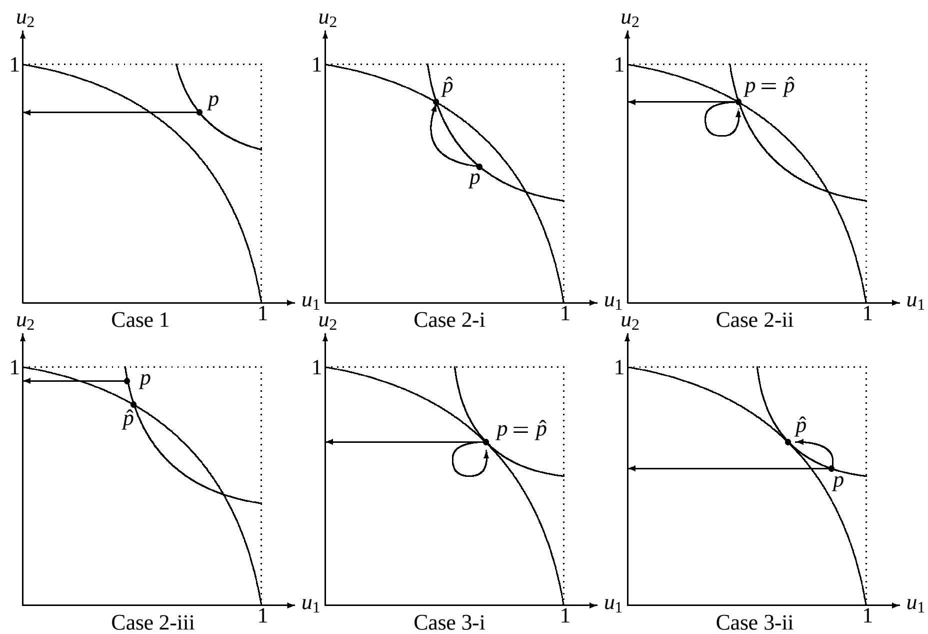


1
$p \in [0,1]^2$ s.t. $p_1 + p_2 \geq 1$
2
$p' \in [0,1]^2$ s.t. $W(p'_1, p'_2) = W(p_1, p_2)$
$i = 1$ if $p' \neq p$ and $i = 2$ otherwise
$(a_i,$ "stay"$)$
$(a_i,$"quit"$)$
$j \neq i$
accept
a'_j
$p'_i \cdot a_i$
a_i
$p'_j \cdot a'_j$



u_2
1
p'
$u(f^W(R), R)$
w^*
p
w
$S(R)$
u_1
1


u_j
1
$\varphi(p')$
p'
u_i
1
Case 1
u_j
1
p'
u_i
1
Case 2
u_j
1
p'
u_i
1
Case 3


u_2
1
p
u_1
1
Case 1
$\hat{p}$
p
Case 2-i
$p = \hat{p}$
Case 2-ii
Case 2-iii
Case 3-i
Case 3-ii