Achieving Efficiency with Manipulative Bargainers

Juan Camilo Gómez*

January 2006

Abstract

Two agents bargain over the allocation of a bundle of divisible commodities. After strategically reporting utility functions to a neutral arbitrator, the outcome is decided by using a bargaining solution concept chosen from a family that includes the Nash and the Raiffa-Kalai-Smorodinsky solutions. When reports are restricted to be continuous, strictly increasing and concave, it has been shown that this kind of "distortion game" leads to inefficient outcomes. We study the distortion game originated when agents are also allowed to claim non-concave utility functions. Contrasting with the previous literature, any interior equilibrium outcome is efficient and any efficient allocation can be supported as an equilibrium outcome of the distortion game. In a similar fashion to the Nash demand game we consider some uncertainty about the opponent's features to virtually implement the Nash bargaining solution.

Key words: bargaining; distortion games; efficiency

JEL Classification Numbers: C72, C78, D74.

*I would like to thank Leo Hurwicz, Yakar Kannai, Marcel Richter, Luis Sánchez-Mier, an associate editor, and the participants in the 2004 Conference on Economic Design (Palma de Mallorca, Spain) for their valuable comments. All remaining errors are mine.

1 Introduction

Two persons settling a dispute often have incentives to make untruthful claims about their preferences. Nevertheless, the application of standard bargaining procedures requires a mediator that is informed about the agents' preferences. It is important to understand if manipulation of private information can affect the desirable properties that the original solution concept has (Thomson [21]). The bargaining literature has dedicated considerable attention to analyze if strategic reports of preferences have incidence on the efficiency of bargaining outcomes. A usual approach to this issue involves defining a "distortion game" that determines the allocation of goods in a pure exchange economy. In such games, a referee decides the final outcome by applying a bargaining concept (such as Nash bargaining or the RKS solution) based on agents' reported utility functions.

Up to now, for such a procedure to generate Pareto-optimal allocations, the reports of agents have been restricted to very particular families of utility functions (e.g. Peters [15], Sobel [18] and [19], and Kibris [9]). If the utility functions are continuous, strictly increasing and concave, there exist "non-pathological" examples of Nash equilibrium reports that imply inefficient outcomes (see Example 3). Instead of restricting, we expand the set of possible reports by including non-concave utility functions. Paradoxically, if the outcome is interior, efficiency is restored. In fact, if agents are slightly uncertain about the tastes of their bargaining partners, we can say even more. Although the arbitrator ignores the true preferences of the bargainers, we design a "perturbed" distortion game that virtually implements the Nash bargaining outcome of the negotiation.

The simplest distortion games[1] that have been studied are those in which a single commodity is distributed. Although efficiency is not an issue for this type of games, it is important to first understand how do agents best-respond in this setting. The fact that bargainers can only claim to have concave (risk-averse) utility functions leads to the existence of a dominant strategy equilibrium. As seen in Crawford and Varian [5],[2] it is optimal for agents to report linear utility functions. This leads to equal division of the commodity.

The multiple commodity case is more complex. In this more general scenario equilibrium outcomes are far from being characterized and some of them are not necessarily Pareto-optimal. In order to achieve efficiency, Sobel [18] requires linear reports from the agents. Thomson [22] reaches a similar result by using quasilinear utility functions. Finally, Peters [15] only allows reports of preference profiles that constitute an equilibrium if agents are telling the truth. Given this precedents, it is natural to suspect that extending the strategy spaces (sets of possible reports) makes the problem intractable. As mentioned before, quite the opposite occurs. After expanding the set of possible reports an interior allocation is an equilibrium outcome if and only if it is Pareto-optimal.

To select among the multiple equilibrium outcomes generated, we adapt our

[1] The distortion game technique was first used by Kurz [10] in reference to a taxation game by Aumann and Kurz [1].

[2] Kannai [8] gives a more general result.

distortion game so that it resembles the Nash demand game [13]. Following Nash's ideas, we assume that players are slightly uncertain about their opponents' preferences and smooth their payoff functions. The procedure we use is closest to the one described by Osborne and Rubinstein [14].[3] Eventhough Nash's demand game and other smoothing models are not concerned with manipulative agents, we are able to preserve their approximate implementation result.

2 The Model

A *bargaining problem* is defined as an ordered pair (S, d) where S is a nonempty, compact, convex subset of $\mathbb{R}^2$, $d = (d_1, d_2) \in S$, and there is at least one $s \in S$ such that $s >> d$.[4] Let $\mathcal{B}$ denote the family of all bargaining problems. A *bargaining solution* is a function $\phi : \mathcal{B} \to \mathbb{R}^2$ such that $\phi(S, d) \in S$. Given $(S, d) \in \mathcal{B}$ define $I_i = \max\{s_i \mid s \in S\}$ for $i = 1, 2$. The bargaining problem $(\{(\frac{s_1-d_1}{I_1-d_1}, \frac{s_2-d_2}{I_2-d_2}) \mid s \in S\}, (0, 0))$ is called the $0-1$ *normalization* of (S, d). Let the *diagonal* of the bargaining problem (S, d) be the set $Diag(S, d) = \{s \in S \mid \frac{s_1-d_1}{I_1-d_1} = \frac{s_2-d_2}{I_2-d_2}\}$. The bargaining solution ϕ is *Pareto efficient* if for any $(S, d) \in \mathcal{B}$, and any $s, t \in S$, $s > t$ implies $t \neq \phi(S, d)$. ϕ is *symmetric* if for any symmetric bargaining problem $(S, d) \in \mathcal{B}$ (i.e. (S, d) satisfies $d_1 = d_2$ and $\forall s \in S$, $(s_2, s_1) \in S$), $\phi(S, d) \in Diag(S, d)$. ϕ is *invariant to positive affine transformations* if for any $(S, d) \in \mathcal{B}$, $a_1, a_2 \in \mathbb{R}_{++}$, and $b_1, b_2 \in \mathbb{R}$, $\phi\big(\{(a_1 s_1 + b_1, a_2 s_2 + b_2) \mid s \in S\}, (a_1 d_1 + b_1, a_2 d_2 + b_2)\big) = (a_1\phi_1(S, d) + b_1, a_2\phi_2(S, d) + b_2)$. ϕ satisfies *midpoint dominance* (also known as *symmetric monotonicity*) if any $0-1$ normalized bargaining problem $(S, d) \in \mathcal{B}$ satisfies $\phi(S, d) \geq (\frac{1}{2}, \frac{1}{2})$. Let Φ be the family of bargaining solutions that satisfy the previous four properties.

We now recall two famous examples of bargaining solutions. The *Nash bargaining solution* [12] is $NB(S, d) = argmax_{s \in S,\, s \geq d}(s_1 - d_1)(s_2 - d_2)$. The *Raiffa - Kalai - Smorodinsky bargaining solution* [7] $RKS(S, d)$ is defined as the unique element of $Diag(S, d)$ whose components are strictly greater than those of any other $t \in Diag(S, d)$. Sobel [18] shows that both $NB(S, d)$ and $RKS(S, d)$ are elements of Φ.

Consider an environment where two agents use an arbitrator to decide the allocation of a bundle of goods $\omega \in \mathbb{R}^n_{++}$. Define $X = \{x \in \mathbb{R}^n \mid \vec{0} \leq x \leq \omega\}$.[5] Each agent i has (true) preferences described by the utility function $u_i : X \longrightarrow \mathbb{R}$, which is assumed to be continuous, strictly increasing, concave, smooth,[6] and $0-1$ normalized so that $u_i(\vec{0}) = 0$ and $u_i(\omega) = 1$. The set of all such functions is denoted by $\mathbf{U}$. Both agents know u_1 and u_2 but the arbitrator does not.

Let $\tilde{\mathbf{U}}$ be defined as the set of reports $\tilde{u} : X \longrightarrow \mathbb{R}$ that are continuous, strictly increasing, quasiconcave, smooth, and $0-1$ normalized. Given $\tilde{u}_1, \tilde{u}_2 \in \tilde{\mathbf{U}}$, define the *utility possibility set* as $UPS(\tilde{u}_1, \tilde{u}_2) = \{v \in [0, 1]^2 \mid \exists(x_1, x_2) \in$

[3]Binmore [3] and van Damme [23] describe other ways of smoothing the payoffs.
[4]Given $a, b \in \mathbb{R}^m$, $a >> b$ means $a_i > b_i$ for all i and $a \geq b$ means $a_i \geq b_i$ for all i.
[5]$\vec{0}$ denotes the vector $(0, \ldots, 0) \in \mathbb{R}^n$.
[6]By smooth we mean that each point on a level surface has a unique supporting hyperplane.

X^2 such that $x_1 + x_2 \leq \omega$ and $v \leq (\tilde{u}_1(x_1), \tilde{u}_2(x_2))\}$. Our first example shows that, as we are not assuming concavity on the reports, the utility possibility set is not necessarily convex.

Example 1. *Let $n = 1$, $\omega = 1$, and $\tilde{u}_1(y) = \tilde{u}_2(y) = y^2$. Then $(0, 1)$ and $(1, 0)$ are in $UPS(\tilde{u}_1, \tilde{u}_2)$, but $(\frac{1}{2}, \frac{1}{2})$ is not. Clearly the UPS is not convex.*

The *outcome correspondence* $\psi : \tilde{\mathbf{U}}^2 \longrightarrow X^2$ is defined as follows. If the utility possibility set $UPS(\tilde{u}_1, \tilde{u}_2)$ is convex, then $\psi(\tilde{u}_1, \tilde{u}_2) = \{(x_1, x_2) \in X^2 \mid (\tilde{u}_1(x_1), \tilde{u}_2(x_2)) = \phi(UPS(\tilde{u}_1, \tilde{u}_2), (0, 0))\}$. Otherwise, $\psi(\tilde{u}_1, \tilde{u}_2) = \{(\vec{0}, \vec{0})\}$. Given $u_1, u_2 \in \mathbf{U}$ and $\phi \in \Phi$, the *distortion game* $G(u_1, u_2, \phi)$ is played by agents reporting utility functions $\tilde{u}_1, \tilde{u}_2$ from $\tilde{\mathbf{U}}$. An arbitrator selects the final allocation of the goods according to ϕ and agents evaluate their payoffs using u_1 and u_2 respectively. The family of all distortion games is denoted by DG.

Given $G(u_1, u_2, \phi) \in DG$, for each $i \in \{1, 2\}$ define the *best response correspondence* $BR_i : \tilde{\mathbf{U}} \longrightarrow \tilde{\mathbf{U}}$ letting $BR_i(\tilde{u}_{-i}) = \{\hat{u}_i \in \tilde{\mathbf{U}} \mid \forall(\hat{x}_i, \hat{x}_{-i}) \in \psi(\hat{u}_i, \tilde{u}_{-i}), \forall\tilde{u}_i \in \tilde{\mathbf{U}}, \forall(\tilde{x}_i, \tilde{x}_{-i}) \in \psi(\tilde{u}_i, \tilde{u}_{-i}), u_i(\hat{x}_i) \geq u_i(\tilde{x}_i)\}$ for any $\tilde{u}_{-i} \in \tilde{\mathbf{U}}$.[7] The pair $(u_1^*, u_2^*) \in \tilde{\mathbf{U}}^2$ is a *Nash equilibrium* of $G(u_1, u_2, \phi)$ if $u_i^* \in BR_i(u_{-i}^*)$ for $i = 1, 2$.

In some cases, for example when both agents report identical and linear utility functions, the outcome correspondence of the distortion game is multivalued. The issue of what outcome to choose is usually dealt with by establishing a "tie-breaking mechanism". With our definition of best response there is no need for such a procedure. In fact, given a report $\tilde{u}_{-i} \in \tilde{\mathbf{U}}$, if $\hat{u}_i \in BR_i(\tilde{u}_{-i})$, any outcome $\hat{x} \in \psi(\hat{u}_i, \tilde{u}_{-i})$ generates the same utility $u_i(\hat{x}_i)$. Thus, in a Nash equilibrium agents are indifferent among all possible outcomes. An important advantage of our enriched strategy space is that it allows existence of Nash equilibria even with our quite stringent version of best response.[8]

It is important to emphasize that our efficiency theorems do not rely on the ability that bargainers may have of choosing a Pareto-optimal allocation from the ones proposed by the outcome correspondence. In previous results that obtain efficient outcomes from linear reports [18], the tie-breaking mechanism is essential. Here, efficiency is directly derived from the strategic interaction between the bargainers.

3 Optimal Strategies

We take advantage of the players' additional choices by analyzing the shape of the UPS generated by their reports. The Pareto frontier of $UPS(\tilde{u}_1, \tilde{u}_2)$ is described, from agent i's perspective, by the function $h(\tilde{u}_i, \tilde{u}_{-i}) : [0, 1] \rightarrow [0, 1]$ defined by $h(\tilde{u}_i, \tilde{u}_{-i})(v_i) = \max\{\tilde{u}_{-i}(\omega - x_i) \mid x_i \in X \text{ and } \tilde{u}_i(x_i) = v_i\}$. The function $h(\tilde{u}_i, \tilde{u}_{-i})$ is well defined because $\tilde{u}_{-i}$ is maximized over a compact set. The following lemma describes the standard properties of the Pareto frontier.

[7] Given $i \in \{1, 2\}$, we use the convention that $-i$ is such that $\{i, -i\} = \{1, 2\}$.

[8] Our definition is more restrictive than, for example, those given in Thomson [20], Sobel [18], or Peters [15].

For the proof, we refer the reader to Billera and Bixby [2] or Chipman and Moore [4].

Lemma 1. *Given two utility functions $\tilde{u}_i, \tilde{u}_{-i} \in \tilde{\mathbf{U}}$, the function $h(\tilde{u}_i, \tilde{u}_{-i})$ is continuous, strictly decreasing, and satisfies $h(\tilde{u}_i, \tilde{u}_{-i})(0) = 1$, $h(\tilde{u}_i, \tilde{u}_{-i})(1) = 0$. Furthermore, if $\tilde{u}_i$ and $\tilde{u}_{-i}$ are (strictly) concave then $h(\tilde{u}_i, \tilde{u}_{-i})$ is (strictly) concave.*

Suppose that, once the reports are given, an agent has the chance to revise her strategy. Assume that she wants to transform the UPS into one whose Pareto-frontier is characterized by a different shape. Our first proposition states that this agent is able to report a utility function such that, without modifying her ordinal preferences, achieves the desired transformation.

Proposition 1. *Given $\tilde{u}_i, \tilde{u}_{-i} \in \tilde{\mathbf{U}}$ and $\hat{h} : [0,1] \longrightarrow [0,1]$ a continuous, strictly decreasing (hence invertible) function such that $\hat{h}(0) = 1$ and $\hat{h}(1) = 0$, let $\hat{u}_i = \hat{h}^{-1} \circ h(\tilde{u}_i, \tilde{u}_{-i}) \circ \tilde{u}_i$. Then $h(\hat{u}_i, \tilde{u}_{-i}) = \hat{h}$. Additionally, for any $x, y \in \mathbb{R}^n_+$, $\tilde{u}_i(x) \geq \tilde{u}_i(y)$ if and only if $\hat{u}_i(x) \geq \hat{u}_i(y)$.*

The remaining part of this section determines what Pareto-frontier shape is obtained in a Nash equilibrium. Our reasoning is based on the crucial role (first studied by Sobel [18]) that midpoint dominance plays in distortion games. Given a report $\tilde{u}_{-i} \in \tilde{\mathbf{U}}$, this property restricts the final bundle of agent i to the compact and convex set defined by $F_i(\tilde{u}_{-i}) = \{x_i \in X \mid \tilde{u}_{-i}(\omega - x_i) \geq \frac{1}{2}\}$. We call this the *midpoint dominance restriction.* If $(\tilde{x}_i, \tilde{x}_{-i}) \in \psi(\tilde{u}_i, \tilde{u}_{-i})$ agent i wants $\tilde{x}_i$ to lie on the frontier of $F_i(\tilde{u}_{-i})$. When the strategy sets just include continuous, strictly increasing, concave utility functions, our next example shows this goal is not always achievable.

Example 2. *Let $n = 1$, $\omega = 1$, $u_1(y) = y$, $u_2(y) = \tilde{u}_2(y) = \sqrt{y}$, and $\phi(S, d) = NB(S, d)$. Then $F_1(\tilde{u}_2) = [0, \frac{3}{4}]$. If non-concave reports are ruled out, the best response for agent 1 is to report $\tilde{u}_1(y) = y$ and obtain $\frac{\sqrt{5}-1}{2}$ units of the good, which is less than $\frac{3}{4}$ units.*

Adding non-concave reports to the strategy space changes things substantially. In this case, agents can force the midpoint dominance restriction to be binding.

Proposition 2. *Given the distortion game $G(u_1, u_2, \phi) \in DG$ and $\tilde{u}_{-i} \in \tilde{\mathbf{U}}$, define $\forall x_i \in \mathbb{R}^n_+$, $\hat{u}_i(x_i) = 1 - (h(u_i, \tilde{u}_{-i}) \circ u_i)(x_i)$. Then $(\hat{x}_i, \hat{x}_{-i}) \in \psi(\hat{u}_i, \tilde{u}_{-i})$ implies $\tilde{u}_{-i}(\hat{x}_{-i}) = \frac{1}{2}$. Furthermore, $\hat{u}_i \in BR_i(\tilde{u}_{-i})$.*

When both agents best respond the utility pair selected by the bargaining solution concept (using the reported utilities) must be $(\frac{1}{2}, \frac{1}{2})$. We use this fact to show that a necessary condition for a pair of reports to be a Nash equilibrium is a linear Pareto-frontier.

Proposition 3. *Let (u_1^*, u_2^*) be a Nash equilibrium of the distortion game $G(u_1, u_2, \phi) \in DG$. Then, $\forall v_i \in [0,1]$ $\forall i \in \{1, 2\}$, $h(u_i^*, u_{-i}^*)(v_i) = 1 - v_i$.*

4 Efficiency of Equilibrium Outcomes

As an introduction to our efficiency results we examine an example, taken from Sobel [18] and also mentioned in Peters [15]. It shows that the outcome of a distortion game may not even be Pareto-optimal when agents are just allowed standard concave reports.

Example 3. *Let* $n = 2$, $\omega = (1,1)$, $u_1(y) = y_1^{\frac{5}{6}} y_2^{\frac{1}{6}}$, $u_2(y) = y_1^{\frac{1}{2}} y_2^{\frac{1}{2}}$, *and* $\phi(S,d) = NB(S,d)$. *If* $u_1^*(y) = \frac{1}{8}(5y_1+3y_2)$ *and* $u_2^*(y) = y_1^{\frac{1}{2}} y_2^{\frac{1}{2}}$ *then* $\psi(u_1^*, u_2^*) = \{((\frac{3}{5}, \frac{1}{3}), (\frac{2}{5}, \frac{2}{3}))\}$. *With just concave reports Sobel shows that* (u_1^*, u_2^*) *is a Nash equilibrium of the described distortion game. Still, the outcome is Pareto dominated by the allocation* $((\frac{2}{3}, \frac{1}{5}), (\frac{1}{3}, \frac{4}{5}))$.

When non-concave reports are allowed, we show that efficiency is restored. The proof technique we use is similar to the one shown in Peters [15].

Proposition 4. *Let* (u_1^*, u_2^*) *be a Nash equilibrium of* $G(u_1, u_2, \phi) \in DG$. *Then, any interior outcome* $(x_1^*, x_2^*) \in \psi(u_1^*, u_2^*)$ *is Pareto optimal with respect to the true utility functions* (u_1, u_2).

The following examples show that it is not possible to dispose of the assumptions we make.

Example 4. *Let* $n = 2$, $\omega = (1,1)$, $u_1(y) = \frac{y_1+y_2}{2}$, $u_2(y) = \frac{2y_1+y_2}{3}$, *and* $\phi \in \Phi$. *Agent 1 reports* $u_1^*(y)$ *defined as follows: If* $\frac{2y_1+y_2}{3} \leq \frac{2}{3}$ *then* $u_1^*(y) = \frac{2y_1+y_2}{4}$. *Otherwise, if* $\frac{2y_1+y_2}{3} > \frac{2}{3}$, $u_1^*(y) = \frac{2y_1+y_2-1}{2}$. *Agent 2 reports* $u_2^*(y) = \frac{y_1+y_2}{2}$. *It can be verified that* (u_1^*, u_2^*) *is a Nash equilibrium for the distortion game* $G(u_1, u_2, \phi)$. *The set of outcomes is* $\psi(u_1^*, u_2^*) = \{((1,0),(0,1))\}$, *but this allocation is Pareto inferior to* $((\frac{7}{12}, \frac{7}{12}), (\frac{5}{12}, \frac{5}{12}))$. *Without interiority the outcome is not necessarily Pareto optimal with respect to the true utility functions.*

Example 5. *Let* $n = 2$, $\omega = (1,1)$, $u_1(y) = \frac{y_1+y_2}{2}$, $u_2(y) = y_1^{\frac{1}{6}} y_2^{\frac{5}{6}}$, *and* $\phi \in \Phi$. *Both agents report* $u_1^*(y) = u_2^*(y) = \min(y_1, y_2)$. *It can be verified that* (u_1^*, u_2^*) *is a Nash equilibrium for the distortion game* $G(u_1, u_2, \phi)$. *The set of outcomes is* $\psi(u_1^*, u_2^*) = \{((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))\}$, *but this allocation is Pareto inferior to* $((\frac{3}{4}, \frac{3}{10}), (\frac{1}{4}, \frac{7}{10}))$. *Notice that the reported utility functions are not strictly monotone, but this is not essential for the counterexample to remain valid. Thus, smoothness is necessary to obtain an efficient outcome.*

Proposition 4 can be interpreted as an analogous version of the first welfare theorem for the distortion game method of allocating goods. It is then natural to ask if the corresponding version of the second welfare theorem also holds. The answer is affirmative.

Proposition 5. *Let the outcome* $(x_1^*, x_2^*) \in X^2$ *be Pareto optimal with respect to the true utility functions* (u_1, u_2). *Then, every distortion game* $G(u_1, u_2, \phi) \in DG$ *has a Nash equilibrium* (u_1^*, u_2^*) *such that* $(x_1^*, x_2^*) \in \psi(u_1^*, u_2^*)$.

Two important concerns arise at this point. First, the results we obtain depend on the outcome correspondence punishing the agents when reports generate a non-convex utility possibility set. This discontinuity of the payoff function puts agents in a situation in which the slightest variation of their strategy might lead to disaster. Of course, bargainers are well aware that too much greed leads to the disagreement outcome, but it is worth trying to express this intuition using less drastic methods. The second concern is the multiplicity of possible outcomes shown in Proposition 5. Any Pareto efficient allocation may result from Nash equilibrium behavior. Many "unfair" commodity distributions, although efficient, might arise from distortion games that allow non-concave reports. It is desirable to refine the set of equilibrium outcomes.

5 Selecting Among Outcomes

We intend to deal with the previous concerns by viewing our bargaining procedure as an application of the Nash demand game (See Nash [13]) within a multi-commodity environment. Using our notation, the Nash demand game is played between two agents with utility functions $u_i \in \mathbf{U}$ that want to distribute a pie of size one. Each agent simultaneously reports a minimum demanded level of utility $\bar{u}_i \in [0, 1]$. If $(\bar{u}_1, \bar{u}_2) \in UPS(u_1, u_2)$, then agents obtain their demands. Otherwise, the outcome is given by the vector $(0, 0)$. It is interesting that exactly the same issues mentioned in the previous paragraph are also applicable for the Nash demand game.

To address these objections Nash smooths the payoffs of the agents and then analyzes Nash equilibria as the smoothed game approximates to the original. Several variants of this procedure exist in the literature. We will adapt the method presented by Osborne and Rubinstein [14] to our distortion game setting. To do so, we must redefine the outcome and best response correspondences.

The *modified outcome correspondence* is defined as follows. If $F_1(\tilde{u}_2) \cap \big(\{\omega\} - F_2(\tilde{u}_1)\big) \neq \emptyset$, then $\psi'(\tilde{u}_1, \tilde{u}_2) = \{(x_1, x_2) \in X^2 \mid x_1 + x_2 \leq \omega \text{ and } \forall i \in \{1, 2\},\ \tilde{u}_i(x_i) = \frac{1}{2}\}$. Otherwise, $\psi'(\tilde{u}_1, \tilde{u}_2) = \{(\vec{0}, \vec{0})\}$. In other words, if it is not feasible to satisfy the demands made by both agents, they get nothing. On the other hand, whenever possible, the arbitrator assigns each agent a bundle that is barely sufficient to satisfy their reported midpoint dominance restrictions. Our first task is to verify that the modified outcome correspondence is well defined.

Lemma 2. *Given any pair of utility functions* $(\tilde{u}_1, \tilde{u}_2) \in \tilde{\mathbf{U}}^2$*,* $\psi'(\tilde{u}_1, \tilde{u}_2) \neq \emptyset$*.*

Although this modified distortion game does not depend on a specific bargaining solution ϕ, it is quite similar to the version described in Section 2. Best-responding agents in the original game are indifferent between the arbitrator using ψ (with respect to some $\phi \in \Phi$) or ψ' to determine the final outcome. Indeed, Proposition 2 shows that best responses imply binding midpoint dominance restrictions. Then, any Nash equilibrium of the original distortion game will remain an equilibrium with the modified outcome correspondence. Further, equilibrium payoffs also remain unchanged.

Now we smooth the payoff functions of the agents. Given $u_1, u_2 \in \mathbf{U}$, the function $P : \mathbb{R}^2_+ \longrightarrow [0,1]$ is a *perturbation with respect to the pair* (u_1, u_2) if it satisfies: (i) If $v \notin UPS(u_1, u_2)$, then $P(v) = 0$, (ii) If v lies in the interior of $UPS(u_1, u_2)$, then $P(v) > 0$, and (iii) P is differentiable (thus continuous) and quasiconcave. We now give some intuition about perturbations. Assume that agents are slightly uncertain about the true utility functions of their opponents or about the size of the bundle ω. As observed in Osborne and Rubinstein [14], the function P can be interpreted as the probability of a deal being reached. When demands from agents grow closer to the boundary of the UPS there is a bigger chance of facing disagreement. Of course, if demands are incompatible the disagreement outcome is certain.

Given $u_1, u_2 \in \mathbf{U}$ and P a perturbation with respect to (u_1, u_2), the *perturbed distortion game* $G(u_1, u_2, P)$ is played by agents reporting utility functions $\tilde{u}_1, \tilde{u}_2 \in \tilde{\mathbf{U}}$. An arbitrator then uses the modified outcome correspondence to determine the allocation of the goods. Agents are now interested on maximizing their expected utility.

For any perturbed distortion game $G(u_1, u_2, P)$ and any agent $i \in \{1, 2\}$, define the *perturbed best response correspondence* by letting $BR_i : \tilde{\mathbf{U}} \longrightarrow \tilde{\mathbf{U}}$ as $BR_i(\tilde{u}_{-i}) = \{\hat{u}_i \in \tilde{\mathbf{U}} \mid \forall (\hat{x}_i, \hat{x}_{-i}) \in \psi(\hat{u}_i, \tilde{u}_{-i}), \forall \tilde{u}_i \in \tilde{\mathbf{U}}, \forall (\tilde{x}_i, \tilde{x}_{-i}) \in \psi(\tilde{u}_i, \tilde{u}_{-i}),$ $u_i(\hat{x}_i)P(u_i(\hat{x}_i), u_{-i}(\hat{x}_{-i})) \geq u_i(\tilde{x}_i)P(u_i(\tilde{x}_i), u_{-i}(\tilde{x}_{-i}))\}$ for any $\tilde{u}_{-i} \in \tilde{\mathbf{U}}$. The pair $(u_1^*, u_2^*) \in \tilde{\mathbf{U}}^2$ is a *Nash equilibrium* of $G(u_1, u_2, P)$ if $u_i^* \in BR_i(u_{-i}^*)$ for $i = 1, 2$. Once again we may show existence of equilibrium thanks to the richness of our strategy space.

Proposition 6. *Any perturbed distortion game $G(u_1, u_2, P)$ has a Nash equilibrium.*

Define a discontinuous function $\bar{P} : \mathbb{R}^2_+ \longrightarrow [0,1]$ by letting $\bar{P}(v) = 1$ if $v \in UPS(u_1, u_2)$ and $\bar{P}(v) = 0$ otherwise. Construct a sequence of perturbed games $\{G^k(u_1, u_2, P^k)\}_{k=1}^{\infty}$ such that P^k converges towards $\bar{P}$. The degree of closeness between a perturbation P^k and the limit $\bar{P}$ is measured by the (Hausdorff) distance between the true UPS and the set of utility pairs that, given P^k, yield an agreement with probability one. In short, the sequence of perturbed games converges towards the modified distortion game. The question at hand is if any sequence of equilibrium outcomes (x_1^{*k}, x_2^{*k}) converges and if so, what is the limit. Our last proposition answers both questions.

Proposition 7. *Let $\{G^k(u_1, u_2, P^k)\}_{k=1}^{\infty}$ be a sequence of perturbed distortion games such that the Hausdorff distance between $UPS(u_1, u_2)$ and the set $\{v \in \mathbb{R}^2_+ \mid P^k(v) = 1\}$ converges to zero as $k \longrightarrow \infty$. For any $k \geq 1$, let (u_1^{*k}, u_2^{*k}) be a Nash equilibrium of $G^k(u_1, u_2, P^k)$ and choose $(x_1^{*k}, x_2^{*k}) \in \psi'(u_1^{*k}, u_2^{*k})$. Then, the limit as $k \longrightarrow \infty$ of $\big(u_1(x_1^{*k}), u_2(x_2^{*k})\big)$ is the Nash solution $NB(UPS(u_1, u_2), (0,0))$.*

To quote Serrano [17] when referring to the Nash demand game, "we learn that in negotiations with uncertainty [...] but where beliefs are concentrated around the truth, Nash equilibrium [...] yield[s] an outcome that gravitates towards the Nash solution." We have managed to demonstrate the same result when the preferences of the agents are unknown to the mediator.

6 Final Remarks

We conclude by describing how the results in this paper are related to three different but closely related research avenues: The Nash program, implementation theory, and distortion games. The Nash program intends to give non-cooperative support to cooperative solution concepts which are usually based on a system of axioms. In the case of the Nash bargaining solution, Nash gives such support by analyzing smoothed demand games. The result of his procedure is a pair of utility levels that complies with the axioms but depends crucially on the preferences of agents. Bargainers then have incentives to misrepresent their true utility function, most probably violating the axioms. For this reason implementation theory takes to the task of creating mechanisms that are founded not on preferences but on outcomes.

The output of the mechanism we describe, the allocation of a multi-commodity bundle, is completely independent of agents' preferences. In that sense, our work can be understood as a contribution to Nash bargaining implementation. The Nash bargaining solution does not satisfy Maskin-monotonicity (c.f. Maskin [11]), a necessary condition for exact implementation of a concept (see Serrano [16]). From this perspective, our approximate implementation result is the best we could have hoped for.

The literature on distortion games coincides with the spirit of implementation as it studies manipulation of private information by bargainers. The difference between them is that implementation theory does not put restrictions on the structure of the game that generates the desired outcomes. Given that a distortion game requires an arbitrator to apply the original concept on the reported utility functions, it comes as no surprise that up to now the related results have been limited to preserve basic properties such as efficiency. The main contribution of this paper is to achieve virtual implementation in the very intuitive but restrictive setting of distortion games.

Appendix

Proof of Proposition 1: As $\hat{h}$ is strictly decreasing, its inverse also has this property. The function $\hat{h}^{-1} \circ h(\tilde{u}_i, \tilde{u}_{-i})$ (applied to $\tilde{u}_i$ to transform it into $\hat{u}_i$) is a monotone transformation because it is the composition of two strictly decreasing functions. Consequently, the ordinal preferences of agent i remain unchanged.

Fix $v_i \in [0,1]$ and let $\bar{x}_i$ maximize $\tilde{u}_{-i}(\omega - x_i)$ subject to $\hat{u}_i(x_i) = v_i$. By definition of $\hat{u}_i$ we have $h(\tilde{u}_i, \tilde{u}_{-i})(\tilde{u}_i(\bar{x}_i)) = \hat{h}(v_i)$. Using the fact that the allocation $(\bar{x}_i, \omega - \bar{x}_i)$ must be Pareto optimal with respect to the reports $(\tilde{u}_i, \tilde{u}_{-i})$, we conclude that $\tilde{u}_{-i}(\omega - \bar{x}_i) = \hat{h}(v_i)$. Hence, $h(\hat{u}_i, \tilde{u}_{-i})(v_i) = \hat{h}(v_i)$.

Proof of Proposition 2: By Proposition 1, the Pareto frontier of $UPS(\hat{u}_i, \tilde{u}_{-i})$ is linear, so $\phi(UPS(\hat{u}_i, \tilde{u}_{-i}), (0,0)) = (\frac{1}{2}, \frac{1}{2})$. Efficiency of ϕ implies that any outcome $(\hat{x}_i, \tilde{x}_{-i}) \in \psi(\hat{u}_i, \tilde{u}_{-i})$ is Pareto optimal with respect to the pair of reports $(\hat{u}_i, \tilde{u}_{-i})$. As $\hat{u}_i$ is a monotone transformation of u_i, the outcome is also optimal with respect to $(u_i, \tilde{u}_{-i})$. We conclude that $\hat{x}_i$ must maximize u_i over $F_i(\tilde{u}_{-i})$. Midpoint dominance of ϕ then implies $\hat{u}_i \in BR_i(\tilde{u}_{-i})$.

Proof of Proposition 3: The utility possibility set $UPS(u_1^*, u_2^*)$ must be convex, otherwise agents have incentives to deviate from the $(\vec{0}, \vec{0})$ outcome. Let $(x_1^*, x_2^*) \in \psi(u_1^*, u_2^*)$. Proposition 2 implies $\phi(UPS(u_1^*, u_2^*), (0,0)) = (\frac{1}{2}, \frac{1}{2})$.

In what follows we show that if $\phi(S,d) = (\frac{1}{2}, \frac{1}{2})$ and (S,d) is a 0-1 normalized bargaining problem, then the Pareto frontier of S is linear. Suppose not. Then, S must contain at least one vector (v_1, v_2) such that $v_1 + v_2 > 1$. Without loss of generality, assume $v_2 > v_1$. Convexity implies that $(\frac{v_2}{1+v_2-v_1}, \frac{v_2}{1+v_2-v_1})$ is in S as it is a linear combination of (v_1, v_2) and $(1,0)$. Furthermore, $v_1 + v_2 > 1$ implies that $\frac{v_2}{1+v_2-v_1} > \frac{1}{2}$. This contradicts the efficiency of ϕ.

Proof of Proposition 4: Let $(x_1^*, x_2^*) \in \psi(u_1^*, u_2^*)$ be such that $x_i^* \in \mathbb{R}^n_{++}$ for $i = 1, 2$. By efficiency of ϕ, (x_1^*, x_2^*) is Pareto optimal with respect to the reports (u_1^*, u_2^*). This means the convex sets $F_1(u_2^*)$ and $\{\omega\} - F_2(u_1^*)$ have disjoint interiors and there exists a separating hyperplane H between them. The fact that $u_1^* \in BR_1(u_2^*)$ implies there is no $(x_1, x_2) \in X^2$ such that $x_1 + x_2 = \omega$, $u_2^*(x_2) > u_2^*(x_2^*) = \frac{1}{2}$, and $u_1(x_1) > u_1(x_1^*)$. Thus, the convex sets $F_1(u_2^*)$ and $\{x_1 \in X \mid u_1(x_1) \geq u_1(x_1^*)\}$ have disjoint interiors and there exists a separating hyperplane H_1 between them. A symmetric argument shows that the convex sets $\{\omega\} - F_2(u_1^*)$ and $\{\omega\} - \{x_2 \in X \mid u_2(x_2) \geq u_2(x_2^*)\}$ have disjoint interiors and there exists a separating hyperplane H_2 between them.

The smoothness and interiority assumptions imply that $H = H_1 = H_2$. Consequently, H also separates the interiors of $\{x_1 \in X \mid u_1(x_1) \geq u_1(x_1^*)\}$ and $\{\omega\} - \{x_2 \in X \mid u_2(x_2) \geq u_2(x_2^*)\}$. This makes allocation (x_1^*, x_2^*) Pareto optimal with respect to the true utility functions, as we wanted.

Proof of Proposition 5: Define u_1^* as a function that represents the same ordinal preferences as u_1 and satisfies $u_1^*(\vec{0}) = 0$, $u_1^*(\omega) = 1$, and $u_1^*(x_1^*) = \frac{1}{2}$. Use Proposition 2 to construct u_2^* in such a way that, without changing the true ordinal preferences of agent 2, $h(u_1^*, u_2^*)$ is linear. As (x_1^*, x_2^*) is Pareto optimal with respect to (u_i, u_{-i}^*), $u_i^* \in BR_i(u_{-i}^*)$ for $i = 1, 2$, so (u_1^*, u_2^*) is a Nash equilibrium of $G(u_1, u_2, \phi)$. Finally, $u_2^*(x_2^*) = \frac{1}{2}$ because (x_1^*, x_2^*) is Pareto optimal with respect to (u_1^*, u_2^*) and $h(u_1^*, u_2^*)$ is linear. Therefore, $(x_1^*, x_2^*) \in \psi(u_1^*, u_2^*)$.

Proof of Lemma 2: Assume there exists $x_1 \in F_1(\tilde{u}_2) \cap \{\omega\} - F_2(\tilde{u}_1)$ as otherwise the result is trivial. Define $x_2 = \omega - x_1$. Then, for any $i \in \{1,2\}$, $\tilde{u}_i(x_i) \geq \frac{1}{2}$. Let $\alpha_i = \min\{\alpha \in [0,1] \mid \tilde{u}_i(\alpha x_i) \geq \frac{1}{2}\}$. We conclude that $(\alpha_1 x_1, \alpha_2 x_2) \in \psi'(\tilde{u}_1, \tilde{u}_2)$ because $\tilde{u}_i(\alpha_i x_i) = \frac{1}{2}$ and $\alpha_1 x_1 + \alpha_2 x_2 \leq x_1 + x_2 = \omega$.

Proof of Proposition 6: Let (v_1^*, v_2^*) maximize the continuous function $v_1 v_2 P(v_1, v_2)$ over the compact set $UPS(u_1, u_2)$. It must be the case that $v_1^* v_2^* P(v_1^*, v_2^*) > 0$, so v_i^* maximizes $v_i P(v_i, v^*_{-i})$ over $[0,1]$. For any $i \in \{1,2\}$ define $u^*_i \in \tilde{U}$ as a monotone transformation of u_i such that $u^*_i(x_i) = \frac{1}{2}$ if and only if $u_i(x_i) = v^*_i$. We claim that (u^*_1, u^*_2) is a Nash equilibrium of $G(u_1, u_2, \phi, P)$. Indeed, suppose that $u^*_i \notin BR_i(u^*_{-i})$. Then, there exist $x^* \in \psi'(u^*_i, u^*_{-i})$, $\tilde{u}_i \in \tilde{U}$, and $\tilde{x} \in \psi'(\tilde{u}_i, u^*_{-i})$ such that $u_i(\tilde{x}_i)P(u_i(\tilde{x}_i), u_{-i}(\tilde{x}_{-i})) > u_i(x^*_i)P(u_i(x^*_i), u_{-i}(x^*_{-i}))$. The definition of u^*_{-i} implies that $u_{-i}(\tilde{x}_{-i}) = u_{-i}(x^*_{-i}) = v^*_{-i}$. This implies that v^*_i does not maximize $v_i P(v_i, v^*_{-i})$ over $[0,1]$, a contradiction.

Proof of Proposition 7: For any $k \in \mathbb{N}$, let (u^{*k}_1, u^{*k}_2) be a Nash equilibrium of the perturbed game $G^k(u_1, u_2, \phi, P^k)$ and $(x^{*k}_1, x^{*k}_2) \in \psi'(u^{*k}_1, u^{*k}_2)$. Define $\pi^{*k} = P^k(u_1(x^{*k}_1), u_2(x^{*k}_2))$. The number π^{*k} is well defined because the best response definition implies that, at a particular Nash equilibrium, any outcome generates the same utility for each agent. Osborne and Rubinstein [14] show that if v^*_i maximizes $v_i P(v_i, v^*_{-i})$ over $[0,1]$ for $i \in \{1,2\}$, then (v^*_1, v^*_2) maximizes $v_1 v_2$ subject to $P(v_1, v_2) \geq P(v^*_1, v^*_2)$. Thus, $(u_1(x^{*k}_1), u_2(x^{*k}_2))$ maximizes $v_1 v_2$ over $\{(v_1, v_2) \in \mathbb{R}^2_+ \mid P^k(v_1, v_2) \geq \pi^{*k}\}$. Define the set $P^k_1 = \{(v_1, v_2) \in \mathbb{R}^2_+ \mid P^k(v_1, v_2) = 1\}$. Then $(u_1(x^{*k}_1), u_2(x^{*k}_2))$ belongs to the region $\{(v_1, v_2) \in \mathbb{R}^2_+ \mid \forall \bar{v} \in P^k_1,\ \ v_1 v_2 \geq \bar{v}_1 \bar{v}_2\} \cap UPS(u_1, u_2)$, which as $k \to \infty$ converges to $NB(UPS(u_1, u_2), (0,0))$.

References

[1] Aumann R. J. and Kurz M. (1977): "Power and Taxes," *Econometrica* Vol. 45, 1137–1161.

[2] Billera L. J. and Bixby R. E. (1973): "A Characterization of Pareto surfaces," *Proceedings of the American mathematical Society* Vol. 41, 261–267.

[3] Binmore K. (1987): "Nash bargaining theory I, II," in: *The economics of bargaining.* Binmore, K. and Dasgupta, P. editors, Cambridge, Basic Blackwell.

[4] Chipman J. S., and Moore J. C. (1972): "Social Utility and the Gains From Trade," *Journal of International Economics* Vol. 2, 157–172.

[5] Crawford V. P., and Varian H. R. (1979): "Distortion of Preferences and the Nash Theory of Bargaining," *Economic Letters* Vol. 3, 203–206.

[6] Hurwicz L. (1972): "On Informationally Decentralized Systems," in: *Decision and organization (a volume in honor of Jacob Marschak), Studies in Mathematical and Managerial Economics,* Vol. 12, 297–336. Amsterdam, North-Holland.

[7] Kalai E. and Smorodinsky M. (1975):, "Other Solutions to Nash's Bargaining Problem" *Econometrica*, vol 43, 513–518.

[8] Kannai Y. (1977): "Concavifiability and Constructions of Concave Utility Functions," *Journal of Mathematical Economics*, vol 4, 1–56.

[9] Kibris O. (2002): "Misrepresentation of Utilities in Bargaining: Pure Exchange and Public Good Economies," *Games and Economic Behavior*, vol 39, 91–110.

[10] Kurz M. (1977): "Distortion of Preferences, Income Distribution, and the Case for a Linear Income Tax," *Journal of Economic Theory*, vol 14, 291–298.

[11] Maskin E. (1977): "Nash Equilibrium and Welfare Optimality," mimeo, MIT.

[12] Nash J. F. (1950): "The Bargaining Problem," *Econometrica*, vol 18, 2, 155–162.

[13] Nash J. F. (1953): "Two-Person Cooperative Games," *Econometrica*, vol 21, 1, 128–140.

[14] Osborne, M.J. and Rubinstein A. (1990): *Bargaining and markets*, Ch 4. New York, Academic Press.

[15] Peters H. (1992): "Self-Optimality and Efficiency in Utility Distortion Games," *Games and Economic Behavior*, vol 4, 252–260.

[16] Serrano R. (1997): "A Comment on the Nash Program and the Theory of Implementation," *Economic Letters*, vol 55, 203–208.

[17] Serrano R. (2004): "Fifty Years of the Nash Program, 1953-2003," Working Paper No. 2004-21, Brown University.

[18] Sobel J. (1981): "Distortion of Utilities and the Bargaining Problem," *Econometrica*, vol 49, 3, 597–619.

[19] Sobel J. (2001): "Manipulation of Preferences and Relative Utilitarianism," *Games and Economic Behavior*, vol 37, 1, 196–215.

[20] Thomson W. (1984): "The Manipulability of Resource Allocation Mechanisms," *Review of Economic Studies*, vol 51, 3, 447–460.

[21] Thomson W. (1988): "The Manipulability of the Shapley-Value," *International Journal of Game Theory*, vol 17, 2, 101–127.

[22] Thomson W. (1994): "Cooperative Models of Bargaining," *Handbook of Game Theory with Economic Applications*, vol 3. New York, NY: North Holland, 1237–1284.

[23] van Damme E. (1987): *Stability and perfection of Nash equilibria*. Berlin - New York, Springer-Verlag.