

Fair Bargains: Distributive Justice and Nash Bargaining Theory

by Marco Mariotti

Department of Economics, Royal Holloway College, University of London,
Egham, Surrey TW20 0EX, UK.

Abstract: the Suppes-Sen dominance relation is a weak and widely accepted criterion of distributive justice. I propose its application to Nash bargaining theory. The Nash Bargaining Solution (NBS) is characterised by replacing the controversial Independence of Irrelevant Alternatives axiom with an axiom embodying the Suppes-Sen principle. The characterisation is more robust than the standard one with respect to variations in the domain of bargaining problems. It is also shown that a subset of Nash's axioms imply the Suppes-Sen relation.

JEL class. number: C71, C78.

I am deeply indebted, with the usual disclaimers, to Paul Madden for comments, bibliographical suggestions, encouragement, and not least for originating my interest in the Suppes-Sen relation. Useful suggestions have also come from Vincenzo Denicolò, Joseph Greenberg, Paola Manzini, Herve Moulin, Martin Osborne, Fioravante Patrone, Hyun Shin, Stefano Zamagni, two anonymous referees and from the seminar audiences at the 1997 Summer Meeting of the Econometric Society in Pasadena and at University College London.

1. Introduction

In the axiomatic theory of bargaining initiated by Nash (1950) one defines a set of admissible bargaining problems (e.g., convex problems) and imposes some desirable requirements (axioms) on the solution function, which picks an element from each problem; the aim is to characterise uniquely such a function. One interpretation of the axioms is as properties that should be satisfied by the choices of a *fair arbitrator* (e.g., Myerson (1990, p. 372-3), Young (1994, ch. 7), Mas Colell *et. al.* (1995, ch. 22.E) and, implicitly, Hammond (1991), p. 203). The axioms that characterise the Nash Bargaining Solution (NBS) include the powerful Independence of Irrelevant Alternative (IIA), which has been extensively discussed and criticised. Indeed, although IIA and its variations may be viewed as a relevant criterion of *rationality* for individual choice, it is difficult to see it as a compelling requirement of *fair arbitration*. As Binmore (1992, p.196) puts it:

“Some authors misunderstand Nash's motives in formulating his bargaining solution and imagine that his axioms can be sensibly interpreted as criteria for a 'fair arbitration scheme' ... other axiom systems have been introduced to characterise other so-called 'bargaining solutions' that do make sense as fair arbitration schemes”.

I agree with Binmore's statement concerning Nash's *axioms*. In this paper I aim to show that - whatever Nash's motives were - there are nonetheless strong reasons to interpret Nash's *solution* as the expression of a fair arbitrator's decisions under certain informational conditions. I will illustrate a striking property of the NBS, which supports this interpretation and allows a characterisation that dispenses with IIA altogether. It turns out that the NBS is the only scale covariant bargaining solution satisfying the criterion of distributive justice known as *Suppes-Sen dominance* (Suppes (1966), Sen (1970), chs. 9 and 9*)^{1,2}.

¹ The most notable modification of Nash's axioms concerning IIA is by Lensberg (1988). He replaces IIA with Stability, which applies when the number of players is a variable.

Given utility vectors x and y , x is said to *Suppes-Sen dominate* (SS-dominate) y if and only if there exists a permutation of x that Pareto dominates y . I refer to the application of this dominance criterion as the Suppes-Sen principle. There are two main ideas behind the Suppes-Sen principle. First, 'fair' decisions should be, in some sense, *impartial*: if x is considered 'more just' than y , this judgement should not depend on the position of any particular pre-specified individual. Secondly, *interpersonal comparisons* of utility levels are meaningful (recall how, in social choice theory, Arrow's (1951) impossibility theorem and in particular Sen's (1970, p. 123-30) analogous result when *cardinal* intensities of preferences are available imply that utilities must satisfy some form of interpersonal comparability if reasonable and defined social choices are to be made). As far as I am aware, the relevance of the Suppes-Sen principle to Nash bargaining theory viewed as a theory of fair arbitration has not been studied. I propose to do so, as follows.

In Nash's bargaining theory a *disagreement point* is given, which is relevant for the outcome of the problem. On my interpretation, a bargaining problem is a special kind of collective decision problem. A fair arbitrator who accepts ordinal interpersonal utility comparisons will want to apply the Suppes-Sen principle to the players' utilities *net* of the disagreement utility. Many scale-dependent bargaining solutions yield SS-undominated outcomes for each problem - consider, for example, the Utilitarian (Myerson, 1981) and the Egalitarian (Kalai, 1977; Roth, 1979) solutions. At first blush, it might appear that also all scale covariant solutions which satisfy the axioms of Pareto optimality and Anonymity (or Symmetry) will be bound to be compatible with the Suppes-Sen principle, the most obvious candidate being perhaps the Kalai and Smorodinsky solution (KSS), identified with *relative egalitarianism*. Surprisingly, this is not the case. There exists only one scale-covariant solution which yields SS-undominated outcomes in each bargaining problem, and this is the NBS.

The validity of the last assertion as a mere mathematical fact is not in doubt, but one important point of interpretation should be discussed. Scale covariant

² Koim (1971) uses the term "fundamental dominance", Blackorby and Donaldson (1977) simply refer to "dominance". The term 'Suppes-Sen dominance' seems to be due to Saposnik (1983).

bargaining solutions can determine solution outcomes without the use of interpersonal utility comparisons of any sort. What is the meaning, then, of adopting a criterion that hinges on such interpersonal comparisons, albeit of an ordinal nature? My methodological premise is that it is *logically meaningful* to compare utilities across individuals, but that it is *practically* difficult, if not impossible, to obtain empirically the necessary scheme of interpersonal scaling³. I quote from Elster and Roemer (1991, p. 10-11):

“Let us assume that there is a *fact of the matter* in an interpersonal comparison of well-being ... It does not follow that we could ever discover it. Statements about the past pose similar problems. We tend to assume that there is a fact of the matter by virtue of which statements about the past are true if true, false if false. We may never be able to *establish* what the fact of the matter is -for example, whether it was raining when Caesar crossed the Rubicon. But that does not affect the *existence* of a fact of the matter. In one sense, other minds are just as inaccessible to us as the past. We need not entertain doubts about their existence and their essential similarity to our own, but we may despair at ever getting the details right”.

Along these lines, consider an arbitrator who does *not* receive sufficiently detailed information to make interpersonal utility comparisons in practice, but who nonetheless believes that the Suppes-Sen principle is meaningful and is a necessary condition of impartial decision making in the (possibly only theoretical) circumstances in which it can be used. The property of the NBS proved in this paper means that by implementing the NBS, the arbitrator can ensure that, *even if* given the necessary additional information, he would not be found in violation of the Suppes-Sen principle.

Beside yielding a novel characterisation of the NBS which dispenses with one of the controversial axioms in Nash's system, the present approach has the added benefit of being robust to the choice of *domain* of the solution function. Nash (1950)

and is followed in more recent treatments such as Madden (1996).

³ Hammond (1991) also doubts the possibility of an empirical derivation of interpersonally comparable utilities, while supporting their use in social decision making. See also the other essays in Elster and Roemer (1991), in particular Weymark (1991).

confined himself to the class of convex problems; his axioms are not consistent on a wider domain. My characterisation, on the contrary, is robust to different -and, one could argue, more realistic- specifications of the domain⁴. This, I also argue, makes bargaining problems more directly comparable to standard social decision problems.

The rest of the paper is organised as follows. The next section contains the main definitions and result. In section 3 I discuss extensively the the Suppes-Sen principle and the informational context I consider appropriate for the proposed interpretation of the NBS. Section 4 studies the extension of the characterisation to non-standard domains. A strengthening of the Suppes-Sen dominance criterion is briefly discussed in section 5, which concludes.

2. Main Characterisation

In Nash's (1950) theory, an n -person bargaining problem is a pair (S, d) , where $S \subseteq \mathbb{R}^n$ and $d \in S$. The interpretation is that S is the set of *feasible utilities* attainable by the players and d is the *disagreement point* which results if no agreement is attained. In order to enhance expositional clarity, in the main text I make two simplifications:

$$(1) \quad n = 2;$$

$$(2) \quad d = 0 \equiv (0, 0).$$

None of the results depends on (1). In the appendix I show in full how some definitions and the main result generalise. Nothing conceptual is lost by assuming (1) but much is gained in readability. As for (2), this is a much used convention which saves on notation and is also immaterial, provided that the assumption of Scale Covariance made below is transformed to include the weak requirement of *translation covariance*⁵.

⁴ This does not mean, of course, that my characterisation is *logically stronger* just because it applies to a larger domain (although it is logically stronger on the same domain as Nash's).

⁵ This requirement is satisfied by all the main solutions.

A bargaining problem can thus be simply described as a set $S \subseteq \mathfrak{R}^2$. Let Π be a collection of bargaining problems. A *solution on Π* is a function $\varphi: \Pi \rightarrow \mathfrak{R}^2$ such that $\varphi(S) \in S$ for all $S \in \Pi$. The following restrictions on S are standard:

- A1) S is compact;
- A2) S is convex;
- A3) there exists $s \in S$ such that $s > \theta^6$.

Restriction (A2) in particular is not trivial; we shall see later how non-convex sets of feasible alternatives can be included in the domain. Let Γ denote the collection of all bargaining problems satisfying (A1) through (A3). Some well-known properties that can be imposed on a solution are the following:

Weak Pareto Optimality (WPO): $s > \varphi(S) \Rightarrow s \notin S$.

Strong Individual Rationality (SIR): $\varphi(S) > \theta$.

Covariance with Positive Scale Transformations (COV): let $\tau: \mathfrak{R}^2 \rightarrow \mathfrak{R}^2$ be a positive, linear, component by component transformation given by $\tau(x) = (\lambda_1 x_1, \lambda_2 x_2)$, with $\lambda_1, \lambda_2 > 0$, for all $x \in \mathfrak{R}^2$, and for any $X \subset \mathfrak{R}^2$ let $\tau(X) = \{y \in \mathfrak{R}^2 \mid y = \tau(x) \text{ for some } x \in X\}$. Then, $\varphi(\tau(S)) = \tau(\varphi(S))$.

Symmetry (SYM): suppose that $s \in S \Rightarrow (s_2, s_1) \in S$. Then, $\varphi_1(S) = \varphi_2(S)$.

Anonymity (AN): let $\pi: \mathfrak{R}^2 \rightarrow \mathfrak{R}^2$ be a map such that $\pi(x) = (x_2, x_1)$ for all $x \in \mathfrak{R}^2$. Then, $\varphi(\pi(S)) = \pi(\varphi(S))$.

Independence of Irrelevant Alternatives (IIA): $S \subseteq T$ and $\varphi(T) \in S \Rightarrow \varphi(T) = \varphi(S)$.

In this paper I introduce the following axiom, discussed extensively in the next section.

Suppes-Sen Proofness (SSP): $(s_2, s_1) > \varphi(S)$ or $s > \varphi(S) \Rightarrow s \in S$.

⁶ Vector inequalities: $x > y$ iff $x_i > y_i$, for $i = 1, 2$.

The *Nash Bargaining Solution (NBS)* $v: \Gamma \rightarrow \mathfrak{R}^2$ is defined by $v(S) \equiv \arg \max_{s \in S \cap \mathfrak{R}^2} s_1 s_2$.

Nash (1950) proved that the NBS is the only solution on Γ that satisfies WPO, IIA, COV and SYM (or AN). I will refer to these four axioms as *Nash's axioms*. The main result is the following:

Theorem 2.1: a solution $\varphi: \Gamma \rightarrow \mathfrak{R}^2$ satisfies COV and SSP if and only if $\varphi = v$.

Proof: given $S \in \Gamma$, suppose that there existed $s \in S$ with $(s_2, s_1) > v(S)$. Then also $s_1 s_2 > v_1(S) v_2(S)$, a contradiction. This, together with the well-known facts that the NBS satisfies COV and WPO, proves the 'if' part of the statement.

For the 'only if' part, let $S \in \Gamma$ and suppose by contradiction that $s \equiv \varphi(S) \neq v(S)$. I will show that then there exists $T \in \Gamma$ such that $\varphi(T)$ is SS-dominated. If there exists $t \in S$ with $t > s$ we are done, so assume that s is weakly Pareto optimal. Distinguish three cases.

Case 1: $s > \theta$. Given any point $x \in \mathfrak{R}^2_{++}$, let $H(x)$ denote the branch of the symmetric hyperbola going through x , that is, $H(x) = \{y \in \mathfrak{R}^2_{++} \mid y_1 y_2 = x_1 x_2\}$. Clearly, there exists $t \in H(s)$ such that $v(S) > t$. Consider now the positive linear transformation τ defined by $\tau_1(s) = \tau_2(t)$ and $\tau_1(t) = \tau_2(s)$. Such a transformation is defined (not uniquely) by $\tau(x) = (\lambda_1 x_1, \lambda_2 x_2)$ for all $x \in \mathfrak{R}^2$ where: $\lambda_1, \lambda_2 > 0$, $\lambda_1 \lambda_2 = \tau_2(s) / \tau_1(s) = s_2 / s_1$. Since $s, t \in H(s)$, these equations have a solution.

Let $\tau(S) \equiv T$. We have $\tau(v(S)) > \tau(t) = (\tau_2(s), \tau_1(s))$. Therefore $\tau(s)$ is SS-dominated by $\tau(v(S))$ in T , and by SSP it must be $\tau(s) \neq \varphi(T)$. However, by COV it must be $\varphi(T) = \varphi(\tau(S)) = \tau(\varphi(S)) = \tau(s)$, a contradiction.

Case 2: $s_1 = 0$ (the case $s_2 = 0$ is treated analogously; note that it cannot be $s = \theta$ if s is weakly Pareto optimal). Since $v(S) > \theta$, there exists $\alpha > 0$ such that $(\alpha, 0) < v(S)$. Let $\lambda_2 > 0$ be such that $\lambda_2 s_2 = \alpha$. Define the transformation τ by $\tau(x) = (x_1, \lambda_2 x_2)$. Now the argument of the previous case applies to $\tau(S)$.

Case 3: $s_1 > 0, s_2 < 0$. Let $t \in \mathcal{U}^2$, with $t_1 < 0$ and $t_2 > 0$, be such such that $s_1 s_2 = t_1 t_2$. In addition, it is clearly possible to choose such a large negative value for t_1 that $t < v(S)$. Again define $\lambda_1, \lambda_2 > 0$ and τ as in case 1 and argue analogously. \square

3. The Interpretation of Suppes-Sen Proofness

In this section I discuss the interpretation of SSP in a context in which interpersonal comparisons of utility cannot be used.

In general, given utility vectors $s, t \in \mathcal{U}^2$, s is said to *SS-dominate* t if $s > t$ or $(s_2, s_1) > t$. Consider first a situation in which utility levels *are* interpersonally comparable. Then requiring that a choice should not be SS-dominated merely combines an 'anonymity' principle with the Pareto optimality principle: in other words, it is a Pareto principle for utility vectors where all information regarding personal identity has been erased. As such, this requirement expresses impartiality in the use of Pareto optimality, and seems hardly objectionable. Mathematically, it amounts to preventing the choice of utility vector which is first-order stochastically dominated, that is:

Observation 3.1: Given $s, t \in \mathcal{U}^2$, s SS-dominates t if and only if $s^* > t^*$, where, for any $u \in \mathcal{U}^2$, u^* denotes the vector u with its components listed in ascending order.

This result is well-known to hold for general vectors in \mathcal{U}^n (see e.g. Madden 1996 and the references therein) and is immediately checked for the two-person case. First-order stochastic dominance is a universally accepted criterion to rank distributions of magnitudes regarded as interpersonally comparable (such as incomes). Observation 3.1. then shows how weak a requirement SS-dominance is in a context of interpersonal comparability of utility. This is confirmed by the fact that, when applied to social welfare functionals, SS-dominance does not discriminate almost at all. All main social welfare functionals, from maximin to utilitarian, will pick SS-undominated alternatives.

Similar remarks are valid for the SSP axiom applied to bargaining solutions which do not satisfy COV. For these solutions, SSP is an appealing requirement, but

not particularly useful in characterisations. Consider now a collective decision-making context in which:

- There is cardinal utility information; that is, each person's utility is represented by a cardinal equivalence class of utility functions.
- Interpersonal comparisons of utility cannot be made; that is, the collective decision should be invariant with respect to independent cardinal rescalings of people's utilities;
- Interpersonal comparisons of utility are meaningful; that is, there exists an (unknown) rescaling of each person's utility which makes utilities interpersonally comparable.

In this informational context, a utility vector which is SS-undominated when expressed using some particular representative utility functions from the equivalence classes representing preferences may become SS-dominated if the rescaling which makes utilities interpersonally comparable were to be applied. Only those utility vectors that are SS-undominated *for all possible scalings* of utility will guarantee that the Suppes-Sen principle is not violated. This is exactly the type of requirement embodied by SSP in the presence of COV: requiring a chosen utility vector not to be SS-dominated while simultaneously identifying all bargaining problems which are cardinal transformations of each other (as COV does) amounts to requiring that a utility vector should not be dominated for *any* scaling of utility. In other words, imposing SSP in the presence of COV amounts to assuming that the arbitrator will want to *ensure* that he would not violate the Suppes-Sen principle whatever utility units turned out to be those that make utility comparable across persons.

This is a stronger requirement than the Suppes-Sen principle itself. I do not claim it is a *compelling* requirement for fair arbitration. However, it is certainly a *reasonable* principle which allows, in view of Theorem 2.1, a transparent ethical interpretation for the NBS: it is the only solution that (i) does make use of interpersonal utility comparisons; (ii) and guarantees that the solution outcome will not be SS-dominated in the unknown 'correct' scaling of utilities. In addition, SSP is also logically implied by a subset of Nash's axioms on comprehensive problems. Let Γ^{com} be the set of all problems satisfying (A1) through to (A3) and

A4) let $s \leq s' \leq s''$; then $s, s'' \in S \Rightarrow s' \in S$.

Observation 3.2: let $\varphi: \Gamma^{\text{com}} \rightarrow \mathfrak{R}^2$ be a solution satisfying IIA, WPO and SYM. Then, it also satisfies SSP.

Proof: suppose not, and let $S \in \Gamma^{\text{com}}$ be a problem such that $s = \varphi(S)$ is SS-dominated. That is, there exists $t \in S$ with $t > s$ or $t > (s_2, s_1)$. Consider the latter case. Define $T = \text{co}\{0, s, (s_2, s_1)\}$. By convexity and comprehensiveness, $T \subseteq S$, and by IIA $\varphi(T) = s$. Since T is symmetric, by SYM $s_1 = s_2$. Then $t > (s_2, s_1)$ implies $t > s$, contradicting WPO. \square

This observation also makes clear how SSP is related to SYM. As a referee noted, SYM and SSP are conceptually similar axioms in the following sense. Symmetry is not a scale invariant property; yet, the property of a problem being symmetric in some choice of scales is scale invariant. Similarly, while the Suppes-Sen principle is not scale invariant, the property of two points being comparable in some choice of scales is scale invariant.

I conclude this discussion by noticing that, in the informational situation I am considering, if an axiom analogous to SSP were applied in a general social choice context rather than in Nash's bargaining problem, it would lead to an impossibility result. Suppose x and y are 'physical' alternatives and that $u_1(x) > u_1(y)$ and $u_2(x) < u_2(y)$, where u_i is i 's cardinal utility function. Then it obvious that one can find a cardinal rescaling v_2 of u_2 (that is, $v_2 = \alpha u_2 + \beta$ for some $\alpha > 0$ and $\beta \in \mathfrak{R}$) such that:

$$u_1(y) < v_2(x) < v_2(y) < u_1(x).$$

But, similarly, one can find another rescaling w_2 of u_2 such that:

$$w_2(x) < u_1(y) < u_1(x) < w_2(y).$$

In the v_2 scaling, the Suppes-Sen principle would lead one to prefer x to y . But in the w_2 scaling, the Suppes-Sen principle would lead one to prefer y to x . This makes it impossible to require that the choice between the physical alternatives x and y should not depend on the scaling of utility chosen to represent x and y in utility space. It is the presence of a 'reference' point used as the origin that allows us to escape this type of impossibility in bargaining theory.

4. Other Domains and Multisolutions

The assumption that S is convex is usually justified by the fact that alternatives are expressed in von Neumann-Morgenstern utilities and that lotteries are available. These two requirements, although fairly standard, are not always palatable. If players are not expected utility maximisers, or if in some underlying game in strategic form no correlating device is available, or simply if players are not willing or able to randomise at all, the feasible set will not be convex or even a continuum⁷. In addition, as Moulin (1996, p. 126) observes, "It is hard to believe that our search for operational criteria of fairness should be confined to a convex world".

Fortunately, the characterisation in terms of SSP given above is quite robust to variations of the basic setting. Since the set of maximisers of the Nash product is not necessarily a singleton when the domain is not convex, in this section I turn to *multisolutions*. Let Π be a collection of bargaining problems. Then a *multisolution on Π* is a correspondence $\varphi: \Pi \rightarrow \mathfrak{R}^2$ such that $\varphi(S) \subseteq S$ for all $S \in \Pi$. The NBS viewed as a multisolution is defined analogously to the solution; the KSS and the ES are always single-valued when they are well-defined. Some axioms are redefined accordingly; a star indicates that they refer to multisolutions:

Weak Pareto Optimality (WPO*): $s > t \in \varphi(S) \Rightarrow s \notin S$.

Strong Individual Rationality (SIR*): $s \in \varphi(S) \Rightarrow s > \emptyset$.

Symmetry (SYM*): suppose that $s \in S \Rightarrow (s_2, s_1) \in S$. Then, $s \in \varphi(S) \Rightarrow (s_2, s_1) \in \varphi(S)$.

Independence of Irrelevant Alternatives (IIA*): $S \subseteq T$ and $\varphi(T) \cap S \neq \emptyset \Rightarrow \varphi(S) = \varphi(T) \cap S$.

Suppes-Sen Proofness (SSP*): $(s_2, s_1) > t \in \varphi(S)$ or $s > t \in \varphi(S) \Rightarrow s \notin S$.

⁷ After two crucial contributions by Kaneko (1980) and Herrero (1989) the convexity assumption has received much attention recently: see, e.g., Conley and Wilkey (1996), Mariotti (1997a,b) and Zhou (1997).

COV and AN remain unchanged.

I consider two interesting domains. The first is Σ , the class of problems S that satisfy (A1) and (A3) in section 2 and such that, in addition, S is comprehensive (A4 in the previous section). The second domain is Θ , the class of problems S that satisfy (A3) of section 2 and such that, in addition, S contains a *finite* number of alternatives. For comparison, I summarise next some results of Mariotti (1997a):

Theorem 4.1: there exists no solution $\varphi: \Sigma \rightarrow \mathfrak{R}^2$ that satisfies Nash's axioms. There exists however a multisolucion $\varphi: \Sigma \cup \Theta \rightarrow \mathfrak{R}^2$ that satisfies WPO*, COV, SYM* and IIA*. This multisolucion is unique and it is the NBS. Finally, the NBS is also the only multisolucion $\varphi: \Theta \rightarrow \mathfrak{R}^2$ that satisfies these axioms.

Single-valuedness is thus incompatible with Nash's axioms. I also note that there exists no characterisation in terms of WPO*, COV, SYM* and IIA* for the NBS multisolucion $\varphi: \Sigma \rightarrow \mathfrak{R}^2$ (Kaneko 1980 has a characterisation on this domain which involves also an upper-semicontinuity axiom). Σ is the natural domain to consider when randomisations are available but the players are not necessarily expected utility maximisers (Rubinstein *et al.*, 1992), or they cannot correlate their strategies in the underlying strategic form description. Θ is the natural domain to consider when randomisations are not available at all. In the present approach, we have:

Theorem 4.2: a multisolucion $\varphi: \Pi \rightarrow \mathfrak{R}^2$, with $\Pi \in \{\Sigma, \Theta, \Sigma \cup \Theta\}$, satisfies COV and SSP* if and only if $\varphi \subseteq v$. In particular, there exist *solutions* $\varphi: \Pi \rightarrow \mathfrak{R}^2$, with $\Pi \in \{\Sigma, \Theta, \Sigma \cup \Theta\}$, which satisfy COV and SSP, and such solutions are all selections from the NBS.

Here, the notation $\varphi \subseteq v$ means: $\varphi(S) \subseteq v(S)$ for all $S \in \Pi$. The proof of this theorem uses the same argument used for theorem 2.1, so I will not repeat it here.

These results suggest that, unlike the standard characterisation, the characterisation of the NBS proposed here is relatively independent of the precise structural properties of the feasible set in the domain. One of the main advantages of this feature is that it makes bargaining problems more directly comparable to standard social decision problems, in the following sense. In the latter type problem one typically has a given set X of 'physical' alternatives, and individual preferences on X

are then allowed to vary. In traditional bargaining theory, to the contrary, also the set X must be allowed to vary; otherwise, one might not be able to obtain in the feasible domain the problems needed for the proof (e.g., in Nash's 1950 case, one needs a symmetric rectangle to apply IIA). My characterisation overcomes this difficulty, because the *only* axiom which involves comparisons of different bargaining problems is COV: it is perfectly possible, then, to think of the set of physical alternatives as fixed⁸.

5. Concluding remarks

In this paper I have supported and characterised the interpretation of the NBS as an expression of distributive justice⁹. It is the only solution that satisfies a 'conservative' criterion of impartiality in arbitration. In particular, the NBS is the only solution that reconciles two powerful yet conflicting needs: on the one hand, interpersonal comparisons of utility should be not used *in its calculation*; on the other hand, basic principles of fairness relying on such comparisons should not be violated by *its outcomes*. I note here that the NBS satisfies a much stronger 'fairness' criterion than SSP, related to second-order stochastic dominance (or Generalised Lorenz dominance)¹⁰. Given $s, t \in \mathfrak{R}^2$, s is said to *GL-dominate* t if there exists $\alpha \in [0, 1]$ such that $(\alpha s_1 + (1-\alpha)s_2, (1-\alpha)s_1 + \alpha s_2) > t$. It is immediately verified that $\alpha = 0$ or $\alpha = 1$ in the inequality implies SS-dominance.

⁸ The only additional structural requirement needed in the bargaining framework is, of course, the presence of a dominated 'reference' alternative.

⁹ The mentioned characterisation of the NBS in terms of *Consistency* by Lensberg (1988) also lends support, from a different perspective, to this interpretation. See Young (1994, ch. 7) for a discussion of this interpretation, Krishna and Serrano (1996) for the relation with *strategic bargaining*, and Thomson (1990) for a general discussion of the Consistency principle. I should also mention the original approach by Gauthier: in Gauthier (1986) he claimed that the KSS is simultaneously an expression of rational bargaining *and* of fairness, but it appears from Gauthier and Sugden (1993) that he now views the NBS as a better candidate to perform that double role.

¹⁰ I thank Herve Moulin for drawing attention to this point. The terminology follows Madden (1996) and Shorrocks (1983).

Generalised Lorenz Proofness (GLP): $(\alpha s_1 + (1-\alpha)s_2, (1-\alpha)s_1 + \alpha s_2) > \varphi(S)$,
 $\alpha \in [0,1] \Rightarrow s \notin S$.

It is easy to verify that the NBS cannot yield a GL-dominated outcome. By definition, the feasible set S is bounded above at $v(S)$ by the symmetric hyperbola through $v(S)$, $H(v(s))$. The set of points $t = (\alpha v_1(S) + (1-\alpha)v_2(S), (1-\alpha)v_1(S) + \alpha v_2(S))$ with $\alpha \in [0,1]$ is the segment joining $v(S)$ and $(v_2(S), v_1(S))$. This segment - connecting a point of $H(v(s))$ with another point which, being symmetric to the first, is also on $H(v(s))$ - lies entirely above $H(v(s))$. Therefore for any point t that GL-dominates $v(S)$ it must be $t \notin S$. Thus:

Corollary 5.1: let $\varphi: \Gamma \rightarrow \mathfrak{R}^2$ be a solution satisfying COV and SSP. Then it also satisfies GLP.

Clearly, the concept of GL-dominance requires *cardinal* interpersonal comparability of utility, not merely ordinal comparability as required by the Suppes-Sen criterion. However, the interpretation of GPL in the presence of COV is based on observations analogous to those of section 3.

It should be emphasised in conclusion that the interpretation I propose will only be valid in circumstances when the axiomatic solution of a bargaining problem à la Nash can be held to be appropriate for issues of fairness. In particular:

- i) there is one distinguished point (d) which is allowed to be relevant for solving the distributional problem;
- ii) cardinal utility information is available;
- iii) the decision can be made on the basis of utility information alone.

Sen (1970), for example, has discussed situations in which the disagreement point should have no bearing on justice issues (note, however, that the interpretation of d as disagreement point is not necessary; all that matters is that there exists a Pareto dominated point for which Arrow's Independence can be violated). Unfortunately, as shown in section 3, the 'conservative' criterion proposed here will lead to an impossibility result when applied to a general social choice context without a reference point. If (ii) and (iii) are not good assumptions (e.g. Roemer 1986, 1990, 1996), other

methods and procedures for deciding fairly will be more useful (see e.g., Young, 1994) and Brams and Taylor, 1996).

Appendix

All the definitions and arguments for the results of the text generalise easily to the n -person case. In this appendix, by way of illustration, the definitions of SSP and GLP and the proof of Theorem 2.1 are given. The generalisations of domains, standard axioms and solutions are obvious. Given $s, t \in \mathfrak{R}^n$, s is said to *SS-dominate* t if $As > t$ for some permutation matrix A ; s is said to *GL-dominate* t if $As > t$ for some bistochastic matrix A .

Suppes-Sen Proofness (SSP): $As > \varphi(S)$ for some permutation matrix $A \Rightarrow s \in S$.

Generalised Lorenz Proofness (GLP): $As > \varphi(S)$ for some bistochastic matrix $A \Rightarrow s \in S$.

Theorem A.2.1: A solution $\varphi: \Gamma \rightarrow \mathfrak{R}^n$ satisfies COV and SSP if and only if $\varphi = v$.

Proof: 'If': I note a stronger property of v , namely that it satisfies GLP. That this is so follows from the fact that the Nash product is a symmetric increasing concave function and from standard characterisation results available in the literature¹¹.

'Only if': let $S \in \Gamma$ and suppose by contradiction that $s \equiv \varphi(S) \neq v(S)$. I will show that then there exists $T \in \Gamma$ such that $\varphi(T)$ is SS-dominated. If there exists $t \in S$ with $t > s$ we are done, so assume that s is weakly Pareto optimal (hence, in particular, not $s \leq 0$). Distinguish three cases.

¹¹ The relevant result for the assertion is: s GL-dominates t if and only if $f(s) \geq f(t)$ for all increasing symmetric quasi-concave real-valued functions f , with strict inequality for some such f . See for instance Madden (1996, Theorem 2). For general surveys see e.g. Moulin (1988), where the relationship with Shur-convexity is noted.

Case 1: $s > 0$. Given any point $x \in \mathfrak{R}^n$, let $H(x)$ denote the symmetric hyperboloid going through x , that is, $H(x) = \{y \in \mathfrak{R}^n \mid \prod_i y_i = \prod_i x_i\}$. Clearly, there exists $t \in H(s)$ such that $v(S) > t$. In addition it is possible to choose t so that $t_i \neq s_i$ for all $i \in \{1, 2, \dots, n\}$. Consider now a positive linear transformation τ defined by $\tau(s) = A\tau(t)$, where A is the $n \times n$ permutation matrix which moves the i^{th} component to the $(i+1)^{\text{th}}$ place (setting $n+1 = 1$). That is:

$$A \equiv \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ & & & \dots & & \\ 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

I show that such a transformation τ exists (not uniquely). Denote S and T the $n \times n$ diagonal matrices with the components of s and t , respectively, on their diagonal, and denote λ the $n \times 1$ vector of coefficients representing τ (that is, $\tau_i(x) = \lambda_i x_i$ for all $x \in \mathfrak{R}^n$, $i \in \{1, 2, \dots, n\}$). It must be proved that the homogeneous system:

$$(A1) \quad S\lambda = AT\lambda$$

has (a class of) strictly positive solutions in λ . (A1) has nontrivial solutions if and only if:

$$(A2) \quad |S - AT| = 0.$$

We have:

$$K \equiv S - AT = \begin{bmatrix} s_1 & 0 & 0 & \dots & 0 & -t_n \\ -t_1 & s_2 & 0 & 0 & \dots & 0 \\ 0 & -t_2 & s_3 & 0 & \dots & 0 \\ & & & \dots & & \\ 0 & \dots & 0 & -t_{n-2} & s_{n-1} & 0 \\ 0 & 0 & \dots & 0 & -t_{n-1} & s_n \end{bmatrix}$$

Expanding along the first row:

$$(A3) \quad |K| = s_1|M_{11}| + (-1)^{n+1}(-t_n)|M_{1n}|,$$

where M_{ij} denotes the minor of K obtained by removing the i^{th} row and j^{th} column. By the properties of triangular matrices (e.g. Birkhoff and MacLane (1953), p. 303), $|M_{11}| = \prod_{i=2}^n s_i$ and $|M_{1n}| = \prod_{i=1}^{n-1} t_i$. Note that the second term on the RHS of (A3) is negative for all n . Therefore (A2) holds if and only if $\prod_i s_i = \prod_i t_i$ or, equivalently, if and only if $t \in H(s)$. Since t was chosen exactly in this way, (A1) has nontrivial solutions.

Suppose now that λ^* is a nontrivial solution of (A1) and that $\lambda^*_i < 0$ (resp. $= 0$) for some $i \in \{1, 2, \dots, n\}$. This means (by inspection of K and the fact that $s, t > 0$) that $\lambda^*_{i-1} < 0$ (resp. $= 0$). Consequently, $\lambda^*_i < 0$ (resp. $= 0$) for all $i \in \{1, 2, \dots, n\}$. The case $\lambda^* = 0$ is excluded by nontriviality. If $\lambda^* < 0$, then $-\lambda^* > 0$ is also a solution. We conclude that the desired τ exists in this case.

Now let $\tau(S) = T$. We have $\tau(v(S)) > \tau(t) = A^{-1}\tau(s)$. Therefore $A\tau(v(S)) > \tau(s)$, and by SSP it must be $\varphi(T) \neq \tau(s)$. However, by COV it must be $\varphi(T) = \varphi(\tau(S)) = \tau(\varphi(S)) = \tau(s)$, a contradiction.

Case 2: There exists $I \subset \{1, 2, \dots, n\}$ with $s_i = 0$ for $i \in I$ and $s_i > 0$ for $i \in \{1, 2, \dots, n\} \setminus I$. Without loss of generality, write s (possibly relabeling the axes) in such a way that the first k components are positive and the other negative: that is, let k be such that $s_i > 0$ for $1 \leq i \leq k$ and $s_i = 0$ for $k < i \leq n$. Now let $t \in \mathfrak{R}^n$ have components with signs as follows: $t_i > 0$ for $1 \leq i < k$ and $i = n$; $t_i = 0$ for $k \leq i < n$. In addition, let $t < v(S)$ (this is possible since $v(S) > 0$). Define the system (A1) as in case 1. The matrix K now has one or more rows whose entries are all zero (certainly the last row, since $t_n = s_n = 0$), therefore (A1) has nontrivial solutions. If λ^* is a nontrivial solution, by the choice of sign of t we now have that $\lambda_{i-1}\lambda_i = s_i M_{i-1}$ whenever $s_i > 0$ (and hence $t_{i-1} > 0$)¹². Therefore the λ^*_i have all the same sign for $1 \leq i \leq k$. Since the other λ^*_i , $k < i \leq n$, are all free variables, the choice $\lambda^* > 0$ is certainly allowed, and the proof for this case concludes as in case 1.

¹² Setting $1 - 1 = n$.

Case 3: There exists $I \subset \{1, 2, \dots, n\}$ with $s_i < 0$ for $i \in I$ and $s_i > 0$ for $i \in \{1, 2, \dots, n\} \setminus I$. Without loss of generality, let k be such that $s_i > 0$ for $1 \leq i \leq k$ and $s_i < 0$ for $k < i \leq n$. Define $s' \equiv (s_1, s_2, \dots, s_{k-1}, -s_k, -s_{k+1}, \dots, -s_{n-1}, s_n)$. Let $t \in \mathcal{S}$ have the following properties:

- (a) $\text{sign } t_i = \text{sign } s'_i$;
- (b) $\prod_i s'_i = \prod_i t_i$;
- (c) $t < v(S)$.

Given (b), (c) is possible by making the negative components of t sufficiently large in absolute value. At this point the argument proceeds in a way analogous to case 1 and will not be repeated. \square

Observation: By reduction (using induction) to echelon form of K it is easy to see that in fact the null-space of K in case 1 has dimension 1 for all n . The transformation τ is therefore subject to exactly the same degree of freedom as in the two-person case.

References

- Arrow, K.J. (1951) "Social choice and individual values", Wiley, New York.
- Binmore, K. (1992) "Fun and games", D.C. Heath and Company, Lexington, Ma.
- Birkhoff, G. and S. MacLane (1953) "A survey of modern algebra", MacMillan, New York.
- Blackorby, C and D. Donaldson (1977) "Utility vs. equity: some plausible quasi-orderings", *Journal of Public Economics* 7: 365-81.
- Brams, S. and A. Taylor (1996) "Fair division", Cambridge University Press, Cambridge.
- Conley, J.P. and S. Wilkie (1996) "An extension of the Nash Bargaining Solution to nonconvex problems", *Games and Economic Behaviour* 13: 26-38.

Elster, J. and J.E. Roemer (1991) "Interpersonal comparisons of well-being", Cambridge University Press, Cambridge.

Gauthier, D. (1986) "Morals by agreement", Oxford University Press, Oxford.

Gauthier, D. and R. Sugden (eds.) (1993) "Rationality, justice and the social contract", Harvester Wheatsheaf, Hemel Hempstead.

Hammond, P.J. (1991) "Interpersonal comparisons of utility: Why and how they are and should be made", ch. 7 of Elster and Roemer.

Herrero, M.J. (1989) "The Nash program: non-convex bargaining problems", *Journal of Economic Theory* 49: 266-77.

Kalai, E. (1977) "Proportional solutions to bargaining situations: interpersonal utility comparisons", *Econometrica* 45: 1623-30.

Kalai, E. and M. Smorodinsky (1975) "Other solutions to Nash's bargaining problem", *Econometrica* 43: 510-18.

Kaneko, M. (1980) "An extension of the Nash bargaining problem and the Nash social welfare function", *Theory and Decision* 12: 135-48.

Kolm, S-C. (1971) "Justice et equite", CEPREMAP, Paris.

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

Lensberg (1988) "Stability and the Nash solution", *Journal of Economic Theory* 45: 330-41.

Madden, P. (1996) "Suppes-Sen dominance, generalised Lorenz dominance and the welfare economics of competitive equilibrium: Some examples", *Journal of Public Economics* 61: 247-62.

Mariotti, M. (1997a) "Nash Bargaining Theory when the number of alternatives can be finite", forthcoming in *Social Choice and Welfare*.

Mariotti, M. (1997b) "Extending Nash's Axioms to Non-Convex Problems", forthcoming in *Games and Economic Behaviour*.

- Mas Colell, A., M.D. Whinston and J.R. Green (1995) "Microeconomic theory", Oxford University Press, New York and Oxford.
- Moulin, H. (1988) "Axioms of cooperative decision making", Cambridge University Press, Cambridge.
- Moulin, H. (1996) "Stand alone and Unanimity tests: A re-examination of fair division", in F. Farina, F. Hahn and S. Vannucci (eds.) "Ethics, rationality and economic behaviour", Clarendon Press, Oxford.
- Myerson, R. (1981) "Utilitarianism, egalitarianism and the timing effect in social choice problems", *Econometrica* 49: 883-97.
- Myerson, R. (1990) "Game theory. Analysis of conflict", Harvard University Press, Cambridge, Ma.
- Nash, J.F. (1950) "The bargaining problem", *Econometrica* 18: 155-62.
- Roemer, J.E. (1986) "The mismatch of bargaining theory and distributive justice", *Ethics* 97: 88-110.
- Roemer, J.E. (1990) "Welfarism and axiomatic bargaining theory", *Recherches Economiques de Louvain* 56: 287-301.
- Roemer, J.E. (1996) "Theories of distributive justice", Harvard University Press, Cambridge, Mass.
- Roth, A. (1979) "Proportional solutions to the bargaining problem", *Econometrica* 47: 775-78.
- Rubinstein, A., Z. Safra and W. Thomson (1992) "On the interpretation of the Nash Bargaining Solution and its extension to non-expected utility preferences", *Econometrica* 60: 1171-86.
- Saposnik, R. (1983) "On evaluating income distributions: rank-dominance, the Suppes-Sen Grading Principle of Justice and Pareto Optimality", *Public Choice* 40: 329-36.
- Sen, A. (1970) "Collective choice and social welfare", Holden-Day, san Francisco.
- Shorrocks, A. (1983) "Ranking income distributions", *Economica* 50: 3-18.
- Suppes, P. (1966) "Some formal models of grading principles", *Synthese* 6: 284-306.
- Thomson, W. (1990) "The consistency principle", in T. Ichiishi, A. Neyman and Y. Tauman (eds.), "Game theory and applications", Academic Press, san Diego.
- Weymark, J.A. (1991) "A reconsideration of the Harsanyi-Sen debate on utilitarianism", ch. 8 of Elster and Roemer (1991).
- Young, H.P. (1994) "Equity in theory and practice", Princeton University Press, Princeton, NJ.
- Zhou, L. (1996) "The Nash bargaining theory with non-convex problems", *Econometrica* 65: 681-86.