The Nash Bargaining Solution and Interpersonal Utility Comparisons

Shiran Rachmilevitch*

February 6, 2012

Abstract

Bargaining theory has a conceptual dichotomy at its core: according to one view, the utilities in the bargaining problem are meaningless numbers (v-N.M utilities), while according to another view they do have concrete meaning (willingness to pay). The former position is assumed by the Nash and Kalai-Smorodinsky solutions, and the latter is assumed by the egalitarian, utilitarian, and equal-loss solutions. In this paper I describe a certain form of equivalence between the *set* consisting of the former solutions and the set consisting of the latter. This equivalence is the result of an attempt to bridge the gap between the aforementioned views; utilizing this equivalence, I derive a new axiomatization of the Nash solution.

Keywords: Bargaining; interpersonal utility comparisons; Nash solution.

JEL Codes: D63; D71

*Department of Economics, University of Haifa, Mount Carmel, Haifa, 31905, Israel. Tel: +972-4-8249588; fax: +972-4-8240059. E-mail address: shiranrach@econ.haifa.ac.il

1 Introduction

The Nash bargaining problem (due to Nash (1950)) is defined as a pair (S, d). The set S, the *feasible set*, consists of all the utility vectors that the players can achieve via unanimous agreement, and $d \in S$, the *disagreement point*, specifies their utilities in case that no such agreement is reached—player i receives the utility payoff d_i in this event.

Regarding the points of S, the current literature offers two interpretations. According to one, utilities are v-N.M utilities: they are determined only up to positive affine transformations, and, in particular, there is no meaning to compare player i's utility payoff to that of player j's. This approach goes back to Nash's (1950) original work. The alternative approach assumes that utilities *are* comparable. This interpretation is especially appealing for situations in which bargaining is over the division of some resource—a "pie"—and a player's overall utility is $v(x) + t$, where x is his share of the pie and t is his wealth. Here, utilities stand for willingness to pay (for the pie), measured on a fixed, common scale. Kalai's (1977) work on the egalitarian solution is the literature's prominent representative of this approach. I will call the latter approach the *interpersonal approach* (IPA) and call the former the *interpersonal free approach* (IPFA).

Concepts related to fairness and efficiency are typically meaningful only under IPA. Think of the utilitarian criterion of maximizing the sum of the players' utilities and the egalitarian criterion of equating these utilities. What sense does it make to add "apples to oranges"? Likewise, what reason is there to equate meaningless numbers? It therefore seems that adopting IPFA, as is widely done in the literature,[1] makes it impossible, from the very onset,

[1] See Thomson (1994) for a comprehensive survey of the literature.

to consider basic ideas of fairness and efficiency, at least in their traditional formulation.

The point of this paper is to argue the opposite. I offer a middleground which combines IPA and IPFA. In this middleground, there is room for both *scale-invariance*, which, in IPFA, comes from the v-N.M assumption, and, at the same time, there is room for operations such as summing up and equating payoffs. Taking this approach, I derive a new characterization of the 2-person Nash bargaining solution.

My idea is this. In the existing literature, under either IPA or IPFA, the bargaining model assumes a fixed set of players. Instead, I offer the following alternative view: imagine that each problem $B = (S, d)$ corresponds to a possibly different set of players, $(1_B, 2_B, \cdots, n_B) \equiv I(B)$. When each problem corresponds to potentially different players, scale-invariance does not stand for the v.N-M assumption anymore—it is a requirement that regards comparisons of different sets of players. Under this interpretation of the model, interpersonal utility comparisons make sense. After all, if the problem B is unique to the players in $I(B)$, we may very well assume that, to begin with, it is given in the scales that capture the right interpersonal comparisons among them. It is therefore meaningful to consider the following solutions to $B = (S, d)$. The utilitarian solution, $U(S, d)$, the egalitarian solution, $E(S, d)$, and the equal-loss solution, $EL(S, d)$. The first is defined to be any selection from $\mathbb{U}(S, d) \equiv \text{argmax}_{S_d} \sum x_i$, where $S_d \equiv \{x \in S | x \geq d\}$,[2] the second is given by $d + \epsilon \cdot \mathbf{1}$,[3] where ϵ is the maximal number such that the aforementioned expression is in S, and the third picks the highest point $y \in S$ such that

[2]Vector inequalities: xRy if and only if $x_i R y_i$ for all i, for both $R \in \{\geq, >\}$.

[3]$\mathbf{1} = (1, \cdots, 1)$; similarly, $\mathbf{0} = (0, \cdots, 0)$.

$a_i - y_i = a_j - y_j$ for all i and j, where $a_i = a_i(S,d) \equiv \max\{x_i | x \in S_d\}$.[4]

The solution U represents the classical idea of utilitarianism, which dates back to Jeremy Bentham. In an environment with a common transferable utility unit (money), as in the "split the pie" problem above, maximization of the sum of utilities is equivalent to Pareto efficiency. The other solutions, E and EL, represent two alternative notions of fairness: equality of gain and equality of sacrifice. All of these solutions are sensible under IPA. However, since they typically yield different recommendations, they present bargaining theory with substantial difficulties: first, in solving bargaining situations, one needs to compromise on either fairness or efficiency; second, to begin with, it is not obvious what is the meaning of a "fair outcome," since there are (at least) two reasonable notions of fairness.

Underlying these difficulties is a more general problem. Suppose that the players consider $\{s^1, \cdots, s^K\}$ as legitimate candidate-solutions. If each solution has its merits, in terms of the axiomatizations it enjoys, the philosophical ideas it expresses, or otherwise, it is not clear which one they should chose. A conservative first step would be to demand that payoffs never fall below the minimum of the recommendations of these solutions; namely, that in each problem (S,d) each player i would receive at least $\min\{s_i^1(S,d), \cdots, s_i^K(S,d)\}$. This is a notion of insurance that guarantees that payoffs will never fall short of a certain bound—a bound which is itself a function of the appealing, but jointly-inconsistent, solutions $(s^k)_{k=1}^K$. Alternatively, one can think of the following noncooperative justification for this idea: the players are seating at the negotiations table, where the agreement $x \in S$ is up for consideration. If $x_i \geq \min\{s_i^1(S,d), \cdots, s_i^K(S,d)\}$ for each player i, then the candidate agree-

[4] $EL(S,d)$ is well-defined for $n = 2$ but may fail to exist when $n \geq 3$.

ment x is robust in the following sense: if some player j complains that x_j is too low, then the other players can reject his complain on the basis that "s^k is a legitimate solution, and under x you receive a payoff which is greater than the one you would have obtained under s^k."

Going back to our original concerns, this rationale leads me to consider the following requirement: to demand that for each bargaining problem (S, d) each player i will receive a payoff which is at least $\min\{U_i(S, d), E_i(S, d), EL_i(S, d)\}$.[5] I say that such a solution satisfies **restricted interpersonal comparisons**. For 2-person problems, a continuous scale-invariant solution exhibits this property if and only if the following is true: for every problem (S, d), each player i receives a payoff which is at least as large as $\min\{N_i(S, d), KS_i(S, d)\}$, where N is the Nash bargaining solution and KS is the Kalai-Smorodinsky solution (due to Kalai and Smorodinsky (1975)). Thus, this paper uncovers a nontrivial link between the major IPA-solutions and the major IPFA-solutions. Based on this link, I characterize the 2-person Nash solution.

The result does not extend to $n \geq 3$. Restricted interpersonal comparisons is not even well-defined in this case, because EL may not exist when $n \geq 3$.[6] Moreover, dropping EL from the axiom—i.e., demanding that each player i receives at least $\min\{U_i(S, d), E_i(S, d)\}$—results in an impossibility: no scale-invariant n-person solution satisfies this requirement for $n \geq 3$.

The rest of the paper is organized as follows. Section 2 lays down the model; in particular, it emphasizes and explains the interpretation that each

[5] To be more precise, the requirement is that there exists a selection U from $\mathbb{U}$ such that the above is satisfied. See subsection 2.1 for the formal definition.

[6] EL is guaranteed to exist if the feasible set is unbounded from below. In the current paper, however, I allow the feasible sets to be compact (i.e., I do not insist on free disposal of utilities), hence existence is an issue.

problem corresponds to a potentially different set of bargainers. Section 3 contains the results.

2 Model

An *n-person bargaining problem* (a problem, for short) is a pair (S, d) such that $S \subset \mathbb{R}^n$ is closed and convex, $d \in S$ is such that $S_d \equiv \{x \in S | x \geq d\}$ is bounded and contains a point x such that $x > d$, and S is d-comprehensive; that is, for all $x \in S$: $d \leq y \leq x \Rightarrow y \in S$. The collection of all these pairs (S, d) is denoted $\mathcal{B}_n$. A problem (S, d) is *smooth* if the Pareto frontier of its feasible set does not contain segments; that is, if for distinct $x, y \in P(S)$ and $\alpha \in (0, 1)$ the point $\alpha x + (1 - \alpha)y$ is not in $P(S)$, where $P(S) \equiv \{x \in S | y \geq x \& y \neq x \Rightarrow y \notin S\}$.

A *solution* is any function $\mu \colon \mathcal{B}_n \to \mathbb{R}^n$ that satisfies $\mu(S, d) \in S$ for all $(S, d) \in \mathcal{B}_n$. The *Nash solution* (due to Nash (1950)), N, is the unique maximizer of $\Pi_{i=1}^n (x_i - d_i)$ over $x \in S_d$. The *Kalai-Smorodinsky solution* (due to Kalai and Smorodinsky (1975)), KS, is given by $(1-\theta)d + \theta a(S, d)$, where θ is the maximal number such that the aforementioned expression is in S, where $a_i(S, d) \equiv \max\{x_i | x \in S_d\}$.[7] Let E and EL denote the egalitarian and equal-loss solutions, respectively, and let $\mathbb{U}$ denote the utilitarian correspondence (see Section 1 above for their definitions).[8]

The following axioms will be of interest in the sequel. In their definition, $(S, d) \in \mathcal{B}_n$ is an arbitrary problem and $\{(S_k, d)\}_{k=1}^{\infty} \subset \mathcal{B}_n$ is an arbitrary sequence of problems with a common disagreement point.

[7] The point $a(S, d)$ is called the *ideal point* of the problem (S, d).

[8] Note that if (S, d) is a smooth problem then $\mathbb{U}(S, d)$ is a singleton; in this case, we can unambiguously talk about *the* utilitarian solution to (S, d).

Disagreement Convexity (D.VEX): $\mu(S, \alpha d + (1-\alpha)\mu(S,d)) = \mu(S,d)$ for all $\alpha \in (0,1]$.

Midpoint Domination (MD): $\mu(S,d) \geq \frac{1}{n}\sum_{i=1}^{n}(a_i(S,d), d_{-i})$.

Scale Invariance (SINV): $\lambda \circ \mu(S,d) = \mu(\lambda \circ S, \lambda \circ d)$ for every positive affine transformation $\lambda \colon \mathbb{R}^n \to \mathbb{R}^n$.[9]

Continuity (CONT): If $\{S_k\}$ converges to S in the Hausdorff metric, then $\{\mu(S_k, d)\}$ converges to $\mu(S,d)$.

D.VEX says that a movement of the disagreement point in the direction of the agreement should not change the agreement and MD says that the agreement should dominate "randomized dictatorship." These ideas are well-known and throughly discussed in the literature.[10]

Now recall the varying-set-of-players interpretation: imagine that every problem $B = (S,d)$ corresponds to a potentially different set of players, $I(B)$. For every B and $I(B)$, the axioms D.VEX and MD assume their standard interpretations. SINV and CONT, however, are of a different nature, since they refer to multiple problems: the former refers to all the positive affine transformations of a given problem and the latter refers to a sequence of problems. Under the varying-set-of-players interpretation, therefore, they refer to problems played by (possibly) different bargainers.

To illustrate the meaning of SINV under this interpretation of the model,

[9] A function $\lambda \colon \mathbb{R}^n \to \mathbb{R}^n$ is a positive affine transformation if $\lambda \circ (x_1, \cdots, x_n) \equiv (\lambda_1 x_1, \cdots, \lambda_n x_n) + t$ for some numbers $\lambda_i > 0$ and $t \in \mathbb{R}$.

[10] See, e.g., Thomson (1994).

and since the bulk of the paper considers the 2-person case, let us look at the class of *2-person division problems*: those $(S,d) \in \mathcal{B}_2$ for which $d = (0,0)$ and $S = \{(y_1,y_2) \in \mathbb{R}^2_+ | y \leq (u_1(x), u_2(1-x)), x \in [0,1]\}$, where the u_i's are concave increasing utility functions with $u_i(0) = 0$, defined on the "pie" $[0,1]$. Assuming that a player's overall utility is quasi-linear, the functions u_i denote willingness to pay (for the pie). Given such a function u_i, one obtains an entire family $\{\lambda u_i | \lambda > 0\}$, parametrized by the *intensity* parameter λ. Then, when coupled with Pareto optimality,[11] SINV simply says that your share of the pie should be independent of the intensity of your opponent's willingness to pay for it. It therefore can be viewed as a principle of justice: you should not take an unfair advantage of your partner just because he happens to have low willingness to pay, and, similarly, he should not do it to you in the opposite case.

A related interpretation, along similar lines, goes as follows. Suppose that after bargaining has been completed, the players face an outside-of-bargaining risk of losing their shares of the pie: player i loses what he got with probability p_i. The combination of SINV and Pareto optimality then says that player i's solution payoff should be independent of p_j, which is an independence principle—it restricts the extent to which outside-of-bargaining circumstances can influence the bargaining outcome.

The interpretation of CONT is straightforward: if two groups of bargainers are "similar," then the agreements that they reach should also be.

[11]i.e., with the requirement (or axiom) that $\mu(S,d) \in P(S)$ for all (S,d).

2.1 Useful results from the existing literature.

The following results, which are stated without their proofs, will be useful in the sequel.

Proposition 1. *(Rachmilevitch 2011a) Let μ be a solution on $\mathcal{B}_2$ that satisfies SINV and CONT. Then, there exists a selection U from $\mathbb{U}$ such that $\mu_i(S,d) \geq min\{U_i(S,d), E_i(S,d)\}$ for all $(S,d) \in \mathcal{B}_2$ and $i \in \{1,2\}$ if and only if $\mu = N$.*

Proposition 2. *(Rachmilevitch 2011a) Let μ be a solution on $\mathcal{B}_n$ that satisfies SINV. Then, there exists a selection U from $\mathbb{U}$ such that $\mu_i(S,d) \geq min\{U_i(S,d), E_i(S,d)\}$ for all $(S,d) \in \mathcal{B}_n$ and $1 \leq i \leq n$ if and only if $n = 2$.*

Proposition 3. *(Rachmilevitch 2011b) Let μ be a solution on $\mathcal{B}_2$ that satisfies SINV. Then $\mu_i(S,d) \geq min\{E_i(S,d), EL_i(S,d)\}$ for all $(S,d) \in \mathcal{B}_2$ and $i \in \{1,2\}$ if and only if $\mu = KS$.*

Proposition 4. *(de Clippel 2007) N is the unique solution on $\mathcal{B}_2$ that satisfies D.VEX and MD.*[12]

3 The main result

Under interpersonal utility comparisons, the operations that were mentioned in Section 1—equating gains, equating loses, and adding utilities of different individuals—are all meaningful. When each problem $B = (S,d)$ is considered in "isolation," for the set of players $I(B)$, it seems mild to assume that this problem, to begin with, is defined in a way that allows for such comparisons.

[12] de Clippel's result is more general than that (it allows the solution to be multi-valued), but for our purpose the aforementioned version will suffice.

The fact that interpersonal utility comparisons give rise to different bargaining solutions, each of which has its merits, but which are jointly inconsistent, leads me to consider the following axiom. In its statement, as usual, (S, d) is an arbitrary problem.

Restricted Interpersonal Comparisons (RIC): There exists a selection $U(S, d) \in \mathbb{U}(S, d)$ such that

$$\mu_i(S, d) \geq \min\{U_i(S, d), E_i(S, d), EL_i(S, d)\} \ \forall i.$$

RIC can be viewed as a notion of insurance that incorporates utilitarianism, egalitarianism, and the equal-loss principle: a lower bound on each player's payoff is obtained by considering, jointly, the three major IPA solutions.

Proposition 5. *Let μ be solution on $\mathcal{B}_2$ that satisfies SINV and CONT. Then μ satisfies RIC if and only if $\mu_i(S, d) \geq min\{N_i(S, d), KS_i(S, d)\}$ for all $(S, d) \in \mathcal{B}_2$ and $i \in \{1, 2\}$.*

Proof. Make the assumptions of the proposition. Assume first that $\mu_i(S, d) \geq \min\{N_i(S, d), KS_i(S, d)\}$ for all $(S, d) \in \mathcal{B}_2$ and $i \in \{1, 2\}$. Then, in view of Proposition 1 and Proposition 3, μ satisfies RIC. Conversely, suppose that it satisfies RIC and let $(S, d) \in \mathcal{B}_2$. We need to show that for each player i the following holds: $\mu_i(S, d) \geq \min\{N_i(S, d), KS_i(S, d)\}$.

By CONT we may assume that (S, d) is smooth. Let λ be the positive affine transformation such that $E(T, d') = U(T, d') = N(T, d') \equiv x$, where $T \equiv \lambda \circ S$ and $d' \equiv \lambda \circ d$.[13] Wlog, suppose that $d = d' = \mathbf{0}$; let $y \equiv KS(T, \mathbf{0})$ and $a \equiv a(T, \mathbf{0})$.

[13]The existence of this positive affine transformation λ was first established by Harsanyi (1959). Shapley (1969) presents a related, and slightly more generally-formulated result.

Suppose first that $a_1 = a_2$. In this case $y = KS(T, \mathbf{0}) = EL(T, \mathbf{0}) = E(T, \mathbf{0}) = U(T, \mathbf{0}) = N(T, \mathbf{0}) = x$, and since μ satisfies RIC, $\mu(T, \mathbf{0}) \geq x$; since $x \in P(T)$, $\mu(T, \mathbf{0}) = x$. By SINV, $\mu(S, d) = \lambda^{-1} \circ x = N(S, d) = KS(S, d)$.

Suppose, on the other hand, that $a_1 \neq a_2$; wlog, suppose that $a_2 > a_1$. In this case, y is to the north west of x. Let $z \equiv \mu(T, \mathbf{0})$. By SINV, it is enough to prove that $z_i \geq \min\{N_i(T, \mathbf{0}), KS_i(T, \mathbf{0})\}$ for both $i \in \{1, 2\}$. That is, that $z_1 \geq y_1$ and that $z_2 \geq x_2$. The second inequality follows directly from RIC, because $EL(T, \mathbf{0})$ is to the north west of x (in fact, it is to the north west of y).[14] Assume by contradiction that $z_1 < y_1$. Let λ' be the positive affine transformation $(A, B) \mapsto (A, \frac{a_1}{a_2} B)$ and let $Q \equiv \lambda' \circ T$. By SINV, $\mu_1(Q, \mathbf{0}) = z_1 < \min\{U_1(Q, \mathbf{0}), E_1(Q, \mathbf{0}), EL_1(Q, \mathbf{0})\}$, in contradiction to RIC.[15] □

We can now turn to the main result.

Theorem 1. *N is the unique solution on $\mathcal{B}_2$ that satisfies RIC, D.VEX, SINV, and CONT.*

Proof. It is well-known that N satisfies D.VEX, SINV, and CONT, and it follows from Proposition 1 that it also satisfies RIC. Conversely, let μ be an arbitrary solution on $\mathcal{B}_2$ that satisfies the four axioms. By Proposition 5, it satisfies $\mu_i(S, d) \geq \min\{N_i(S, d), KS_i(S, d)\}$ for all $(S, d) \in \mathcal{B}_2$ and $i \in \{1, 2\}$. Therefore, since both N and KS satisfy MD, μ satisfies MD. By Proposition 4, the combination of MD and D.VEX implies $\mu = N$. □

[14]This argument implicitly relies on the smoothness of S. In its absence, we cannot conclude that $z_2 \geq x_2$, because the selection U in terms of which RIC is defined may select a point to the south east of x.

[15]Note that $\lambda' \circ y = EL(Q, \mathbf{0}) = E(Q, \mathbf{0})$; the fact that the utilitarian solution must weakly move to the right when λ' is applied to T completes the proof.

The axioms in Theorem 1 are independent. KS satisfies all of them besides D.VEX, E satisfies all of them besides SINV, and the disagreement solution $\mu(S,d) \equiv d$ satisfies all of them besides RIC. I now turn to describe a solution that satisfies all the axioms but CONT.

Let $S_x \equiv \text{conv hull}\{(0,0),(0,1),(x,1),(2x,0)\}$ and let $D_x \equiv \{\beta \times (0,0) + (1-\beta) \times (x,1) | 0 \leq \beta \leq 1\}$ be the diagonal that connects the origin to $(x,1)$. Define the following solution, μ^*, as follows. For (S,d) such that $S = S_x$ and d is to the left of D_x,[16] let $\mu^*(S,d) = (x,1)$. If (S,d) is such that there exists a positive affine transformation λ such that $(\lambda \circ S, \lambda \circ d)$ is a problem of the kind that was described in the previous sentence, let $\mu^*(S,d) = \lambda^{-1} \circ (x,1)$. For any other (S,d), set $\mu^*(S,d) = N(S,d)$. Clearly μ^* satisfies D.VEX and SINV. As for RIC, we only need to establish it for the case where $S = S_x$ for some $x > 0$ and d is to the left of D_x. If $x < 1$, then $\mu^*(S,d)$ is the unique utilitarian point in S_d. If, on the other hand, $x \geq 1$, then $\mu^*(S,d)$ is weakly to the right of $E(S,d)$.

For multi-person bargaining, i.e., for $n \geq 3$, RIC must be amended, because EL may fail to exist in this case. The following strengthening of the axiom then presents itself: there exists a selection $U(S,d) \in \mathbb{U}(S,d)$ such that $\mu_i(S,d) \geq \min\{U_i(S,d), E_i(S,d)\}$ for all problems (S,d) and all players i. As stated in Proposition 2, this strengthening results in an impossibility when $n \geq 3$. As we saw in Proposition 1, on the other hand, this stronger axiom, when combined with SINV, pins down the Nash solution for $n = 2$.

Acknowledgments I am grateful to Ehud Kalai, Alan Miller, and Anna Rubinchik for helpful comments.

[16] i.e., there is an $z \in D_x$ with $z_1 \geq d_1$.

4 References

[1] de Clippel, G., (2007), “An Axiomatization of the Nash Bargaining Solution,” *Social Choice and Welfare*, **29**, 201-210.

[2] Harsanyi, J,C., (1959), “A bargaining model for the cooperative n-person games.” In: Tucker A.W., Luce R.D. (eds) Contributions to the theory of Games IV. Princeton University Press, Princeton, 325-355.

[3] Kalai, E. (1977), “Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons,” *Econometrica*, **45**, 1623-1630.

[4] Nash, J.F, (1950),“The Bargaining Problem,” *Econometrica*, **18**, 155-162.

[5] Shapley, L.S., (1969), “Utility Comparison and the Theory of Games.” In: *La Décision: Agrégation et Dynamique des Ordres de Préférence,* Editions du CNRS, Paris, pp. 251-263.

[6] Rachmilevitch, S. (2011a), “Fairness, Efficiency, and the Nash Bargaining Solution,“ Working paper.

[7] Rachmilevitch, S. (2011b), “Fairness in Bargaining and the Kalai-Smorodinsky Solution,“ Working paper.

[8] Thomson, W. (1994), “Cooperative models of bargaining.” In: Aumann R.J., Hart S. (eds) Handbook of game theory, vol 2. North-Holland, Amsterdam, 1237-1284.