

Solidarity, distributive justice, and interpersonal utility comparisons

Shiran Rachmilevitch*

October 6, 2014

Abstract

Rawlsianism advocates maximizing the utility of the worst-off individual and utilitarianism advocates maximizing the overall utility-sum across individuals. I consider weaker variants of these principles in Nash's (1950) bargaining model: *utilitarian monotonicity* (resp. *Rawlsian monotonicity*) requires an expansion of the bargaining set not to lead to a decrease in the value of the utilitarian (resp. Rawlsian) objective. Within the class of scale invariant bargaining solutions, only "pathological" ones (e.g., dictatorship) satisfy any of these conditions. A bargaining solution satisfies *distributive justice monotonicity* if an expansion of the bargaining set does not lead to a weak decrease in the values of both the utilitarian and the Rawlsian objectives, with a strict decrease in (at least) one of them. The Nash solution is the unique scale invariant solution that satisfies this condition and the axiom *conflict-freeness*, which requires the ideal point (the utility pointwise maximum) to be selected whenever it is feasible.

Keywords: Distributive justice; Nash solution; Rawlsianism; Utilitarianism.

JEL Codes: C78; D74.

*Department of Economics, University of Haifa, Mount Carmel, Haifa, 31905, Israel. Email: shiranrach@econ.haifa.ac.il Web: <http://econ.haifa.ac.il/~shiranrach/>

1 Introduction

Rawlsianism and *utilitarianism* are fundamental principles of distributive justice, whose place in welfare economics and political philosophy is paramount.¹ The former school of thought advocates maximizing the utility of society’s worst-off individual, the latter advocates maximizing the overall utility-sum across individuals.² Both rely on *interpersonal comparisons of utility*. Indeed, if utilities were not comparable, then their summation, for instance, would bear no ethical significance. Such comparisons, however, have been deemed problematic by mainstream economics ever since the “ordinalist revolution” in the 1930’s (see Mandler (1999) for a survey). In light of this state of affairs, my goal in the present paper is twofold: (a) to explore the tension between the utilitarian and Rawlsian standpoints on the one hand and the rejection of interpersonal comparisons on the other hand, and (b) to try to resolve this tension. I will carry out the study using the simplest social choice model that lends itself to this kind of analysis—a simplified version of Nash’s (1950) bargaining problem. As we shall see, the attempt to resolve the aforementioned tension will lead us to the Nash bargaining solution.

A *Nash bargaining problem* is a compact and convex set $S \subset \mathbb{R}_+^2$, which is non-trivial and comprehensive. Not-triviality means that there is some $x \in S$ with $x > \mathbf{0} \equiv (0, 0)$ ³, and comprehensiveness means that $\{y \in \mathbb{R}_+^2 : y \leq x\} \subset S$, for all $x \in S$. The interpretation is that two players face a set of feasible v-N-M utility-agreements, S , and they will both get zero utility unless they unanimously agree on some $x \in S$, in which case each player i receives the utility payoff x_i . Under this

¹See Fleurbaey, Salles, and Weymark (2008).

²I use “Rawlsianism” for what is commonly referred to as “egalitarianism,” since the latter term highlights the drive for equality, whereas I intend to focus on the advancement of the well-being of the worst-off. It seems fitting to use “Rawlsianism,” after this view’s leading representative over the past half century (Rawls (1971)).

³Vector inequalities: uRv means that $u_i R v_i$ for each i for both $R \in \{>, \geq\}$, and $u \not\geq v$ iff $[u \geq v$ and $u \neq v]$.

interpretation, non-triviality means that there are strict incentives to avoid disagreement and comprehensiveness means that utilities can be freely disposed (up to the disagreement level). Let \mathcal{S} denote the collection of all problems. A *bargaining solution* is a function $\sigma: \mathcal{S} \rightarrow \mathbb{R}_+^2$ such that $\sigma(S) \in S$ for all $S \in \mathcal{S}$.

If the utility numbers are arbitrary numerical representations of the players' v-N-M preferences, then they are devoid of any objective significance, and, consequently, utilities of different players cannot (and should not) be compared. Mathematically, this incomparability is expressed by *scale invariance*. A solution σ has this property if for every problem S and any pair of positive linear transformations $l = (l_1, l_2)$ (namely, $l_i(a) \equiv t_i a$ for some constant $t_i > 0$) it is true that $\sigma(l \circ S) = l \circ \sigma(S)$.

Scale invariance is satisfied by the main pillar of bargaining theory—the aforementioned *Nash solution* (Nash (1950)), which assigns to each S the point $x \in S$ that maximizes the utility-product $x_1 \cdot x_2$. By contrast, scale invariance is incompatible with both Rawlsianism and utilitarianism. More formally, utilitarian and Rawlsian bargaining solutions are defined as follows: the utilitarian objective is $U(x) \equiv x_1 + x_2$, the Rawlsian objective is $R(x) \equiv \min\{x_1, x_2\}$, and a utilitarian (Rawlsian) solution maximizes U (R) over S for every $S \in \mathcal{S}$. There are many utilitarian (Rawlsian) solutions, but they only differ in how they make their selection from the bargaining problem in case that the utility-sum (utility-minimum) is maximized at multiple points. Since any Rawlsian or utilitarian solution is not scale invariant, it seems that there is a profound tension between interpersonal incomparability of the individual utilities and the basic building blocks of distributive justice. *Can we express meaningful distributive justice restrictions in the bargaining model while maintaining interpersonal incomparability of the individual utilities?*

Before we embark on the task that this question lays in front of us, it is legitimate to ask whether we should at all take utilitarianism and Rawlsianism as our guideline principles for distributive justice. Why be guided by notions that are deeply rooted in interpersonal comparisons when utilities are assumed to be interpersonally

incomparable?

The answer is that we could (and perhaps even should) take utilitarianism and Rawlsianism as relevant principles to our ethical reasoning, if we took utilities to be comparable *in principle*, even if not *in practice*. For example, utilities are comparable in principle but not in practice when there is a fixed “rate of exchange” between them (namely, utility is *linearly comparable*), but this rate of exchange is unknown.⁴ Under this view both utilitarianism and Rawlsianism make sense, but they simply cannot be implemented.⁵ Appreciating the merits of utilitarianism and Rawlsianism, while not committing to comparing utilities in the same way that a utilitarian or a Rawlsian would, is therefore a meaningful and coherent view.

I therefore set out to look for weaker versions of utilitarianism and Rawlsianism. These versions are required to be sufficiently weak so as to allow for scale invariance, but, at the same time, they are also required to express the utilitarian/Rawlsian spirit. This motivates the following monotonicity conditions. A solution σ is *utilitarian monotonic* if for any two nested problems, $S \subset T$, it is true that $U(\sigma(S)) \leq U(\sigma(T))$. Similarly, it is *Rawlsian monotonic* if $S \subset T$ implies that $R(\sigma(S)) \leq R(\sigma(T))$. Both monotonicity conditions link the utilitarian/Rawlsian agendas with the principle of *solidarity*. To see how, consider first Rawlsian monotonicity. Suppose that the problem S expands to T . Then, since everybody can benefit from the expansion of options, solidarity requires that everybody indeed benefits, at least weakly. Rawlsian monotonicity is a weakening of the requirement that the only reason to compromise on solidarity is to increase the value of the Rawlsian objective. Namely, it is a weakening of the following condition:

⁴Utility may be comparable but not linearly comparable; namely, z utils of player i equal $\phi(z)$ units of player j , where ϕ is some increasing but non-linear function. Knowing ϕ only up to a parameter is the general version of not knowing the correct rate of exchange in the case of linear transferability.

⁵For more on the idea that utility may be comparable in principle but not in practice, see Hammond (1991) and Mariotti (1999), as well as the introduction in Elster and Roemer (1991).

$$(\star)_R \quad \forall S, T \in \mathcal{S} : S \subset T \ \& \ \neg[\sigma(S) \leq \sigma(T)] \Rightarrow R(\sigma(T)) > R(\sigma(S)).$$

Under some mild conditions, Rawlsian monotonicity implies $(\star)_R$. As for utilitarian monotonicity, similar results hold with respect to the condition $(\star)_U$:

$$(\star)_U \quad \forall S, T \in \mathcal{S} : S \subset T \ \& \ \neg[\sigma(S) \leq \sigma(T)] \Rightarrow U(\sigma(T)) > U(\sigma(S)).$$

Unfortunately, utilitarian and Rawlsian monotonicity turn out to be rather strong, and they fail to bridge the gap between the incomparability of the individual utilities and the utilitarian/Rawlsian principles. More specifically, any well-behaved scale invariant solution violates both monotonicity conditions.⁶

In light of this state of affairs, I consider an alternative condition, *distributive justice monotonicity*. It requires that when a problem S expands to T , the following holds: $\neg[(U(\sigma(T)), R(\sigma(T))) \preceq (U(\sigma(S)), R(\sigma(S)))]$. That is, an expansion of the feasible set of utility allocations cannot result in a weak decrease in the values of both the utilitarian and Rawlsian objectives, with a strict decrease in (at least) one of them. Distributive justice monotonicity is a weakening of the combination of utilitarian and Rawlsian monotonicity.

The positive result of this paper is that the Nash solution is the unique scale invariant solution that satisfies distributive justice monotonicity and *conflict freeness*—the requirement that the solution assigns to each S the point $a(S)$ whenever it is feasible to do so, where $a(S)$, the *ideal point* of S , is defined by $a_i(S) \equiv \max\{s_i : s \in S\}$.

The rest of the paper is organized as follows. Section 2 focuses on the impossibility of a well-behaved, scale invariant, utilitarian monotonic solution. Section 3 does the same for Rawlsian monotonicity. Section 4 discusses solidarity. Section 5 is dedicated to the characterization of the Nash solution on the basis of distributive justice monotonicity, and Section 6 briefly concludes.

⁶The precise meaning of “well-behaved” will be clarified shortly, in Sections 2 and 3.

2 Utilitarian monotonicity: Impossibility I

Unlike utilitarianism, utilitarian monotonicity *is* compatible with scale invariance. Here are two examples of utilitarian monotonic and scale invariant solutions. The first is the *i-th dictatorial solution*, which assigns to each problem the point Me^i , where M is the maximum possible and e^i is such that $e_i^i = 1$ and $e_j^i = 0$. The second, the *disagreement solution*, assigns to each problem the origin, $\mathbf{0}$. Obviously, none of these solutions is appealing. Unfortunately, this is no coincidence: below I show that utilitarian monotonicity is incompatible with scale invariance, provided that other, fairly mild and standard requirements are imposed on the solution.

The obvious drawback of the dictatorial solution is that it violates *symmetry*. A solution satisfies *symmetry* if for every $S \in \mathcal{S}$ that is symmetric with respect to the 45° line, the point selected by the solution belongs to this line. The obvious drawback of the disagreement solution is that it is not *weakly Pareto optimal*. A solution satisfies *weak Pareto optimality* if for every $S \in \mathcal{S}$ the solution makes its selection out of $WP(S) \equiv \{x \in S : y > x \Rightarrow y \notin S\}$. The following result shows why we had to resort to the aforementioned pathological solutions in order to show that utilitarian monotonicity and scale invariance can coexist.

Theorem 1. *There does not exist a utilitarian monotonic solution that satisfies scale invariance, symmetry, and weak Pareto optimality.*

To prove Theorem 1, the following lemma will be useful. For the lemma, the following axiom is needed. A solution σ satisfies *conflict-freeness* if $\sigma(S) = a(S)$ whenever $a(S) \in S$.

Lemma 1. *If a solution satisfies weak Pareto optimality, symmetry, and scale invariance, then it satisfies conflict-freeness.*

Proof. Let σ be a solution as above, and assume by contradiction that there is some (a, b) such that $x \equiv \sigma(S) \neq (a, b)$, where $S = \{y \in \mathbb{R}_+^2 : y \leq (a, b)\}$. Then, wlog,

by weak Pareto, $y = (a, c)$ for some $c < b$. By rescaling 1's utility by b/a we get that $\sigma(V) = (b, c)$, where V is the image of S under this rescaling; in particular, V is symmetric. However, (b, c) is not on the 45° line, in contradiction to symmetry. \square

I will also make use of the following fact (which is easy to verify): if a solution σ satisfies scale invariance, symmetry, and weak Pareto optimality, and $T \in \mathcal{S}$ is a triangle, then $\sigma(T)$ is the midpoint of T 's hypotenuse. Call this property *midpoint outcome on a linear frontier*, or MOL for short (Anbarci (1998)).

Proof of Theorem 1: Assume by contradiction that σ is a solution with all these properties. Let $V \equiv \{x \in \mathbb{R}_+^2 : x \leq (2, 1)\}$. Let l_α be the line that goes through $(2, 1)$ and has slope α , where $\alpha \in (-1, -\frac{1}{2})$. Let $l_\alpha^* \equiv l_\alpha \cap \mathbb{R}_+^2$ and let $T_\alpha \in \mathcal{S}$ be the triangular problem such that $WP(T_\alpha) = l_\alpha^*$. By Lemma 1, $\sigma(V) = (2, 1)$ and therefore, by utilitarian monotonicity, $\alpha > -1$ implies that $\sigma(T_\alpha)$ is weakly to the right of $(2, 1)$. However, since $\alpha < -\frac{1}{2}$, MOL implies that $\sigma(T_\alpha)$ is strictly to the left of $(2, 1)$ —a contradiction. \square

The axioms in Theorem 1 are independent. The Nash solution satisfies all of them except utilitarian monotonicity. The disagreement and dictatorial solutions satisfy all of them except weak Pareto optimality and symmetry, respectively. Any symmetry-respecting utilitarian solution satisfies all of them except scale invariance.

As seen in Theorem 1's proof, the only role of the axioms different from utilitarian monotonicity is to guarantee conflict-freeness and MOL. The following corollary is therefore an immediate consequence of Theorem 1.

Corollary 1. *There does not exist a utilitarian monotonic solution that satisfies conflict-freeness and midpoint outcome on a linear frontier.*

In this corollary, too, the axioms are independent. The Nash solution satisfies all of them except utilitarian monotonicity. The *midpoint solution*—the one that assigns

to each $S \in \mathcal{S}$ the point $\frac{1}{2}a(S)$, satisfies all of them except conflict-freeness. Finally, a utilitarian solution satisfies all of them except MOL.

Both conflict-freeness and MOL are particular instances of the following more general axiom (due to Anbarci (1998)). A solution σ satisfies the *balanced focal point property* if $\sigma(S) = (a, b)$, provided that $S = \text{conv}\{\mathbf{0}, (a, b), (\lambda a, 0), (0, \lambda b)\}$ for some $\lambda \in [1, 2]$. Note that conflict-freeness obtains for $\lambda = 1$ and MOL obtains for $\lambda = 2$. The rationale behind this property (for a general $\lambda \in [1, 2]$) is that (a, b) represents a point of equal concessions, in the sense that the areas to its north-west and south-east are the same. Every solution that satisfies weak Pareto optimality, symmetry, and scale invariance satisfies the balanced focal point property. The following is an immediate consequence of Corollary 1.

Corollary 2. *There does not exist a utilitarian monotonic solution that satisfies the balanced focal point property.*

3 Rawlsian monotonicity: Impossibility II

Here is a “Rawlsian counterpart” of Theorem 1.

Theorem 2. *There does not exist a Rawlsian monotonic solution that satisfies scale invariance, symmetry, and weak Pareto optimality.*

Proof. Assume by contradiction that σ is a solution with all these properties. Let $S \equiv \text{conv}\{\mathbf{0}, (1, 0), (0, b)\}$, where $b > 1$. By MOL, $R(\sigma(S)) = \frac{1}{2}$. Any point on S 's frontier is of the form $(t, -bt + b)$, for some $t \in [0, 1]$. Pick a $t^* \in (\frac{b}{1+b}, 1 - \frac{1}{2b})$ and look at $V \equiv \{x \in \mathbb{R}_+^2 : x \leq (t^*, -bt^* + b)\}$. By Lemma 1, $\sigma(V) = (t^*, -bt^* + b)$ and by the choice of t^* , $R(\sigma(V)) = -bt^* + b = b(1 - t^*)$. Since $V \subset S$, Rawlsian monotonicity implies $b(1 - t^*) \leq \frac{1}{2}$ and therefore $1 - \frac{1}{2b} \leq t^*$ —a contradiction. \square

It is easy to verify the independence of the axioms in Theorem 2. In the proof of Theorem 2, like in that of Theorem 1, the only role of the axioms different from

Rawlsian monotonicity is to guarantee conflict-freeness and MOL. Therefore, the following result is immediate.

Corollary 3. *There does not exist a Rawlsian monotonic solution that satisfies the balanced focal point property.*

Remark 1: Utilitarian and Rawlsian monotonicity are independent. That utilitarian monotonicity does not imply Rawlsian monotonicity is easy to see: any utilitarian solution satisfies the former but not the latter. That Rawlsian monotonicity does not imply utilitarian monotonicity is seen in the following example, σ^* :

$$\sigma^*(S) \equiv \begin{cases} \mathbf{0} & \text{if } S \cap \{(2, 1), (1.4, 1.4)\} = \emptyset \\ (2, 1) & \text{if } S \cap \{(2, 1), (1.4, 1.4)\} = \{(2, 1)\} \\ (1.4, 1.4) & \text{otherwise} \end{cases}$$

It is easy to check that σ^* is Rawlsian monotonic but not utilitarian monotonic.

Remark 2: Both Rawlsian and utilitarian monotonicity are implied by *strong monotonicity*, which requires $\sigma(S) \leq \sigma(T)$ whenever $S \subset T$. The converse, however, does not hold: utilitarian/Rawlsian monotonicity does not imply strong monotonicity. This is seen in the following example, σ^{**} :

$$\sigma^{**}(S) \equiv \begin{cases} (a_1(S), 0) & \text{if } a_1(S) \geq a_2(S) \\ (0, a_2(S)) & \text{otherwise} \end{cases}$$

It is easy to check that σ^{**} is both utilitarian and Rawlsian monotonic, but not strongly monotonic.

4 Solidarity

Strong monotonicity is an axiomatic expression of the solidarity-principle “noone gets hurt when the set of feasible utility allocations expands.” As mentioned in the

Introduction, Rawlsian monotonicity is implied by the requirement that the only reason to compromise on strong monotonicity is to increase the value of the Rawlsian objective, a requirement abbreviated $(\star)_R$. To see this, consider a $(\star)_R$ -respecting solution, σ , and assume by contradiction that it violates Rawlsian monotonicity. That is, there are $S \subset T$ such that $R(\sigma(T)) < R(\sigma(S))$. Note that this strict inequality implies $\neg[\sigma(S) \leq \sigma(T)]$, which, in turn, leads to a contradiction, since the combination of $S \subset T$, $\neg[\sigma(S) \leq \sigma(T)]$, and $(\star)_R$ implies $R(\sigma(T)) > R(\sigma(S))$. Conversely, under some mild conditions Rawlsian monotonicity implies $(\star)_R$. These conditions are the following.

A solution σ is *continuous* if for every sequence in \mathcal{S} , $\{S_n\}$, that converges in the Hausdorff topology to some $S \in \mathcal{S}$, it is true that $\sigma(S_n) \rightarrow \sigma(S)$. A solution σ satisfies *destruction proofness* if $S \subset T$ implies $\neg[\sigma(T) \not\leq \sigma(S)]$. If a solution does not satisfy destruction proofness, then there are circumstances given which it is worthwhile for the players to cooperate in order to undermine the feasibility of some agreements.

Theorem 3. *Let σ be a continuous solution that satisfies Rawlsian monotonicity and destruction proofness. Then σ satisfies $(\star)_R$.*

Proof. Let σ be a solution as above, and assume by contradiction that there are S and T such that $S \subset T$, $\neg[\sigma(S) \leq \sigma(T)]$, but, nevertheless, $R(\sigma(T)) \leq R(\sigma(S))$. Let $x \equiv \sigma(S)$ and $y \equiv \sigma(T)$. By Rawlsian monotonicity, $R(x) = R(y)$.

Note that x is not on the 45° line; otherwise, Rawlsian monotonicity would imply $x \leq y$, in contradiction to $\neg[x \leq y]$. Wlog, suppose that $x_2 > x_1$. This implies that $y_1 > y_2$ (note that (i) $y_1 = y_2$ is ruled out by destruction proofness, and (ii) the combination of $y_2 > y_1$ and $R(y) = R(x)$ implies $y \geq x$, which is impossible).

For each $\lambda \in [0, 1]$ let $V_\lambda \equiv (1-\lambda)S + \lambda T$, and let $v_\lambda \equiv \sigma(V_\lambda)$. Let d_i be the distance of i from the 45° line for each $i \in \{x, y\}$. Let $\delta \equiv \frac{\min\{d_x, d_y\}}{2}$. Given $c \in \mathbb{R}^2$ and $r > 0$, let $B(c)_r$ denote the closed ball with center c and radius r . Let $A \equiv \{\lambda : v_\lambda \in B_\delta(x)\}$ and $B \equiv \{\lambda : v_\lambda \in B_\delta(y)\}$. By continuity, there is a $\bar{\lambda} < 1$ such that $\lambda \in B$ for all $\lambda \in [\bar{\lambda}, 1]$. Also, $0 \in A$. Let $\lambda^+ \equiv \sup A$ and $\lambda^- \equiv \inf B$. By continuity and the

non-emptiness of A and B , $\lambda^+ \in A$ and $\lambda^- \in B$. Since $A \cap B = \emptyset$, $\lambda^+ < \lambda^-$. Let $\lambda^* \in (\lambda^+, \lambda^-)$. By Rawlsian monotonicity $R(v_{\lambda^*}) \geq R(x)$, therefore $R(v_{\lambda^*}) > R(x)$. Therefore $R(v_{\lambda^*}) > R(y)$ —in contradiction to Rawlsian monotonicity. \square

Utilitarian monotonicity is a weakening of $(\star)_U$. With some extra assumptions one obtains a “utilitarian analog” of Theorem 3 (sufficient conditions are (1) to strengthen destruction proofness to weak Pareto optimality, and (2) to restrict attention to problems on which the utility-sum is maximized at a single point). For brevity, I omit the details.

5 Distributive justice monotonicity: A characterization of the Nash solution

The following condition is a weakening of the combination of utilitarian and Rawlsian monotonicity. A solution satisfies *distributive justice monotonicity* if $S \subset T$ implies $\neg[(U(\sigma(T)), R(\sigma(T))) \preceq (U(\sigma(S)), R(\sigma(S)))]$. In words, distributive justice monotonicity requires an expansion of options not to lead to a weak decrease in the values of both the utilitarian and Rawlsian objectives, with a strict decrease in (at least) one of them.

Lemma 2. *The Nash solution satisfies distributive justice monotonicity.*

Proof. Let $S, T \in \mathcal{S}$ be such that $S \subset T$. Let $(a, b) \equiv N(S)$ and $(x, y) \equiv N(T)$. Assume by contradiction that $(U((a, b)), R((a, b))) \preceq (U((x, y)), R((x, y)))$. In particular,

$$a + b \geq x + y.$$

Since $(a, b) \neq (x, y)$ and the maximizer of the Nash product is unique, $xy > ab$, hence $\sqrt{x} \cdot \sqrt{y} > \sqrt{a} \cdot \sqrt{b}$. Therefore $a + b - 2\sqrt{a} \cdot \sqrt{b} > x + y - 2\sqrt{x} \cdot \sqrt{y}$, which implies $(\sqrt{a} - \sqrt{b})^2 > (\sqrt{x} - \sqrt{y})^2$, hence $|\sqrt{a} - \sqrt{b}| > |\sqrt{x} - \sqrt{y}|$. Wlog, suppose that $a \geq b$ and $x \geq y$. Therefore,

$$\sqrt{a} - \sqrt{b} > \sqrt{x} - \sqrt{y}. \quad (1)$$

Now, since $a \geq b$ and $x \geq y$ we have that $R((a, b)) = b$ and $R((x, y)) = y$, and per our assumption $y \leq b$ and so $\sqrt{y} \leq \sqrt{b}$. Combining this inequality with (1), we obtain $\sqrt{a} - \sqrt{b} > \sqrt{x} - \sqrt{b}$, and so $a > x$. We conclude that $N(S) = (a, b) \not\geq (x, y) = N(T)$, which is obviously impossible. \square

Theorem 4. *A solution satisfies conflict-freeness, scale invariance, and distributive justice monotonicity if and only if it is the Nash solution.*

Proof. Since N satisfies scale invariance and conflict-freeness, it is enough, by Lemma 2, to prove uniqueness. Let then σ be a solution with all the above properties. Let $S \in \mathcal{S}$. By scale invariance we may assume that $N(S) = E(S) = (1, 1)$. Let $V \equiv \{x \in \mathbb{R}_+^2 : x \leq (1, 1)\}$. By conflict-freeness, $\sigma(V) = (1, 1)$. Since S is weakly below the hyperplane $\{(x, y) : x + y = 2\}$ and $V \subset S$, distributive justice monotonicity implies $\sigma(S) = (1, 1) = N(S)$. \square

The axioms in Theorem 4 are independent. The disagreement solution satisfies all of them besides conflict-freeness. The Kalai-Smorodinsky solution—the one that assigns to each S the point $\theta a(S)$, where θ is the maximum possible—satisfies all of them besides distributive justice monotonicity. A solution that assigns for every S a maximizer of $\frac{1}{2}U(x) + \frac{1}{2}R(x)$ over $x \in S$ satisfies all the axioms besides scale invariance ($\frac{1}{2}U + \frac{1}{2}R$ can be maximized at multiple points, because its level curves are linear on both sides of the 45° line).

Theorem 4 provides a distributive-justice-foundation for the Nash solution. In this sense, the result which is closest to it in the existing literature is by Mariotti (1999), who characterized the Nash solution on the basis of scale invariance and *Suppes-Sen Dominance*, or SSD (Suppes (1966), Sen (1970)). SSD requires that if x is the selected agreement and (a, b) is some feasible point, then both $(a, b) > x$ and $(b, a) > x$ are

false.⁷ SSD stems from the combination of the Pareto principle and impartiality, since it boils down to an application of weak Pareto optimality after all information about the players’ identities has been erased. It follows from Mariotti’s result and Theorem 4 that under the restriction to scale invariant solutions, the combination “conflict-freeness+distributive justice monotonicity” is equivalent to SSD (as they both imply N). Without scale invariance the equivalence disappears, but there is still a general logical relation between the properties.

Proposition 1. *Every solution that satisfies conflict-freeness and distributive justice monotonicity also satisfies SSD.*

Proof. Let σ satisfy conflict-freeness and distributive justice monotonicity, let S be a problem, and let $x \equiv \sigma(S)$. Let $(a, b) \in S$. We will see that both $(a, b) > x$ and $(b, a) > x$ are false.

Suppose first that $(a, b) > x$. Let $V \equiv \{s \in S : s \leq (a, b)\}$. By conflict-freeness $\sigma(V) = (a, b)$; therefore, $\sigma(S) = x$ constitutes a violation of distributive justice monotonicity, a contradiction.

Now consider the possibility $(b, a) > x$. Let $V' \equiv \{s \in S : s \leq (b, a)\}$. By conflict-freeness $\sigma(V') = (b, a)$ and therefore, by distributive justice monotonicity, $\min\{x_1, x_2\} > \min\{a, b\}$. Wlog, suppose that $\min\{x_1, x_2\} = x_1$.

If $a \leq b$ then $x_1 > a > x_2 \geq x_1$ —a contradiction. If $b \leq a$ then $x_1 > b$, which is also impossible. □

The “converse” of Proposition 1 does not hold. In fact, there exist solutions that satisfy SSD but violate both conflict-freeness and distributive justice monotonicity. Here is an example of such a solution. Let $\tilde{S} \equiv \{s \in \mathbb{R}_+^2 : s \leq (1, 1)\}$ and consider:

$$\sigma^{***}(S) \equiv \begin{cases} (1, \frac{1}{2}) & \text{if } S = \tilde{S} \\ N(S) & \text{otherwise} \end{cases}$$

⁷This condition also appears, under different names, in Kolm (1971) and Blackorby and Donaldson (1977).

Clearly σ^{***} violates conflict-freeness (on \tilde{S}) and it is easy to see that it satisfies SSD: the only problem on which it could possibly violate SSD is \tilde{S} (otherwise it coincides with N), but there does not exist any $s \in \tilde{S}$ such that $s > (1, \frac{1}{2})$ or $(s_2, s_1) > (1, \frac{1}{2})$. To see that it violates distributive justice monotonicity, note that $\sigma(V) = (\frac{3}{4}, \frac{3}{4})$, where $V \equiv \{s \in \tilde{S} : s \leq (\frac{3}{4}, \frac{3}{4})\}$, so in the move from V to \tilde{S} the value of the utilitarian objective remains the same and that of the Rawlsian objective decreases.

Another result from the literature that is related to Theorem 4 is by Shapley (1969), who showed that the Nash solution is the only one that jointly satisfies the utilitarian and Rawlsian objectives for some rescaling of the individual utilities. However, the philosophy underlying Shapley’s work is significantly different from the one underlying Theorem 4: Shapley’s reasoning relies heavily on the interpersonal comparability and interpersonal transferability of the utilities, whereas the line of thinking that culminates in Theorem 4 deliberately tries to avoid these assumptions.⁸

6 Conclusion

The main positive result of this paper is a characterization of the Nash solution that is based on a weak ethical condition, distributive justice monotonicity. Distributive justice monotonicity implies a weak form of substitution between utilitarianism and Rawlsianism: since this axiom rules out certain joint compromises on utilitarianism and Rawlsianism, it can be interpreted as saying that we “need to have at least one of them.” It is interesting to note that alongside this weak substitutability, there is (sometimes) a form of *complementarity* between the two principles, in the sense that there are cases in which common-sense reasoning has both utilitarian and Rawlsian aspects to it, and they both support the same course of action. A classic example can be found in the parable about the Good Samaritan: when two strangers meet by the side of the road and one of them is in dire need of help, it is natural to justify the

⁸For a detailed discussion of Shapley’s work, see Yaari (1981).

other's duty to help on the basis of a *utilitarian* argument (Walzer (1983), p.33)—if the risks and costs of giving help are low relative to the benefits it provides to the needy party, help must be provided; what we have here is a utilitarian argument that is being invoked because of a *Rawlsian* reason—the call to help the worst-off individual.⁹

References

- [1] Anbarci, N. (1998), Simple characterizations of the Nash and Kalai/Smorodinsky solutions, *Theory and Decision*, **45**, 255-261.
- [2] Blackorby, C., and Donaldson, D. (1977), Utility vs. equity: some plausible quasi-orderings, *Journal of Public Economics*, **7**, 365-381.
- [3] Elster, J., and Roemer, J.E. (Eds.), *Interpersonal Comparisons of Well-Being*, Cambridge University Press (1991).
- [4] Fleurbaey, M., Salles, M., and Weymark, J.A. (eds.) *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, Cambridge University Press, (2008).
- [5] Hammond, P. J. (1991), Interpersonal comparisons of utility: why and how they are and should be made, in J. Elster and J. E. Roemer (eds.), *Interpersonal Comparisons of Well-Being*, Cambridge University Press.
- [6] Kalai, E. (1977), Proportional solutions to bargaining situations: interpersonal utility comparisons, *Econometrica*, **45**, 1623-1630.

⁹To be precise, the Good Samaritan story cannot be modeled as a bargaining problem, since its starting point—the status quo, or disagreement point—is on the boundary of the feasible set. The only movement is along that boundary, where utility is transferred from the able to the needy. A basic feature of our model, by contrast, is that the disagreement point is set to the origin, and the surplus to be shared can benefit both players simultaneously. Despite not being formally modelable as a bargaining problem, I believe that the Good Samaritan tale is an effective illustration of a case where utilitarian and Rawlsian considerations have a common practical implication.

- [7] Kalai, E. and Smorodinsky, M. (1975), Other solutions to Nash's bargaining problem, *Econometrica*, **43**, 513-518.
- [8] Kolm, S-C., Justice et Equite, CEPREMAP, Paris (1971).
- [9] Mandler, M. (1999), Dilemmas in Economic Theory: Persisting Foundational Problems of Microeconomics, Oxford University Press.
- [10] Mariotti, M. (1999), Fair bargains: distributive justice and Nash bargaining theory, *Review of Economic Studies*, *66*, 733-741.
- [11] Nash, J.F. (1950), The bargaining problem, *Econometrica*, **18**, 155-162.
- [12] Rawls, J. A Theory of Justice, Belknap (1971).
- [13] Sen, A. (1970), Collective Choice and Social Welfare, Holden-Day.
- [14] 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, 251-263.
- [15] Suppes, P. (1966), Some formal models of grading principles, *Synthese*, **6**, 284-306.
- [16] Walzer, M. Spheres of Justice, Basic Books (1983).
- [17] Yaari, M.E. (1981), Rawls, Edgeworth, Shapley, Nash: theories of distributive justice re-examined, *Journal of Economic Theory*, **24**, 1-39.