1996

Let's Agree that All Dictatorships are Equally Bad

Uzi Segal

ISSN:0318-725X
ISBN:0-7714-1926-0

RESEARCH REPORT 9608

Let's Agree that All Dictatorships are Equally Bad

by

Uzi Segal

October 1996

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@sscl.uwo.ca

Let's Agree that All Dictatorships are Equally Bad*

Uzi Segal[†]

October 1, 1996

Abstract

A social policy is a rule which assigns each possible set of endowments an allocation of these endowments among members of society. This paper assumes that individuals have preferences over private consumption and preferences over all possible social policies. I offer a set of axioms which imply that the best social policy is to maximize a weighted sum of individual utility levels. The weight of an individual given a certain bundle of resources is the inverse of the maximal utility gain this person may enjoy from this bundle. The key axiom is that all individuals agree that giving all the resources of the economy always to the same person is bad, regardless of who that person is. Members of society may have different preferences over social policies, but they all agree that the above social policy is best.

*I gratefully acknowledge many stimulating discussions with Kim Border. I also benefited from suggestions by Edi Karni, Graham Loomes, Joe Ostroy, Zvi Safra, William Thomson, and Dan Vincent. I thank SSHRCC for financial support.

†Department of Economics, University of Western Ontario, London N6A 5C2, Canada. E-mail: segal@sscl.uwo.ca

1 Introduction

At the core of all approaches to social choice lies the dissonance between selfish individuals who are seeking the largest possible share of the community wealth for themselves, and the necessity for a compromise. Harsanyi [9, 10] offered two different solutions to this problem. The first requires members of society to give the same weight to the well-being of each one of them, and is achieved through the invention of the following hypothetical lotteries. Let p be a social policy, that is, a lottery yielding social state s_j with probability p_j. Each member of society will perceive it as a lottery that yields him the outcome "be person i at social state j" with probability p_j/n. The expected utility of such a lottery is $\sum_i \sum_j p_j u_i(s_j)/n = \frac{1}{n} \sum u_i(p)$.[1] In the literature. this is called the impartial observer theorem (see Weymark [17, Section 5]).

An alternative approach assumes the existence of a social order over policies, that is, over lotteries over social states. Individual and social preferences satisfy the axioms of expected utility, and in addition it is assumed that if all members of society prefer lottery p to q, then so does society. It follows that social preferences can be represented by a weighted sum of individual utilities of the form $\sum a_i u_i$.

Both approaches are unsatisfactory. The first requires a person to evaluate his and everyone else's well-being equally.[2] Arguably, it also requires a person to be able to identify with each other member of society to such a degree that he can understand and evaluate each social outcome from other people's perspective. Many social thinkers argue that the lack of such ability taints gender and racial relations in modern societies. Moreover. this approach forces all the coefficients $a_1, \ldots, a_n$ to be the same, thus ruling out the possibility of, for example, affirmative action policies.

The second approach relies on the existence of a social order, whose existence is doubtful. Which economic agent holds these preferences? On the other hand, if these preferences are an aggregate of individual preferences, then we should not make independent assumptions about them. And certainly we should not make assumptions that contradict individual preferences (as is done, for example, in Epstein and Segal [7]).

[1] It should be noted that the representation theorem does not assume that all individuals face the same social lottery; see Harsanyi [10].

[2] Since people know who they are, this must be the meaning of putting oneself behind a veil of ignorance.

The two approaches mentioned above offer opposite views on how to solve the conflict between individual selfishness and the necessity for a compromise. The first tries to internalize this conflict by assuming a mental exercise where individuals try to identify with other members of society. The second approach makes both components explicit and removes any elements of preferences for social welfare away from individual preferences.[3] Alternatively, it requires all members of society to agree on one social order.

This paper seeks to make concerns for social justice an explicit part of the personal characteristics of members of society. This is done by assuming that individuals have preferences over social policies. Since a policy is defined for all possible initial endowments, such preferences are defined not over the allocations of one set of social endowments, but over functions that assign each possible initial endowments a possible allocation. These preferences involve some limits on individual selfishness, but these limits seem quite benign. The key axiom is that all individuals agree that giving all the resources of the economy always to the same person is bad, regardless of who that person is. As a result, I am able to show that even though personal tastes and notions of justice may differ, members of society will nevertheless unanimously agree on one social policy being the best. This optimal policy maximizes a weighted sum of individual utility levels, where the weight of an individual given a certain bundle of resources is the inverse of the maximal utility gain this person may enjoy from this bundle. Note that these weights change from one set of social endowments to another.

This paper differs from other models in some major aspects, the most important of which is the fact that it shows the ability to reach social agreement even when members of society have completely different notions of justice. This is partially achieved through the extension of the domain of social preferences. Whereas most models of social choice deal with the allocation of a given set of endowment, this paper deals with preferences over allocation rules that apply to all possible endowments simultaneously.

The main theorem is presented in section 2. I discuss some possible objections to the paper's analysis in section 3, and relate the results to the literature on section 4. All proofs are in an appendix.

[3] Formally, individual preferences in Harsanyi's [9] model are over social policies. But it is clear from the discussion there (see section IV) that the individual preferences are sensitive only to the person's own consumption.

2 Social Welfare

Consider the following social choice problem. A group of people has to find a way to allocate a bundle of goods between them. Society is to choose a (possibly degenerate) lottery p over such allocations, provided the sum of allocated bundles, over individuals, in each outcome of p will not exceed the available resources. Such a lottery induces lotteries over individual consumption bundles. Each person has preferences over such lotteries. Which allocation should society pick?

In this framework, the chosen lottery may be a function of individual preferences over lotteries over private consumption and of the given bundle of goods. One may ask how will it react to changes in individual preferences, as is done, for example, by Maskin [12] or Dhillon and Mertens [5]. Alternatively, one can ask how it will react to changes in the bundle of goods available to society (Yaari [18]).[4] In this paper I adopt the second approach.

Consider a given n-person society. A social state $x = (x_1, \ldots, x_n)$ is an allocation of t goods between members of society, where $x_i \in \mathbb{R}^t$ is the outcome of person i. For $i = 1, \ldots, n$, person i has a preference relation $\succeq_i$ over lotteries over consumption bundles (with $\succ_i$ and $\sim_i$ for the strict and the indifference relations). Assume

E (Expected Utility) For every i, the preference relation $\succeq_i$ satisfies the axioms of expected utility.

It follows that the preference relation $\succeq_i$ can be represented by $V_i(x^1, p^1; \ldots; x^m, p^m) = \sum_{j=1}^m p^j u_i(x^j)$, where $(x^1, p^1; \ldots; x^m, p^m)$ is an arbitrary lottery with outcomes in $\mathbb{R}^t$. I assume that the functions u_i are strictly monotonic, that is, $x \gneqq y$ implies $u_i(x) > u_i(y)$. The functions u_i are unique up to linear transformations, so pick for each person one such utility function. It turns out that in the present model, the choice of von Neumann–Morgenstern utility functions makes no difference (see below).

Let $0 \leqq \underline{\omega}, \overline{\omega} \in \mathbb{R}^t$ be lower and upper bounds to all imaginable resources society may have, and let $\mathcal{G} = \{\omega : \underline{\omega} \leqq \omega \leqq \overline{\omega}\}$. For each $\omega \in \mathcal{G}$, let $\mathcal{L}(\omega)$ be all the possible lotteries over allocations of ω or less between society's n members. The lottery $L \in \mathcal{L}(\omega)$ induces the lottery L_i over consumption

[4] For a similar distinction in the bargaining problem, see Rubinstein, Safra, and Thomson [14].

bundles for consumer i, $i = 1, \ldots, n$. A social policy f is a function assigning each $\omega \in \mathcal{G}$ an element of $\mathcal{L}(\omega)$. (So a policy may allocate less than what is available to society, but not more). Denote by $f_i(\omega)$ the outcome person i receives under f when the available resources to the economy are ω. Note that it may be a lottery.

It may be argued that social policies should be continuous. That is, small changes in ω should result in a small change in the chosen lottery. But there are social policies, like marginal tax rates, that are discontinuous. I will therefore assume that policies are only measurable functions,[5] and denote by $\mathcal{F}$ the set of all such policies. Denote by f^{i*} the policy in $\mathcal{F}$ yielding person i the outcome ω and all other players zero, for all ω. So for all ω, $f^{i*}(\omega) = (0, \ldots, \omega, \ldots, 0)$, $i = 1, \ldots, n$.

Two policies can be mixed as follows. For $f, g \in \mathcal{F}$ and $\alpha \in [0,1]$, the policy $(f, \alpha; g, 1-\alpha)$ assigns the endowments ω the social policy $f(\omega)$ with probability α and the social policy $g(\omega)$ with probability $1-\alpha$. Since $f(\omega)$ and $g(\omega)$ are lotteries over possible allocations, so is $[(f, \alpha; g, 1-\alpha)](\omega)$. In other words, $(f, \alpha; g, 1-\alpha)$ is a social policy in $\mathcal{F}$.

It is evident from the vast literature concerning social choice that different policies may have different sources of appeal. It is thus natural to compare policies not through their outcomes for one given ω, but as functions defined for all possible values of ω. I assume that each member of society has complete and transitive preferences $\succcurlyeq_i$ on $\mathcal{F}$. These preferences represent individual notions of justice, and may of course vary from one individual to another. Denote by $\gg_i$ and $\approx_i$ the strict and indifference relations obtained from $\succcurlyeq_i$, respectively. On $\succcurlyeq_i$ assume:

C (Continuity) $\succcurlyeq_i$ is continuous.

M (Monotonicity) If for all $j = 1, \ldots, n$ and for all $\omega \in \mathcal{G}$, $f_j(\omega) \succeq_j g_j(\omega)$, then $f \succcurlyeq_i g$. If, in addition, there exists a set of positive measure $\mathcal{G}' \subset \mathcal{G}$ such that for every j and for every $\omega \in \mathcal{G}'$, $f_j(\omega) \succ_j g_j(\omega)$, then $f \gg_i g$.

I (Independence) $f \succcurlyeq_i g \iff \forall h \in \mathcal{F}$ and $\forall \alpha \in (0,1]$, $(f, \alpha; h, 1-\alpha) \succcurlyeq_i (g, \alpha; h, 1-\alpha)$.

D (Dictatorship Indifference) $f^{1*} \approx_i \cdots \approx_i f^{n*}$. Moreover, let $\mathcal{G}_1, \ldots,$

[5] See Appendix A for details concerning the mathematical properties of the model.

$\mathcal{G}_n$ be a partition of $\mathcal{G}$.[6] If $f \in \mathcal{F}$ is such that for every $\omega \in \mathcal{G}_j$, $f(\omega) = f^{j*}(\omega)$, then $\forall j$, $f \succcurlyeq_i f^{j*}$.

The continuity assumption relates to the orders $\succcurlyeq_i$, and does not imply continuity of policies. It should be emphasized that condition I does not follow from the assumption that all n players have von Neumann–Morgenstern utilities, as these utility functions represent individual preferences over uncertain consumption, while the relations $\succcurlyeq_i$ are over policies. Recall that such policies are functions whose outcomes are individual lotteries over bundles of commodities. As for condition D, its first part states indifference between all n possible pure dictatorships. The term dictatorship is to be understood as a situation where one person can impose his selfish preferences over society. In this case, such a person will always receive all the social resources. The second part suggests that "mixed dictatorship," in the sense that each person is a dictator only for some values of ω, cannot be worse than a pure one.

Condition D seems plausible if $\succcurlyeq$ is the preference relation an arbitrator has over policies. Impartiality is an essential part of arbitration, and it seems to be at least partially captured by this axiom. But condition D is much stronger than that, as it assumes that each individual is indifferent between all forms of pure dictatorships. Obviously, if members of society are to agree on what social policy to employ, it must be based on individual willingness to consider each other's well-being. Unlike the first of Harsanyi's [9] models, where individuals pretend not to know who they are and therefore give the same weight to everyone, condition D is much weaker, as it requires indifference between the extreme cases of dictatorships. This assumption requires no knowledge of other people preferences, except for the fact that they are strictly monotonic.[7] Also note that condition D does not state indifference regarding who will receive everything for just one value of ω, but for all such values.

Condition D assumes that people realize that dictatorship is morally wrong on its own ground, and not only because someone else may be a dictator. The higher is the value of ω/n, and the higher is the number of commodities t, the more obscene the idea of giving all of it to one person will seem. Hence the requirement that $\omega \geqq \underline{\omega}$. Moreover, if people realize that

[6] $\mathcal{G}_1, \ldots, \mathcal{G}_n$ is a partition of $\mathcal{G}$ if $\cup_{i=1}^n \mathcal{G}_i = \mathcal{G}$ and $i \neq j$ implies $\mathcal{G}_i \cap \mathcal{G}_j = \varnothing$.

[7] Possible allocation rules in situations where some people do not care for all goods are discussed in Yaari and Bar-Hillel [19].

some degree of selflessness is required for agreement to be reached, and that any condition that is strong enough to guarantee agreement will have to be symmetric, then condition D seems to be a minimal requirement.

For each possible lottery $L \in \mathcal{L}(\omega)$ over allocations of ω, the functions $u_1, \dots, u_n$ determine a utility allocation between the n members of society. Define $S(\omega)$ to be the set of utility allocations that can be obtained from lotteries in $\mathcal{L}(\omega)$. That is, $S(\omega) = \{(u_1(L_1), \dots, u_n(L_n)) : L_1, \dots, L_n$ are induced from $L \in \mathcal{L}(\omega)\}$. Since $u_1, \dots, u_n$ are von Neumann–Morgenstern utilities and all lotteries over allocations are possible, $S(\omega)$ is convex. And since giving everybody zero is a possible allocation, $S(\omega)$ is comprehensive with respect to the point $z^* = (u_1(0), \dots, u_n(0))$. That is, $x \in S(\omega)$ and $z^* \leqq y \leqq x$, imply $y \in S(\omega)$.

Although each social policy $f \in \mathcal{F}$ specifies for each ω a unique utility allocation in $S(\omega)$, the opposite is not true, as different policies may yield the same allocation of utilities. However, by monotonicity all such policies are indifferent to each other in the relation $\succcurlyeq_i$.

Define the following set of policies $\mathcal{F}^*$. The policy f is in $\mathcal{F}^*$ if for each ω it yields the utility allocation $(s_1, \dots, s_n)$, satisfying

$$(s_1, \dots, s_n) \in \arg\max_{v \in S(\omega)} \sum_{i=1}^{n} \frac{v_i}{u_i(\omega) - u_i(0)} \tag{1}$$

Policies in $\mathcal{F}^*$ maximize for each bundle ω a weighted sum of the individual utilities, where the weight of person i is the inverse of the maximal gain in utility he may receive from this endowment, that is, $1/[u_i(\omega) - u_i(0)]$.

This functional form was suggested in the context of the bargaining problem by Cao [4], and in the context of social choice by Dhillon and Mertens [5], who call it relative utilitarianism. I discuss the differences between their approach and the present one in section 4 below.

Let $H(\omega)$ be the hyperplane through the points $(u_1(\omega), 0, \dots, 0), \dots, (0, \dots, 0, u_n(\omega))$, and let $H^*(\omega)$ be the highest hyperplane that is parallel to $H(\omega)$ and tangent to $S(\omega)$. The policies in $\mathcal{F}^*$ pick for ω a lottery over allocations that yields a utility distribution at such a tangency point. Fig. 1 depicts the case $n = 2$ for one particular ω. In this picture, z^* denotes the utility allocation obtained from giving both players 0, a and b denote the utility allocations corresponding to $(\omega, 0)$ and $(0, \omega)$, respectively, and c

denotes the utility allocation under any policy in $\mathcal{F}^*$. These policies enjoy the following property.

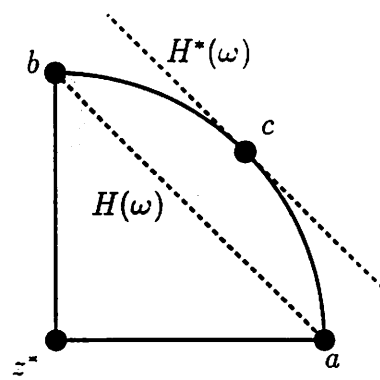

Figure 1: The case $n = 2$.

Theorem 1 *Suppose that the preference relation $\succcurlyeq_i$ satisfies conditions C, M, I, and D. Then all policies in $\mathcal{F}^*$ are $\succcurlyeq_i$-best policies in $\mathcal{F}$.*

To make sure this is not an empty theorem, one has to show that the four axioms are consistent, that is, that there are preference relations satisfying all of them. This is done in the following example.

Example 1 Let μ be a measure with full support over $\mathcal{G}$. Define $f \succcurlyeq g$ iff

$$\int_{\mathcal{G}} \sum_{i=1}^{n} \frac{u_i(f_i(\omega))}{u_i(\omega) - u_i(0)} d\mu(\omega) \geqslant \int_{\mathcal{G}} \sum_{i=1}^{n} \frac{u_i(g_i(\omega))}{u_i(\omega) - u_i(0)} d\mu(\omega)$$

It is easy to verify that $\succcurlyeq$ satisfies conditions C, M, I, and D. ∎

If the outer boundary of $S(\omega)$ is flat, then the utility allocation of eq. (1) may be not uniquely defined. However, if all the utility functions u_i are strictly concave, then the outer boundary is strictly concave, and all policies in $\mathcal{F}^*$ yield only one possible utility allocation for every ω. This utility allocation is considered best by all members of society, despite the fact that the preferences $\succcurlyeq_i$ may differ. Society should then use such a policy.

3 Discussion

In this section I discuss some possible objections to the model presented above.

ISN'T THIS MODEL ESSENTIALLY HARSANYI'S? No. It is true that Harsanyi's [9] second model assumes that individuals and society have expected utility preferences, connected by a Pareto assumption which is parallel to condition M. However, these preferences are different from those used here. In Harsanyi, individual and social preferences are over allocations of a given ω between members of society, although individual preferences are assumed to depend only on individual consumption (see [9, section IV]). Likewise here, individual preferences are over individual consumption, but the preferences $\succcurlyeq_i$, which are supposed to parallel Harsanyi's social preferences, are over social policies that apply to all values of ω simultaneously. The distinction is not just that the domain of preferences is different from that of Harsanyi's, but that the domains of "private" and "social" preferences are different. (This is why the proof of Theorem 1 cannot use standard social choice theorems).[8] To illustrate the difference between the models, observe that in Example 1, the induced order on possible allocations of a given ω is flat, and all allocations are equally attractive.

IS THIS A MODEL OF UTILITARIANISM? No, is what Sen [15], Weymark [17, Section 6], and Roemer [13, Ch. 4] would probably say. For utilitarianism, one needs the ability to compare individual utility levels, which this model, like many others (e.g., Harsanyi [9], Maskin [12], or Epstein and Segal [7]) cannot. Moreover, the present model cannot even compare relative utilities, as the n von Neumann–Morgenstern indexes can be chosen independently of each other (cf. Weymark [17, p. 302]). Also, the optimal policy is not sensitive to the particular choice of utilities. If u_i is replaced with $u_i' = a_i u_i + b_i$ with $a_i > 0$, then the vector $(a_1 s_1 + b_1, \ldots, a_n s_n + b_n)$ satisfies eq. (1) with respect to the utilities $u_1', \ldots, u_n'$ and the set $\mathcal{F}^*$ remain the same. Therefore, a statement like "person i's utility is twice that of person j" is meaningless in the present context. Although this is true, it should also be noted that no

[8] In Harsanyi's [10] impartial observer theorem the domains of the observer's preferences and those of the individual preferences are different. But the representation theorem there eventually follows from the fact that individual preferences over lotteries are expected utility. This is not the case here, where assumption E does not imply assumption I.

other model of utilitarianism is consistent with the above axioms, as follows from Lemma 1. In other words, even though the aim of Theorem 1 is to find an optimal policy, it also determines individual weights. If, independently of this model, interpersonal comparisons of utility are possible, they must use the weights obtained by Theorem 1.[9]

Lemma 1 *If f is a $\succcurlyeq_i$-best policy, then it agrees almost everywhere with a policy in $\mathcal{F}^*$. Therefore, if an optimal policy yields for all ω the utility allocation*

$$(s_1, \ldots, s_n) \in \arg\max_{v \in S(\omega)} \sum_{i=1}^{n} a_i(\omega) v_i$$

then for almost all ω, $a_i(\omega) = 1/[u_i(\omega) - u_i(0)]$.

WHY SHOULD THE PREFERENCES $\succcurlyeq_i$ AGREE WITH THE INDIVIDUAL PREFERENCES $\succeq_i$? Consider the following sterilised problem.[10] There is a certain number of dialysis machines, which is less than the number of people in need of them. There are (at least) two ways in which society can allocate these machines. *1.* First come first serve: When someone needs a machine he joins the line. He may die before he receives a machine, but if he gets one, he will use it as long as he needs it (that is, until he dies). *2.* First in first out: When someone needs a machine he gets the machine that was used by the longest served patient. He will then use the machine until he becomes the longest user and a new person needs a machine.

If the economy is sufficiently large and the ratio between the number of patients and the number of machines is bounded away from one, then under both systems all machines will be constantly occupied. In other words, the expected value of the length of time t a new patient will use a machine is the same under both systems. The first, however, is an uncertain distribution over t, while the second is close to yielding the average $\bar{t}$ with probability one (excluding the possibility of dying of other causes before the patient becomes the longest user). There is a number of studies suggesting a widespread aversion to gambling with one's life and health (see e.g. Bombardier et al. [2]

[9] As argued by Yaari [18], the fact that individual weights vary with the social resources is consistent with utilitarianism.

[10] I am especially thankful to Graham Loomes for many discussions of this example.

or Gafni and Torrance [8]). If individuals are risk averse and they do not have to worry about the allocation mechanism, then they should clearly prefer the second system. Societies tend to prefer the first system over the second, hence a violation of condition M.

Formally, this example does not pose a problem to Harsanyi's "Pareto" assumption, as in his model, individual preferences are over social allocations, and may not represent preferences over personal consumption. However, as mentioned above, Harsanyi's view is that these individual preferences are sensitive only to individual consumption. The dialysis machine example thus challenges all models that use versions of the "Pareto" assumption.

I believe that this example implies that the present model, and other models, should not be applied to situations where attitudes towards an allocation mechanism go beyond the distribution of goods induced by it. In this particular case, there are hidden costs involved in the allocation mechanisms that are not explicit in the induced utility distributions. Condition M, and Harsanyi's "Pareto" assumption, consider only individual preferences over possible allocations. Not surprisingly, they cannot grasp the complexity of the two mechanisms described above.

ISN'T THIS MODEL VULNERABLE TO DIAMOND'S CRITICISM? Diamond [6] claims that if society is indifferent between giving a certain good to person 1 and giving it to person 2, it is better to randomize over these two policies. Likewise, one can argue here that a random dictator is better than a deterministic one. I disagree. One reason Diamond's argument seems so plausible is that randomization soothes away conflicts by treating similar agents equally. But in this model, if society has to choose a dictator, then there is no conflict, because by condition I, everyone agrees that all dictators are equally bad. If there is no conflict, there is nothing to be gained by randomization.

Another justification for randomization preferences is that there are situations where ex post equality is impossible, for example, when an indivisible good is to be allocated. Randomization yields ex ante equality, which is better than no equality at all. However, if all good are divisible, and if the individual utility functions are concave, then both ex ante and ex post equality are feasible without randomization.

4 Conclusion

This paper presents a model where individuals who have different tastes, and different notions of justice, will nevertheless be able to agree on one policy as being the best. This is done by separating preferences over consumption from preferences over social policies. A related approach is suggested by Karni and Safra [11], but they assume that all social preferences are symmetric between individuals. (Example 1 shows that this paper's axioms do not imply such symmetry). Also, like Harsanyi [9], these preferences are over the possible allocations of one given bundle ω. (In their paper, ω is one unit of a nondivisible good).

The analysis of this paper can be applied to the bargaining problem. An n person bargaining problem is a pair (S, d), where $S \subset \mathbb{R}^n$ is the set of possible von Neumann–Morgenstern utility allocations obtained from the game, and d is the disagreement point. A solution is a function F assigning each game (S, d) a point in S. Following Border and Segal [3], one can define preferences over such solutions. The axioms of section 2 can be applied to such preferences. (Condition D will require that $F^{1*} \approx_i \cdots \approx_i F^{n*}$, where F^{j*} always yields player $k \neq j$ his disagreement utility level d_k, and player j his highest possible utility from each game). Similarly to Theorem 1, the set of $\succcurlyeq_i$-best solutions include those solutions that maximize a weighted sum of the utilities for each game, where the weight of player i is the inverse of the difference between the maximal utility level he can reach in the game and his disagreement utility level. This solution was first suggested by Cao [4], but it suffers from some undesired properties, like violation of disagreement point monotonicity (see Thomson [16, p. 1261]).

An essential assumption of the present paper is that the domain of the preferences $\succcurlyeq_i$ is a set of social policies that apply to all values of $\omega \in \mathcal{G}$. In this model, individuals know who they are and who else belongs to their society, but they do not know what resources will be available to them in the future. In this, I follow Yaari [18], where the dependence of social policy on the resources of the economy is explicit. Other models hold social endowments fixed, and analyze social choice as a function of preferences. This is Sen's [15, p. 1124] understanding of Harsanyi [9], and it is explicit in Dhillon and Mertens [5], where axioms are based on changing utilities and the number of players. Such models follow from Arrow's impossibility theorem (and are also related to the existence of incentive compatible mechanisms), where policies (and mechanisms) should apply to all utility portfolios. Since this paper is more interested in justice and possible cooperation, it seems natural to hold individuals fixed, and vary the resources.

The weakest part of the model is condition D. But if we seek unanimity, some consideration for other people's well-being must be assumed. In this paper people realize that it is in their self interest that society will reach an agreeable allocation of its resources. It is therefore essential that each member of society will accept the fact that he will not be able to attain the highest possible utility level the economy may provide him (that is, $u_i(\omega)$), and that compromise is necessary for an agreement. In this context, I believe condition D to be quite convincing. Another reason why condition D may be acceptable is that it is very unlikely anyone will ever have to make a choice between pure dictatorships. This may make people more sympathetic to the moral aspects of this condition, and to its implications.

A The Structure of Policies

Assume n individuals and t commodities, so an allocation is a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^{n \times t}$. The set $\mathcal{G} = \{\omega : \underline{\omega} \leqq \omega \leqq \overline{\omega}\}$ is endowed with the Lebesgue measure λ on $\mathbb{R}^k$. For $\omega \in \mathcal{G}$, let $X(\omega) = \{x : \sum_{i=1}^n x_i \leqq \omega\}$ be the set of possible allocations of ω or less. The set of lotteries over $X(\omega)$, denoted $\mathcal{L}(\omega)$, is endowed with the topology of weak convergence. Note that $X(\omega) \subseteq X(\overline{\omega})$ and that $\mathcal{L}(\omega) \subseteq \mathcal{L}(\overline{\omega})$. A policy f assigns to each $\omega \in \mathcal{G}$ a lottery $f(\omega) \in \mathcal{L}(\omega)$. Policies are assumed to be Borel-measurable.

Since $X(\overline{\omega})$ is a compact metric space, the set $\mathcal{L}(\overline{\omega})$ with the topology of weak convergence is a compact metrizable space (see Aliprantis and Border [1, Lemma 3.69]). Denote this metric $d_{\mathcal{L}}$. The set of all policies $\mathcal{F}$ is a metric space under the metric

$$d_{\mathcal{F}}(f,g) = \sup_{\omega \in \mathcal{G}} d_{\mathcal{L}}(f(\omega), g(\omega))$$

The preference relation $\succcurlyeq_i$ over $\mathcal{F}$ is assumed to be continuous with respect to this metric (see condition C).

For $\omega \in \mathcal{G}$, let $S(\omega) = \{(u_1(L_1), \ldots, u_n(L_n)) : L_1, \ldots, L_n$ are the individual lotteries induced from $L \in \mathcal{L}(\omega)\}$. From a policy $f \in \mathcal{F}$ define a function $\varphi_f : \mathbb{R}^t \to \mathbb{R}^n$, given by $\varphi_f(\omega) = (u_1(f_1(\omega)), \ldots, u_n(f_n(\omega)))$. ($f_i(\omega)$, which may be a lottery, is the outcome person i receives under f when the available resources to the economy are ω). Let $\Phi = \{\varphi_f : f \in \mathcal{F}\}$. Since the utility functions $u_1, \ldots, u_n$ are continuous, each $\varphi \in \Phi$ is measurable. By condition M, if $\varphi_f = \varphi_g$, then $f \approx_i g$ for every i. It follows that the order $\succcurlyeq_i$ on $\mathcal{F}$ induces a natural order on Φ, which with a slight abuse of notations will also be denoted $\succcurlyeq_i$. That is, $\varphi_f \succcurlyeq_i \varphi_g$ iff $f \succcurlyeq_i g$.

For $\varphi, \psi \in \Phi$, define

$$d(\varphi, \psi) = \sup_{\omega \in \mathcal{G}} \| \varphi(\omega) - \psi(\omega) \|$$

Conditions C, M, I, and D on $\succcurlyeq_i$ over $\mathcal{F}$ easily translate into conditions on $\succcurlyeq_i$ over Φ. For condition D*, define $\varphi^{i*}(\omega)$ to give player i the utility level $u_i(\omega)$, and every other player $j \neq i$, $u_j(0)$.

C* $\succcurlyeq_i$ is a closed subset of $\Phi \times \Phi$.

M* If for all $\omega \in \mathcal{G}$, $\varphi(\omega) \geqq \psi(\omega)$, then $\varphi \succcurlyeq_i \psi$. If, in addition, there exists a set of positive measure $\mathcal{G}' \subset \mathcal{G}$ such that for every j and for every $\omega \in \mathcal{G}'$, $\varphi_j(\omega) > \psi_j(\omega)$, then $\varphi \gg_i \psi$.

I* $\varphi \succcurlyeq_i \psi \iff \forall \rho \in \Phi$ and $\forall \alpha \in (0,1]$, $\alpha\varphi + (1-\alpha)\rho \succcurlyeq_i \alpha\psi + (1-\alpha)\rho$.

D* $\varphi^{1*} \approx_i \cdots \approx_i \varphi^{n*}$. Moreover, let $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be a partition of $\mathcal{G}$. If $\varphi \in \Phi$ is such that for every $\omega \in \mathcal{G}_j$, $\varphi(\omega) = \varphi^{j*}(\omega)$, then $\forall j$, $\varphi \succcurlyeq_i \varphi^{j*}$.

B Proofs

Proof of Theorem 1 The theorem is proved through a sequence of lemmas. For simplicity, I omit the index i from $\succcurlyeq_i$.

Lemma 2 *Let $\mathcal{G}_1, \mathcal{G}_2$ be a partition of $\mathcal{G}$. Let $\varphi^1, \psi^1, \varphi^2, \psi^2 \in \Phi$ such that on $\mathcal{G}_1$, $\varphi^i = \psi^i$, $i = 1, 2$, and on $\mathcal{G}_2$, $\varphi^1 = \varphi^2$ and $\psi^1 = \psi^2$. Then $\varphi^1 \succcurlyeq \varphi^2 \iff \psi^1 \succcurlyeq \psi^2$.*

Proof: For every ω, $\frac{1}{2}\varphi^1 + \frac{1}{2}\psi^2 = \frac{1}{2}\varphi^2 + \frac{1}{2}\psi^1$. By condition I*, $\varphi^1 \succcurlyeq \varphi^2$ iff $\frac{1}{2}\varphi^1 + \frac{1}{2}\psi^1 \succcurlyeq \frac{1}{2}\varphi^2 + \frac{1}{2}\psi^1 = \frac{1}{2}\varphi^1 + \frac{1}{2}\psi^2$ iff $\psi^1 \succcurlyeq \psi^2$. □

Lemma 3 *Let $\varphi^{1*}, \ldots, \varphi^{n*}$ be as in condition D^*. let $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be a partition of $\mathcal{G}$, and let $\varphi \in \Phi$ such that on $\mathcal{G}_i$, $\varphi = \varphi^{i*}$, $i = 1, \ldots, n$. Then $\forall i$, $\varphi \approx \varphi^{i*}$. In other words, the weak preference sign in the second part of condition D^* must be indifference.*

Proof: Assume first that there are $j \neq k$ such that for $i \notin \{j, k\}$, $\mathcal{G}_i = \varnothing$. Define $\varphi^1 = \varphi^{j*}$, $\psi^2 = \varphi^{k*}$, and $\psi^1 = \varphi$. Also. let $\varphi^2 = \varphi^{k*}$ on $\mathcal{G}_j$ and $\varphi^2 = \varphi^{j*}$ on $\mathcal{G}_k$. By Lemma 2. $\varphi^{j*} \succcurlyeq \varphi^2 \iff \varphi \succcurlyeq \varphi^{k*}$. By condition D*, $\varphi^2, \varphi \succcurlyeq \varphi^{j*} \approx \varphi^{k*}$. Hence $\varphi \approx \varphi^{j*}$.

We prove the lemma by induction on ℓ, under the assumption that for $i \notin \{j_1, \ldots, j_\ell\}$, $j_1 < \cdots < j_\ell$, $\mathcal{G}_i = \varnothing$. We proved it already for the case $\ell = 2$. Suppose it holds for $2 \leqslant \ell \leqslant n-1$, and prove for $\ell + 1$. Define $\varphi^1 = \varphi$ on $\cup_{i=1}^{\ell} \mathcal{G}_{j_i}$, and $\varphi^1 = \varphi^{j_1*}$ on $\mathcal{G}_{j_{\ell+1}}$, $\psi^2 = \varphi^{j_{\ell+1}*}$. and $\psi^1 = \varphi$. Also, let $\varphi^2 = \varphi^{j_{\ell+1}*}$ on $\cup_{i=1}^{\ell} \mathcal{G}_{j_i}$ and $\varphi^2 = \varphi^{j_1*}$ on $\mathcal{G}_{j_{\ell+1}}$. By the induction hypothesis. $\varphi^1 \approx \varphi^2 \approx \varphi^{1*}$ (the ℓ sets are $\mathcal{G}_{j_1} \cup \mathcal{G}_{j_{\ell+1}}, \mathcal{G}_{j_2}, \ldots \mathcal{G}_{j_\ell}$). Hence by Lemma 2, $\varphi = \psi^1 \approx \psi^2 = \varphi^{j_{\ell+1}*}$. □

Let $z^* = (u_1(0), \ldots, u_n(0))$. For $\omega \in \mathcal{G}$ and $x \in S(\omega)$, let $\mathbf{L}$ be the line through z^* and x and let H be the plane

$$\sum_{i=1}^{n} \frac{x_i - u_i(0)}{u_i(\omega) - u_i(0)} = 1$$

Denote the intersection point of $\mathbf{L}$ and H by d and define $\alpha = \alpha(x, \omega) = \| x - z^* \| \, / \, \| d - z^* \|$. Let $\beta = \beta(x, \omega) \in \mathbb{R}^n_+$ be given by $\beta_i = (d_i - u_i(0))/(u_i(\omega) - u_i(0))$. Clearly. $\alpha \in [0, n]$ and $\sum \beta_i = 1$. In the sequel. for $a \in \mathbb{R}$. $[a]$ denotes the largest integer not bigger than a.

For $x \in S(\omega)$, define $\langle x \rangle^k$ and $\langle d \rangle^k$ by

1. $\langle d \rangle^k$ is in H, and satisfies

$$\langle d \rangle^k - z^* = \left(\frac{[k\beta_1(x, \omega)]}{k}, \ldots, \frac{[k\beta_{n-1}(x, \omega)]}{k}, \frac{k - \sum_{i=1}^{n-1} [k\beta_i(x, \omega)]}{k} \right)$$

2. $\langle x \rangle^k$ is on the line through z^* and $\langle d \rangle^k$, and satisfies

$$\frac{\| \langle x \rangle^k - z^* \|}{\| \langle d \rangle^k - z^* \|} = \frac{\max\{0, [k\alpha(x, \omega)] - n\}}{k}$$

It follows that $\langle x\rangle^n$ is in $S(\omega)$. Define a function $\theta^k: \Phi \to \Phi$ by $\theta^k(\varphi)(\omega) = \langle\varphi(\omega)\rangle^k$. Since the sets $S(\omega)$ are convex, there are policies $f^k \in \mathcal{F}$ that will generate these utility allocations. Denote $\langle\varphi\rangle^k = \theta^k(\varphi)$.

Definition 1 *The two functions $\varphi, \psi \in \Phi$ are said to be equivalent to each other if for every $\omega \in \mathcal{G}$, $\sum(\varphi_i(\omega) - u_i(0))/(u_i(\omega) - u_i(0)) = \sum(\psi_i(\omega) - u_i(0))/(u_i(\omega) - u_i(0))$. This relation is denoted $\varphi I \psi$.*

Fact 1 *Let $\varphi, \psi \in \Phi$. If $\varphi I \psi$, then $\langle\varphi\rangle^k I \langle\psi\rangle^k$.*

Fact 2 *Let $\varphi \in \Phi$. Then $\langle\varphi\rangle^k \underset{k\to\infty}{\longrightarrow} \varphi$.*

Lemma 4 *Let $\mathcal{G}_1, \mathcal{G}_2$ be a partition of $\mathcal{G}$ and let $\varphi, \psi \in \Phi$ such that*

1. *On $\mathcal{G}_1$, $\varphi = \psi$;*
2. *For $\omega, \omega' \in \mathcal{G}_2$,*
   (a) *$\alpha^* := \alpha(\varphi(\omega), \omega) = \alpha(\varphi(\omega'), \omega') = \alpha(\psi(\omega), \omega) = \alpha(\psi(\omega'), \omega')$;*
   (b) *$\beta^{1*} := \beta(\varphi(\omega), \omega) = \beta(\varphi(\omega'), \omega')$; and*
   (c) *$\beta^{2*} := \beta(\psi(\omega), \omega) = \beta(\psi(\omega'), \omega')$.*

Then $\varphi \approx \psi$.

Proof: Suppose first that $\alpha^* \leqslant 1$. By Lemma 3, $\varphi^{1*} \approx \rho^i$, $i = 1, \ldots, n$, where $\rho^i = \varphi^{1*}$ on $\mathcal{G}_1$, and $\rho^i = \varphi^{i*}$ on $\mathcal{G}_2$. By condition I*, $\varphi^1 := \sum \beta_i^{1*}\rho^i \approx \psi^1 := \sum \beta_i^{2*}\rho^i$. Let $\mathbf{0} \in \Phi$ be the "zero policy", assigning each $\omega \in \mathcal{G}$ the vector of utility levels z^* (which means that everyone receives the allocation 0). Again by condition I*, $\varphi^2 := \alpha^*\varphi^1 + (1 - \alpha^*)\mathbf{0} \approx \psi^2 := \alpha^*\psi^1 + (1 - \alpha^*)\mathbf{0}$. On $\mathcal{G}_2$, $\varphi^2 = \varphi$ and $\psi^2 = \psi$. (Clearly, if $\alpha(\varphi(\omega), \omega) = \alpha(\psi(\omega), \omega)$, and $\beta(\varphi(\omega), \omega) = \beta(\psi(\omega), \omega)$, then $\varphi(\omega) = \psi(\omega)$). Also, on $\mathcal{G}_1$, $\varphi^2 = \psi^2$. Therefore, since on $\mathcal{G}_1$, $\varphi = \psi$, it follows by Lemma 2 that $\varphi \approx \psi$.

Suppose now that $\alpha^* > 1$. Note that on $\mathcal{G}_1$, $\varphi^1 = \psi^1$. Let $\varphi^3, \psi^3 = \alpha^*\varphi^{1*} + (1 - \alpha^*)\mathbf{0}$ on $\mathcal{G}_1$, and on $\mathcal{G}_2$, let $\varphi^3 = \varphi^1$ and $\psi^3 = \psi^1$. By Lemma 2, $\varphi^3 \approx \psi^3$. Also, let $\varphi^4 = \psi^4 = \varphi^{1*}$ on $\mathcal{G}_1$, and on $\mathcal{G}_2$, let $\varphi^4 = \varphi$ and $\psi^4 = \psi$. Now $\varphi^3 = (1/\alpha^*)\varphi^4 + [1 - (1/\alpha^*)]\mathbf{0}$, while $\psi^3 = (1/\alpha^*)\psi^4 + [1 - (1/\alpha^*)]\mathbf{0}$, hence $\varphi^4 \approx \psi^4$. Again by Lemma 2, $\varphi \approx \psi$. □

Lemma 5 *If $\langle\varphi\rangle^k I \langle\psi\rangle^k$, then $\langle\varphi\rangle^k \approx \langle\psi\rangle^k$.*

Proof: let

- $\mathcal{G}_i^1 = \{\omega : \alpha(\langle\varphi\rangle^k(\omega),\omega) = \alpha(\langle\psi\rangle^k(\omega),\omega) = i/k\}$, $i = 0,\dots,nk$
- For $j = (j_1,\dots,j_n)$, $\mathcal{G}_j^2 = \{\omega : \beta(\langle\varphi\rangle^k(\omega),\omega) = (j_1/k,\dots,j_n/k)\}$, where $j_1 = 0,\dots,k$; $j_2 = 0,\dots,k-j_1$; $\dots$; $j_{n-1} = 0,\dots,k-\sum_{i=1}^{n-2} j_i$; and $j_n = k - \sum_{i=1}^{n-1} j_i$.
- For $\ell = (\ell_1,\dots,\ell_n)$, $\mathcal{G}_\ell^3 = \{\omega : \beta(\langle\psi\rangle^k(\omega),\omega) = (\ell_1/k,\dots,\ell_n/k)\}$, where $\ell_1 = 0,\dots,k$; $\ell_2 = 0,\dots,k-\ell_1$; $\dots$; $\ell_{n-1} = 0,\dots,k-\sum_{i=1}^{n-2} \ell_i$; and $\ell_n = k - \sum_{i=1}^{n-1} \ell_i$.

Let m^* be the number of all possible combinations of $\alpha(\langle\varphi\rangle^k(\omega),\omega)$ and $\beta(\langle\varphi\rangle^k(\omega),\omega)$ (which is also the number of all possible combinations of $\alpha(\langle\psi\rangle^k(\omega),\omega)$ and $\beta(\langle\psi\rangle^k(\omega),\omega)$). Let $\mathcal{G}_1,\dots,\mathcal{G}_{m^*}$ be all the possible intersections of the form $\mathcal{G}_i^1 \cap \mathcal{G}_j^2 \cap \mathcal{G}_\ell^3$. Of course, for some m, $\mathcal{G}_m$ may be empty. For $m = 0,\dots,m^*$, define $\varphi^m = \psi^m = \mathbf{0}$ on $\cup_{i>m}\mathcal{G}_i$, and on $\cup_{i\leqslant m}\mathcal{G}_i$, $\varphi^m = \langle\varphi\rangle^k$ and $\psi^m = \langle\psi\rangle^k$. Note that $\varphi_{m^*} = \langle\varphi\rangle^k$ and $\psi_{m^*} = \langle\psi\rangle^k$. We prove by induction that for all $m = 0,\dots,m^*$, $\varphi^m \approx \psi^m$. The claim is trivially true for $m = 0$. Suppose it holds for m, and prove for $m+1$.

Define $\chi \in \Phi$ such that on $\cup_{i\neq m+1}\mathcal{G}_i$, $\chi = \varphi^{m+1}$, and on $\mathcal{G}_{m+1}$, $\chi = \psi^{m+1}$. By Lemma 4, $\chi \approx \varphi^{m+1}$. Also, it follows by Lemma 2 and the induction hypothesis that $\chi \approx \psi^{m+1}$. Therefore, $\varphi^{m+1} \approx \psi^{m+1}$. □

Suppose $\varphi I \psi$. By Fact 1, for every k, $\langle\varphi\rangle^k I \langle\psi\rangle^k$, therefore $\langle\varphi\rangle^k \approx \langle\psi\rangle^k$. By Fact 2 and the continuity of $\succcurlyeq$, it follows that $\varphi \approx \psi$. Theorem 1 now follows from condition M*. ■

Proof of Lemma 1 Let $f \in \mathcal{F}$, and suppose that there is $\mathcal{G}' \subset \mathcal{G}$ such that $\lambda(\mathcal{G}') > 0$, and such that on $\mathcal{G}'$, $\varphi_f(\omega) \notin \arg\max_{v\in S(\omega)} \sum_{i=1}^n v_i/(u_i(\omega)-u_i(0))$. By Lemma 4 and condition M*, f is not optimal. ■

References

[1] Aliprantis, C. D. and K. C. Border. 1994. *Infinite dimensional analysis.* Berlin: Springer–Verlag.

[2] Bombardier, C. et al. 1982. "Comparison of three preference measurement methodologies in the evaluation of a functional status index," in R. Deber and G. Thompson (eds), *Choices in health care: Decision making and evaluation of effectiveness*, Toronto: University of Toronto Press.

[3] Border, K. C. and U. Segal. 1995. Preferences over solutions to the bargaining problem. *Econometrica*, forthcoming.

[4] Cao, X. 1981. Preference functions and bargaining solutions. *IEEE* 164–171.

[5] Dhillon, A. and J.-F. Mertens. 1992. Relative Utilitarianism. Mimeo.

[6] Diamond, P. A. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy*, 75:765–766.

[7] Epstein, L. G. and U. Segal. 1992. Quadratic social welfare functions. *Journal of Political Economy* 100:691–712.

[8] Gafni, A. and G.W. Torrance. 1994. "Risk attitude and time preference in health," *Management Science* 30:440–451.

[9] Harsanyi, J. C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63:309–321.

[10] Harsanyi, J. C. 1977. *Rational behavior and bargaining equilibrium in games and social situations.* Cambridge: Cambridge University Press.

[11] Karni, E. and Z. Safra. 1995. Individual sense of justice and social welfare functions. Mimeo.

[12] Maskin, E. 1978. A Theorem of Utilitarianism. *Review of Economic Studies* 45:93–96.

[13] Roemer, J. E. 1996. *Theories of Distributive Justice*. Cambridge MA: Harvard University Press.

[14] 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* 60:1171–1186.

[15] Sen, A. 1986. Social choice theory. In K.J. Arrow and M.D. Intriligator (eds.): *Handbook of Mathematical Economics*, vol. III, ch. 22. Amsterdam: North Holland.

[16] Thomson, W. 1994. Cooperative models of bargaining. In R.J. Aumann and S. Hart (eds.): *Handbook of Game Theory*, vol. II, ch. 35. Amsterdam: North Holland.

[17] Weymark, J.A. 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In J. Elster and J.E. Roemer (eds.): *Interpersonal Comparisons of Well-Being*, ch. 8. Cambridge: Cambridge University Press.

[18] Yaari, M. E. 1981. Rawls, Edgeworth, Shapley, Nash: Theories of distributive justice re-examined. *Journal of Economic Theory* 24:1–39.

[19] Yaari, M. E. and M. Bar-Hillel. 1984. On dividing justly. *Social Choice and Welfare* 1:1–24.