MORAL SENTIMENTS AND SOCIAL CHOICE[*]

Edi Karni[†] and Zvi Safra[‡]

April 1, 2007

Corresponding author: Edi Karni, Dept. of economics, Johns Hopkins University, Baltimore MD 21218, USA

Running title: Moral sentiments and social choice

[*] While working on the research that is summarized in this paper the authors visited several institutions. Edi Karni visited the Berglas School of Economics, Tel Aviv University and GRID, at Ecole Normale Superieur de Cachan. Zvi Safra visited CERMCEM and EUREQua, at the University of Paris I and LIP6, at the University of Paris 6. The hospitality of these institutions is gratefully acknowledged. We are also grateful to the editor and two anonymous referees for their very useful comments and suggestions. Zvi Safra thanks The Israel Science Foudation (grant 1299/05) for its support.

[†] Johns Hopkins University (karni@jhu.edu).

[‡] The College of Management and Tel Aviv University (zvisa@colman.ac.il).

Abstract

We examine the implications, for social choice, of individuals having an intrinsic sense of fairness. Taking the viewpoint that social justice reflects the moral attitudes of the constituent members, we analyze the effect of the intensity of the individual sense of fairness on the solution of Nash bargaining over random allocation procedures. We use a stylized model of university admission policies to illustrate our approach. We show that even if social policies are ultimately determined by the bargaining power of the different groups, a society whose members have a common notion of fairness tends to implement fairer admission policies when the intensity of the sense of fairness of individual members increases.

1 A Positive Approach to Normative Economics

Social policies and institutions are shaped by the power of the constituent members to influence these policies and institutions. A well recognized source of power is the conviction of individuals that the policies they support are just. In general, different individuals may hold distinct ideas of fairness with varying degrees of conviction. Hence the design of policies and institutions ultimately depends on the degree to which the idea of fairness is shared by the individual members, on the intensity of their moral conviction, and on the mechanism by which individual preferences are translated into social decisions.

In this paper we investigate the implications of individual concern for fairness in shaping social policies. We assume that such policies are the outcome of bargaining among social groups with diverse interests that may or may not subscribe to a common notion of fairness. Consequently, policies are shaped by the relative bargaining power of the different groups which depends, among other things, on the intensity of their moral conviction. In other words, to the extent that these policies are compatible with some notion of fairness, it is because the individual members of the groups that subscribe to this notion of fairness have regard for and are willing to act upon it. To give the analysis a concrete context and interpretation we use, by way of illustration, stylized admission policies at selective colleges and universities. However, we do not regard this work as a contribution towards understanding the actual formulation of admission policies. In fact, we do not presume that the Nash bargaining model [15] is the best framework for the analysis of admission policies. Rather we use the specific context to illustrate the ways by which individuals sense of fairness manifest itself in the context of the Nash bargaining model to shape social policies. While the context is specific, the approach taken here is general and may be applied, with appropriate modifications, to the analysis of other social policies and institutions using the Nash bargaining solution or, for that matter, other procedures that map individual preferences into social policies, that seem appropriate.

Our analysis highlights several aspects of the issue: the effect on policies of the ideas of fairness and the degree to which they are shared among individuals belonging to the same society, the intensity with which these ideas are held by various individuals, and the interaction between individual preferences incorporating a sense of fairness and the social decision making process. We assume that people possess an intrinsic sense of fairness.[1] This means that acting consistently with one's notion of what is right is a self-rewarding activity. Put differently, a sense of fairness is a moral sentiment, that is, an emotion and acting virtuously produces a gratifying feeling.[2] In Karni and Safra [10] we developed an axiomatic model of individual behavior incorporating this idea. In that work we considered individual choice among procedures that rely on the outcome of lots to allocate an indivisible good among different claimants. We show below that policies whose implication for specific individuals depend on their position in the population distribution of some characteristics may be modeled using a similar analytical framework.

[1] This idea has a long history that goes back to St. Anselm (see discussion and references in Jasso, [7]). Karni and Safra [10] provides additional arguments and further references.

[2] See Hume [6].

Even if there is agreement on the ranking of alternative policies by their fairness, individuals may still differ according to how strongly they feel about the issue of fairness. The analysis of the impact of such differences on social policies requires quantifying the intensity of the sense of fairness. In Karni and Safra [11] we developed measures of the intensity of individual sense of fairness. Here we illustrate the usefulness of these measures by applying them to the analysis of procedures intended to allocate an indivisible good among different claimants and to a stylized procedure of college admission.

There are many procedures by which social policies may be decided; *ceteris paribus* the outcome may depend on the particular procedure employed. We consider here a class of procedures characterized by the sole requirement that for a policy to be adopted it must be agreed upon by all interested parties. We model this agreement as the outcome of bargaining among different social groups with conflicting interests. Formally, we adopt the Nash bargaining model as our analytical framework and the Nash bargaining solution as our main analytical tool. According to this approach, a change of policy is justified if percentage-wise the utility gain from the change to one of the parties is larger than the percentage-wise utility loss to the others. A policy is chosen if no change is justified.[3] We show that in bilateral bargaining situations, other things being equal, an increase in the intensity of the sense of fairness of members of any group has the effect of making the Nash bargaining solution fairer according to the notion of fairness held by that group. Therefore, the effect of a more intense sense of fairness of members of any group on the well-being of members of the other group depends on its initial position. Judging by the notion of fairness of the members whose sense of fairness intensified, a group that was deprived of its fair share benefits and a group that enjoyed unfair privileges suffers.

The situation is more complicated in the case of a multilateral bargaining. When the intensity of the sense of fairness of members of a given group tends to infinity, the Nash

[3] For a discussion of bargaining as a social choice process, see Young [17].

bargaining solution tends to the most fair outcome according to the notion of fairness held by that group. However, this tendency is not always monotonic. There are situations in which, locally, an increase in the intensity of the sense of fairness of a given group makes the Nash bargaining solution less fair according to the notion of fairness held by members of that group. We present an example that demonstrates the feasibility of such situations.

In the next section we describe the analytical framework and the method by which college and university admission policies are embedded in this framework. A version of the Nash bargaining model applicable to our framework is developed in Section 3. In that section we also present our main results on the comparative statics effects of an increase in the intensity of the sense of fairness. Section 4 contains a discussion of the results and points out some of their implications and Section 5 concludes the paper.

2 The Model

2.1 Individual preferences and the intrinsic sense of fairness

The study of social choice when individuals possess an intrinsic sense of fairness was first undertaken in Karni [9]. There the context was the need to allocate, by lot, an indivisible good (or bad) between two claimants. The approach involved choosing a particular lot (a random allocation procedure) by which to determine who gets the good. Individuals are assumed to have preferences over random allocation procedures, reflecting their self-interest as well as an inherent concern for the fair treatment of others. Building upon this idea Karni and Safra [10] developed an axiomatic model of self-interest seeking moral individuals which is applied here to the analysis of social choice.

Let $N = \{1, ..., n\}, 2 < n < \infty$, be a set of individuals constituting a society that must choose a procedure by which to allocate, among its members, one unit of an indivisible good. Because the ex post allocations are necessarily unfair, the problem is to select a random allocation procedure that permits fairer ex ante treatment of the eligible individuals. Formally, let e^i, be the unit vector in $\mathbb{R}^n$ representing the ex post allocation in which individual i is assigned the good and denote by X the set of all ex post allocations (i.e., $X = \{e^i \mid 1 \leq i \leq n\}$). Let P be the $n-1$ dimensional simplex representing the set of all probability distributions on X. In the present context elements of P have the interpretation of random allocation procedures.

Individuals are characterized by two distinct binary relations on P: A preference relation $\succcurlyeq^i$, representing individual i's actual choice behavior and the fairness relation $\succcurlyeq^i_F$, representing individual i's moral value judgment. The relation $\succcurlyeq^i$ has the usual interpretation, namely, for any pair of allocation procedures $\mathbf{q}$ and $\mathbf{q}'$ in P, $\mathbf{q} \succcurlyeq^i \mathbf{q}'$ means that, if he were to choose between $\mathbf{q}$ and $\mathbf{q}'$, individual i would choose $\mathbf{q}$ or would be indifferent between the two. The relation $\succcurlyeq^i_F$ has the interpretation of 'being fairer than' and $\mathbf{q} \succcurlyeq^i_F \mathbf{q}'$ means that, according to individual i's moral value judgment, the allocation procedure $\mathbf{q}$ is at least as fair as the allocation procedure $\mathbf{q}'$. It is assumed that the sense of fairness is a moral sentiment that, jointly with concern for self-interest, governs the individual's choice behavior among random allocation procedures.

In Karni and Safra [10], we used the juxtaposition of the preference relation and the fairness relation to derive a new binary relation $\succcurlyeq^i_S$ on P representing the self-interest motive implicit in the individual choice behavior. Broadly speaking, an allocation procedure $\mathbf{q}$ is preferred over another allocation procedure $\mathbf{q}'$ from a *self-interest point of view* if the two allocation procedures are equally fair and $\mathbf{q}$ is preferred over $\mathbf{q}'$. We also show necessary and sufficient conditions under which the self-interest motive is represented by an affine function $\kappa^i : P \rightarrow \mathbb{R}$ (with a slight abuse of notations, we also use κ^i to denote the gradient of the affine function), the moral value judgment is represented by a strictly quasi-concave function $\sigma^i : P \rightarrow \mathbb{R}$, and the preference relation $\succcurlyeq^i$ is represented by a utility function $V^i : \kappa^i (P) \times \sigma^i (P) \rightarrow \mathbb{R}$. Thus, for all $\mathbf{q}, \mathbf{q}' \in P$, $\mathbf{q} \succcurlyeq^i \mathbf{q}'$ if and only if $V^i ((\kappa^i \cdot \mathbf{q}, \sigma^i (\mathbf{q})) \geq V^i ((\kappa^i \cdot \mathbf{q}', \sigma^i (\mathbf{q}'))$. In addition, we characterize the case in which function V^i is additively separable in the self-interest and fairness components. Formally,

$$V^i\left((\kappa^i\cdot\mathbf{q},\sigma^i\left(\mathbf{q}\right)\right)=h^i\left(\kappa^i\cdot\mathbf{q}\right)+\sigma^i\left(\mathbf{q}\right), \tag{1}$$

where h^i is a monotonic increasing function. This representation is unique up to positive cardinal unit-comparable transformation, namely, if $\left(\tilde{h}^i,\tilde{\kappa}^i,\tilde{\sigma}^i\right)$ represent $\succcurlyeq^i$ and is additively separable then $h^i\circ\kappa^i=c\tilde{h}^i\circ\tilde{\kappa}^i+a_\kappa$ and $\sigma^i=c\tilde{\sigma}^i+a_\sigma$, $c>0$.

We define *pure self-interest* as the case in which $\kappa^i=e^i$ (that is, each individual self-interest component depends solely on his own probability q_i of winning the good). We assume throughout that all individual preferences display pure self-interest.

In Karni and Safra [11] we developed measures that make it possible to compare the intensity of the sense of fairness of different individuals. In other words, we defined and characterized (on the set of all possible individuals) the relation of 'possessing a more intense sense of fairness' for the additive and nonadditive models. Such interpersonal comparisons require that the ordinal preferences and fairness relations of the individuals being compared be themselves comparable. Put differently, the preference-fairness relations pairs $(\succcurlyeq,\succcurlyeq_F)$ and $\left(\hat{\succcurlyeq},\hat{\succcurlyeq}_F\right)$ are *comparable* if they incorporate the same idea of fairness and induce the same self-interest relation (that is, if $\succcurlyeq_F=\hat{\succcurlyeq}_F$ and $\succcurlyeq_S=\hat{\succcurlyeq}_S$). For comparable preference-fairness relations with corresponding functional representations (h,κ,σ) and $\left(\hat{h},\kappa,\hat{\sigma}\right)$ the definition of $\left(\hat{\succcurlyeq},\hat{\succcurlyeq}_F\right)$ possessing a stronger sense of fairness than $(\succcurlyeq,\succcurlyeq_F)$ is given in Karni and Safra [11].[4] To simplify the exposition below, we assume that the preference-fairness relation pair $\left(\hat{\succcurlyeq},\hat{\succcurlyeq}_F\right)$ displays a more intense sense of fairness than the preference-fairness relation pair $(\succcurlyeq,\succcurlyeq_F)$ if $\hat{h}=h$ and $\hat{\sigma}=\lambda\sigma$, where $\lambda>1$. In this formulation, λ is a measure of the intensity of the sense of fairness. This formulation is somewhat less general than the measure developed in Karni and Safra [11], in the sense that the ratio f'/g' is constant and

[4] According to Karni and Safra [11], Theorem 3, $\left(\hat{\succcurlyeq},\hat{\succcurlyeq}_F\right)$ possessing a stronger sense of fairness than $(\succcurlyeq,\succcurlyeq_F)$ is equivalent to the existence of monotonic increasing functions f,g satisfying $h=f\circ\hat{h}$, $\sigma=g\circ\hat{\sigma}$ and $f'(\hat{h}(\kappa\cdot\mathbf{q}))\geq g'(\hat{\sigma}(\mathbf{q}))$.

equal to λ. In the context of the present work this entails no essential loss of generality. To see this observe that the local nature of the Theorem below implies that all the arguments involving changes in λ can be reproduced for the general case by changing this ratio pointwise, in the same direction.

Note that if all individuals subscribe to the same moral value judgment, then this moral value judgment may be interpreted as a criterion for decision making from behind a veil of ignorance. In this case the fairness relation is analogous to Harsanyi's [4] concept of social preference relation and to his concept of preference relation of an impartial observer (Harsanyi, [3] and [5]). Even if the moral value judgment is common to all groups, the intensity of the sense of fairness may still vary among them. In view of this observation, it is worth emphasizing that, unlike Harsanyi's purely normative approach, we are taking a positive approach to social choice. According to our approach, moral considerations are combined with self-interests to produce outcomes representing the resolution of conflicting interests.

2.2 Group preferences

In what follows we analyze bargaining in a society with a given social structure. We assume that individuals belonging to the same social group have the same preference-fairness relations over some relevant subset of the set of allocation procedures, whose interpretation depends on the particular problem at hand. Formally, let $\{N_j\}_{j=1}^{m}$ be a partition of N , where each N_j represents a distinct social group consisting of n_j individuals. Let $\Psi = \{\mathbf{q} \in P | \, i, k \in N_j \Rightarrow q_i = q_k\}$ be the set of social procedures that do not distinguish among individuals belonging to the same group. We assume that members of the same group have the same preferences on Ψ. In other words, for all $j = 1, ..., m$ and for all $i, k \in N_j$, $\succcurlyeq^i = \succcurlyeq^k$ and $\succcurlyeq^i_F = \succcurlyeq^k_F$ on Ψ. As in this case $q_i = q_k$, the pure self-interest components of individuals belonging to the same group, restricted to Ψ, are the same. Note that the assumption that individuals belonging to the same social group have identical fairness relation and identical preference relation means that they are comparable in the sense of Karni and Safra [11]). It is possible, therefore, to perform a comparative statics analysis of the effect of increasing sense of fairness of members of the same social group using the measures of intensity of the sense of fairness developed by Karni and Safra.

We identify Ψ with the $m-1$ dimensional simplex Δ by using the isomorphism $\delta : \Psi \to \Delta$ satisfying $\mathbf{p} = \delta(\mathbf{q})$ if, for $j \in \{1, ..., m\}$, $p_j = \sum_{k \in N_j} q_k = n_j q_i$ for some $i \in N_j$. Clearly, $\mathbf{q} = \delta^{-1}(\mathbf{p})$ if $q_i = \frac{p_j}{n_j}$, for $i \in N_j$. Taking into account the assumptions made in the preceding subsection, individual i's preference relation on Δ is represented by

$$U^i(\mathbf{p}) = V^i(\delta^{-1}(\mathbf{p})) = h^i\left(\frac{p_j}{n_j}\right) + \lambda^i \sigma^i\left(\delta^{-1}(\mathbf{p})\right) \tag{2}$$

where $i \in N_j$ and λ^i is the parameter representing the intensity of the sense of fairness. By construction, $U^i = U^k$ (and $\lambda^i = \lambda^k$) whenever $i, k \in N_j$. Henceforth we use the index j to denote the various components of the utility functions of individuals belonging to group N_j.

For convenience we assume, henceforth, that σ^j is non-positive and that its maximal value in Δ is zero. Moreover, since subsequently we invoke the Nash bargaining solution, we assume that for each j both h^j and σ^j are concave functions. This assumption plays, in our model, a role analogous to that of risk-aversion in the original Nash bargaining model. We assume further that the functions h^j and σ^j are differentiable.

2.3 Example: College admission policies

Let $N = \{1, ..., n\}$ be a society consisting of a finite and large number of individuals where, as above, n_j denotes the number of individuals belonging to the social group N_j. Consider an institution, for instance a college, that has a limited number b ($b < n$) of openings for new students. Assume that there exists a test whose score is positively correlated with the students' college performance. Each social group is characterized by a distribution function F_j over $[0, 1]$, the range of the possible test scores. Suppose that, for reasons to be discussed below, the distributions of distinct groups are different. A *feasible admission policy* is an m-tuple $s = (s_1, ..., s_m) \in [0,1]^m$, where s_j is the cutoff score for admission of members of group N_j, that satisfies the feasibility constraint $\sum_{j=1}^{m} n_j (1 - F_j(s_j)) = b$.[5] The set of all feasible admission policies is denoted by $S(b)$. Admission policies must be decided upon *ex-ante*, namely, before individuals have a chance to observe their test scores and when they are indistinguishable from other members of their group.

Next, using the preceding construction, we define the subset $\Psi \subset P$ of personal probabilities. With each s we associate a vector of probabilities $\mathbf{q}(s) \in \mathbb{R}^n$: for $i \in N_j$, let $\pi_i(s) = (1 - F_j(s_j))$ and define $q_i(s) = \pi_i(s) / \sum_{k \in N} \pi_k(s)$. The probability q^i is interpreted as the (ex ante) probability that individual i of group N^j be included in a random draw (of one person) from the student population. Note that the denominator is equal to b. Ψ is the set of all these vectors. In this way $S(b)$ is embedded in P and, as in the general analysis of subsection 2.2, is then identified with the set Δ using the transformation δ. The utilities U^j are defined as in equation (2).[6]

Affirmative action policies are intended to achieve greater parity of opportunities. If the disparity in the performance of members of different social groups, and hence their opportunities, is the result of discrimination, affirmative action policies that apply different performance thresholds for college admission to different social groups are perceived as just. Redressing past injustice, however, is not the sole moral imperative that may figure in the design of college admission policy. The merit system, which imposes a uniform (nondiscriminatory) admission standard based on performance, amounts to equal treatment of all candidates. The merit system represents a competing moral value judgment that may be

[5] We assume that n is sufficiently large to ensure that the probabilities are close to the empirical distribution. We only consider equality since preferences are later assumed to increase with admission probabilities.

[6] The transformation from $S(b)$ to Δ is one-to-one and onto. This implies that no information is lost by the transformation. Hence any preference relation on $S(b)$ can be represented by a corresponding preference relation on Δ, the set of agreements in the bargaining problem to be discussed in section 3 below.

applied to the design of college admission policy. Formally, the *merit policy* s^m is an admission policy satisfying $s_k^m = s_j^m = r$, for all k, j. If college performance is positively correlated with the social value-added of higher education then s^m is socially efficient. The merit policy induces the admission probability vector $\mathbf{p}^m = \delta(\mathbf{q}(s^m))$. Similarly, the *proportional representation policy* s^{pr} is the admission policy satisfying $\pi_j\left(s_j^{pr}\right) = \frac{b}{n}$, for all j, with the induced admission probability vector $\mathbf{p}^{pr} = \delta(\mathbf{q}(s^{pr}))$. A fairness relation may take into consideration both the efficiency and the equality of opportunity. In particular, if the differences in the test scores among the groups is a manifestation of unequal opportunities then the moral value judgments may involve, in addition to consideration of efficiency, the need to redress past injustice. For instance, moral value judgment involving trade off between these two components may be represented by σ that assumes the following functional form:

$$\sigma(\mathbf{q}) = \ell(d(\mathbf{q}, \mathbf{q}(s^m))) + \ell(d(\mathbf{q}, \mathbf{q}(s^{pr})))$$

where d is the Euclidean metric and ℓ is a monotonic decreasing and concave real-valued function. Note, however, that our approach is general and can accommodate other concepts of fairness pertaining to college admission policies.

3 Social Choice as a Nash Bargaining Solution

3.1 The bargaining model

The analysis that follows is based on the premise that social policy is ultimately the outcome of bargaining among the groups involved. To model the situation we apply the Nash bargaining solution adopted to the analytical framework of the preceding section. The Nash bargaining solution is based on the assumption that the utilities of the bargainers are unique up to positive linear transformation. This is justified if the bargainers preference relations over the set of lotteries on the agreed upon payoffs are linear in the probabilities. In our framework, the agreements themselves, for instance, the admission policies, are identified with lotteries on random allocations procedures of the available slots. However, the utility assigned to the random allocation procedures is not linear in the probabilities that figure in the design of these procedures. In particular, it is reasonable to suppose that representation of the fairness relation is quasi-concave in the probabilities. Thus, while we do not assume that the utility function representing the individual attitudes toward random allocation procedures are linear in the probabilities of these procedures, we nevertheless assume that the utility functions are unique up to positive linear transformations. We justify this assumption by adopting the approach to bargaining due to Rubinstein, Safra and Thomson [16] which, as we explain next, is natural for the problem at hand. We note in passing, that while other approaches to bargaining solutions may be used to model the resolution of social conflict and the Nash bargaining solution has an appealing normative feature, namely, with appropriate natural scale factors, it is the unique solution implementing, simultaneously, the egalitarian and utilitarian solutions (see discussion in Myerson [14] 8.3).

Let Δ be the set of all possible agreements and denote by $\mathbf{d}$ the disagreement point. Extend the domain of U^j to $\Delta \cup \{\mathbf{d}\}$.[7] Anticipating the analysis that follows, the interpretation of the disagreement point requires some care. We are concerned with situations in which a more intense sense of fairness would lead social groups to reject agreements that they deem to be unfair even at the cost of disagreement. In the limit, when the sense of fairness is infinitely strong, all but the fairest policy are rejected. Note that in this case, if there is no consensus regarding what constitutes the fairest policy, the solution will be disagreement. In view of this consideration the interpretation of the disagreement depends on the specifics of the problem at hand, which may include the institutional and legal environment. For example, if the bargaining is over the procedure of allocating an indivisible good, a disagreement may mean that nobody gets the good. The analogous result in the

[7] Note that the concavity of $h^j \circ \delta^{-1}$ and $\sigma^j \circ \delta^{-1}$ on Δ follows from the concavity of h^j and σ^j on Ψ.

example of college admission policies is that if no agreement is reached the college will suffer a cut-off of funding forcing it to close. In this case, the probability of admission of all groups is zero and the disagreement point is the origin of $\mathbb{R}^m$. We assume that, given the absence of resources, this corresponds to the fairest treatment of the different groups. We discuss the more general case in Section 4.2. As commonly assumed in bargaining models, we suppose that for every j there exists some random allocation procedure in Δ that is indifferent to the disagreement point and that there exist some random allocation procedures that are strictly preferred over the disagreement point by all groups.

A breakdown risk is a pair $(\mathbf{p};\alpha) \in \Delta \times [0,1] := B$, where α denotes the probability that the bargaining process will end with an agreement $\mathbf{p}$; otherwise, with a probability $(1-\alpha)$ it will end with disagreement. We extend the choice set to include breakdown risks and suppose that each player's preference relation, $\succsim^j$, is extended to the set B by the following homogeneity axiom of Rubinstein, Safra, and Thomson [16].

Homogeneity: For all $\alpha, \alpha', \gamma \in [0,1]$ and $\mathbf{p}, \mathbf{q} \in\Delta$, if $(\mathbf{p};\alpha) \succsim^j (\mathbf{q};\alpha')$ then $(\mathbf{p};\gamma\alpha) \succsim^j (\mathbf{q};\gamma\alpha')$.

Consequently, the utility function U^j in equation (2) can be extended to B by:

$$U^j(\mathbf{p};\alpha) = \alpha U^j(\mathbf{p}) + (1-\alpha) U^j(\mathbf{d}) \tag{3}$$

$U^j(\mathbf{p})$ is as in (2) and $U^j(\mathbf{d})$ is arbitrarily chosen.

Of particular interest is the case in which $\mathbf{d}$ is as fair as the fairest random allocation procedure, according to group N_j's fairness relation. Hence, by assumption, $\sigma^j(\mathbf{d}) = 0$. This implies that the utility of the disagreement point is independent of λ^j, the parameter representing the intensity of the sense of fairness. Invoking the assumption that $\sigma^j \leq 0$, the set of individually rational agreements shrinks (towards the most fair point of group j) as λ^j increases (that is, for $\lambda^j > \bar{\lambda}^j$, $\{\mathbf{p} \in \Delta \mid h^j\left(\frac{p_j}{n_j}\right) + \lambda^j\sigma^j(\delta^{-1}(\mathbf{p})) - U^j(\mathbf{d}) \geq 0\} \subset \{\mathbf{p} \in \Delta \mid h^j\left(\frac{p_j}{n_j}\right) + \bar{\lambda}^j\sigma^j(\delta^{-1}(\mathbf{p})) - U^j(\mathbf{d}) \geq 0\}$).

The idea that a stronger sense of fairness makes the disagreement point relatively more attractive is reminiscent of certain explanations of the results of experiments with ultimatum games. In these games, a fixed amount of money has to be divided between two players. One player proposes a division, which the second player must either accept or reject. If the proposal is accepted, the game is terminated and the money is paid out according to the proposed division. If the proposal is rejected, the game is terminated and the two players get nothing. In many experiments it turns out that the proposer offers the responder a substantial part of the sum to be divided, and in some experiments divisions that left the responder with small fraction of the total amount were rejected. One explanation of these observations is that individuals have a sense of fairness and are willing to reject what they consider to be grossly unfair divisions, namely, enforce a disagreement, even at a cost to themselves (see Camerer, [2]). Extension of this argument leads to the conclusion that proposed divisions that are acceptable to some responders will be rejected by responders who have stronger sense of fairness, suggesting that it makes the disagreement point relatively more attractive.

Under the assumption that, for all j, h^j and σ^j are strictly concave functions our bargaining problems are 'well behaved', that is, the image of B in the utilities space is convex. Thus our model conforms to a Nash bargaining structure and the (n-person) symmetric Nash bargaining solution, $N(B, \mathbf{d})$, is defined by

$$N(B, \mathbf{d}) = \arg\max\{\prod_{j=1}^{m} \left(U^j(\mathbf{p}; \alpha) - U^j(\mathbf{d})\right)^{n_j} \mid U^j(\mathbf{p}; \alpha) - U^j(\mathbf{d}) \geq 0,\ j = 1, ..., m\} \quad (4)$$

Our assumptions imply that the solution is unique. Clearly, it is attained for $\alpha = 1$. Notice that we model a bargaining problem involving n individuals. While in equation (4) the exponent n_j may suggest that individuals belonging to larger social groups possess greater bargaining power, in fact their individual probabilities of obtaining the good may decline with the size of the group. It is also worth noting that the symmetric treatment, implicit in our formulation, serves to simplify the exposition and may be generalized to include asymmetries.

3.2 The power of moral conviction

We examine next the implications of variations in the intensity of the sense of fairness on social policies. A concern for fairness changes the parameters of the acceptance set of possible agreements and the solution of the bargaining problem. Moreover, the implications of a heightened sense of fairness for the choice of random allocation policies depend on the nature of the social decision-making procedure.

Because all known bargaining solutions pick Pareto optimal outcomes (e.g., the Nash and the Kalai-Smorodinsky [8] solutions), when the intensity of the sense of fairness of members of a given group tends to infinity the solution tends to the fairest outcome according to the notion of fairness held by that group. This is an immediate implication of the shrinkage of the set of individually rational outcomes towards the fairest outcome of that group. However, this tendency of the Nash bargaining solution to shift towards the fairest outcome is not everywhere monotonic. There exist situations in which, locally, such an increase in the intensity of the sense of fairness makes the Nash bargaining solution less fair according to the notion of fairness held by members of that group.

To understand these effects, it is best to start by considering bargaining between two social groups, where the influences are most transparent. In Section 3.2.1 we show that, in this case, an increase in the sense of fairness of all individuals belonging to a given group always increases the degree of fairness of the Nash solution according to the notion of fairness held by members of this group. In Section 3.2.2 we discuss the case of bargaining among more than two groups. We present there an example which demonstrates that, locally, an increase in the sense of fairness of all individuals belonging to a given group decreases the degree of fairness of the Nash solution according to the notion of fairness held by members of this group.

3.2.1 The bilateral case

In this case there are two social groups, consisting of n_1 and n_2 members, respectively, and the set Δ is therefore one-dimensional simplex. To simplify the exposition we denote points in Δ, like $(p, 1-p)$, by p and write $h^i\left(\frac{p}{n_i}\right)$ for short. Note that this change of notation means that h^1 increases with p while h^2 decreases with p.[8] Using the remaining degree of freedom in the utility representation, we normalize the utility functions so that $U^j(\mathbf{d}) = 0$, for all j. The Nash bargaining solution is therefore the solution of

$$\underset{\{p:U^1(p),U^2(p)\geq 0\}}{\arg\max} \left[h^1\left(\frac{p}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p)\right)\right]^{n_1}\left[h^2\left(\frac{p}{n_2}\right) + \lambda^2\sigma^2\left(\delta^{-1}(p)\right)\right]^{n_2} \tag{5}$$

and the necessary and sufficient condition for maximum is:

$$\frac{n_1\left[\frac{\mathrm{d}}{\mathrm{d}p}h^1\left(\frac{p}{n_1}\right) + \lambda^1\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right)\right]}{h^1\left(\frac{p}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p)\right)} = -\frac{n_2\left[\frac{\mathrm{d}}{\mathrm{d}p}h^2\left(\frac{p}{n_2}\right) + \lambda^2\frac{\mathrm{d}}{\mathrm{d}p}\sigma^2\left(\delta^{-1}(p)\right)\right]}{h^2\left(\frac{p}{n_2}\right) + \lambda^2\sigma^2\left(\delta^{-1}(p)\right)}. \tag{6}$$

Following Aumann and Kurz [1] we define the *boldness functions* $b^j : \Delta \times \mathbb{R}_+ \to \mathbb{R}$, $j = 1, 2$ such that b^1 is the left-hand side, and b^2 is the right-hand side, of equation (6). The numerator of a boldness function is the marginal gain to group j from pushing for a more favorable solution and the denominator is the potential loss of such a push since it may result in disagreement. By concavity, b^1 decreases and b^2 increases with respect to p.

Consider next the implications of an increase in the intensity of the sense of fairness of all individuals belonging to one of the social groups.

Theorem *If $\sigma^j(\mathbf{d}) = 0$ then an increase in the intensity of the sense of fairness of members of group j implies that the policy corresponding to the Nash bargaining solution is fairer according to the notion of fairness held by members of this group.*

[8] More generally, elements of Δ may be normalized such that $p_m = 1 - \Sigma_{i=1}^{m-1} p_i$ and all derivatives are taken with respect to the first $m-1$ variables. Under this normalization, the functions σ^i and $h^i \circ \kappa^i$ are defined over the projection of Δ over $\mathbb{R}^{m-1}$. For an elaborate discussion of issues involve in defining 'probability derivatives' see Machina [13]. Note, however, that we choose a different approach for the multivariate case.

The case in which $\sigma^j(\mathbf{d}) \neq 0$ is discussed in section 4.2 below. Note that under the Theorem's hypothesis, if the two groups have a common notion of fairness, then an increase in the intensity of the sense of fairness of either or both groups results in a fairer policy.

Proof. We show that an increase in the intensity of the sense of fairness of members of group 1 increases the fairness of the policy under the Nash bargaining solution.

Differentiating the boldness function of members of group 1 with respect to λ^1 we obtain:

$$\frac{\partial}{\partial\lambda^1} b^1\left(p, \lambda^1\right) = \frac{n_1\left[h^1\left(\frac{p}{n_1}\right)\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right) - \sigma^1\left(\delta^{-1}(p)\right)\frac{\mathrm{d}}{\mathrm{d}p}h^1\left(\frac{p}{n_1}\right)\right]}{\left[h^1\left(\frac{p}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p)\right)\right]^2} \tag{7}$$

Clearly, the sign of this derivative is determined by the sign of the numerator. We show next that

$$h^1\left(\frac{p}{n_1}\right)\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right) - \sigma^1\left(\delta^{-1}(p)\right)\frac{\mathrm{d}}{\mathrm{d}p}h^1\left(\frac{p}{n_1}\right) \gtreqless 0 \iff p \lesseqgtr p^F \tag{8}$$

where p^F is the fairest point according to group 1.

Consider $p < p^F$. By the concavity of σ^1, $\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right) > 0$. This, together with the positivity and monotonicity of $h^1\left(\frac{p}{n_1}\right)$ (that follows from $U^1\left(p\right) \geq 0$), implies that $h^1\left(\frac{p}{n_1}\right)\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right) - \sigma^1\left(\delta^{-1}(p)\right)\frac{\mathrm{d}}{\mathrm{d}p}h^1\left(\frac{p}{n_1}\right) > 0$.

If $p > p^F$ then $\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right) < 0$. By $U^1\left(p\right) \geq 0$, for λ^1 sufficiently close to 0, $h^1\left(\frac{p}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p)\right) > 0$. Since $\sigma^1\left(\delta^{-1}(p)\right) < 0$, for λ^1 sufficiently large $h^1\left(\frac{p}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p)\right) < 0$. Moreover, for every finite λ^1,

$$h^1\left(\frac{p^F}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p^F)\right) \geq 0 \text{ and } \frac{\mathrm{d}}{\mathrm{d}p}\left[h^1\left(\frac{p}{n_1}\right) + \lambda^1\sigma^1\left(\delta^{-1}(p)\right)\right]\Bigg|_{p=p^F} > 0. \tag{9}$$

Define $\bar{\lambda}^1\left(p\right)$ by the equation $h^1\left(\frac{p}{n_1}\right) + \bar{\lambda}^1\left(p\right)\sigma^1\left(\delta^{-1}(p)\right) = 0$. The concavity of the utility $h^1 + \bar{\lambda}\sigma^1$ implies that $\frac{\mathrm{d}}{\mathrm{d}p}\left[h^1\left(\frac{p}{n_1}\right) + \bar{\lambda}^1\sigma^1\left(\delta^{-1}(p)\right)\right] < 0$ at p. Hence

$$\frac{h^1\left(\frac{p}{n_1}\right)}{\sigma^1\left(\delta^{-1}(p)\right)} = -\bar{\lambda}^1\left(p\right) < \frac{\frac{\mathrm{d}}{\mathrm{d}p}h^1\left(\frac{p}{n_1}\right)}{\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right)}. \tag{10}$$

Therefore $h^1\left(\frac{p}{n_1}\right)\frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right)-\sigma^1\left(\delta^{-1}(p)\right)\frac{\mathrm{d}}{\mathrm{d}p}h^1\left(\frac{p}{n_1}\right)<0$.

An increase in the intensity of the sense of fairness of members of group 1, amounts to an increase in λ^1. By equation (8), this rotates the graph of the boldness function of members of group 1 around p^F (this is the situation depicted in Figure 1).

Let p^N be the initial Nash bargaining solution. If members of group 1 are treated favorably relative to their fairest procedure (that is, $p^F < p^N$), then the new Nash solution is at a smaller value of p, hence closer to p^F (see Figure 1).

If the initial Nash bargaining solution p^N is a policy in which members of group 1 are treated unfavorably relative to the fairest procedure (that is, $p^F > p^N$), then the new Nash solution is at a larger value of p, hence closer to p^F. ■

3.2.2 The multilateral case

Unlike in the bilateral bargaining case, in the multilateral case it is not true that an increase in the intensity of the sense of fairness of members of any group always implies that the policy under the Nash bargaining solution is fairer according to the notion of fairness held by members of that group. In the example below we examine the effect of an increase in the sense of fairness of one group when initially none of the groups displays any concern for fairness. As above, utility functions are normalized so that $U^j(\mathbf{d})=0$, for all j.

Example: greater intensity of the sense of fairness causing a decrease in fairness of the policy.

Let $m=3$ and consider the problem:

$$\max_{\{\mathbf{p}\in\Delta:U^j(\mathbf{p})\geq 0,\, j=1,2,3\}}\sum_{j=1}^{3}n_j\log\left[h^j\left(\frac{p_j}{n_j}\right)+\lambda^j\sigma^j\left(\delta^{-1}(\mathbf{p})\right)\right]\quad\text{subject to }\ 1-\sum_{j=1}^{3}p_j=0$$

Suppose that $\lambda^j=0$, $j=1,2,3$ and let μ be the Lagrange multiplier. Then the necessary

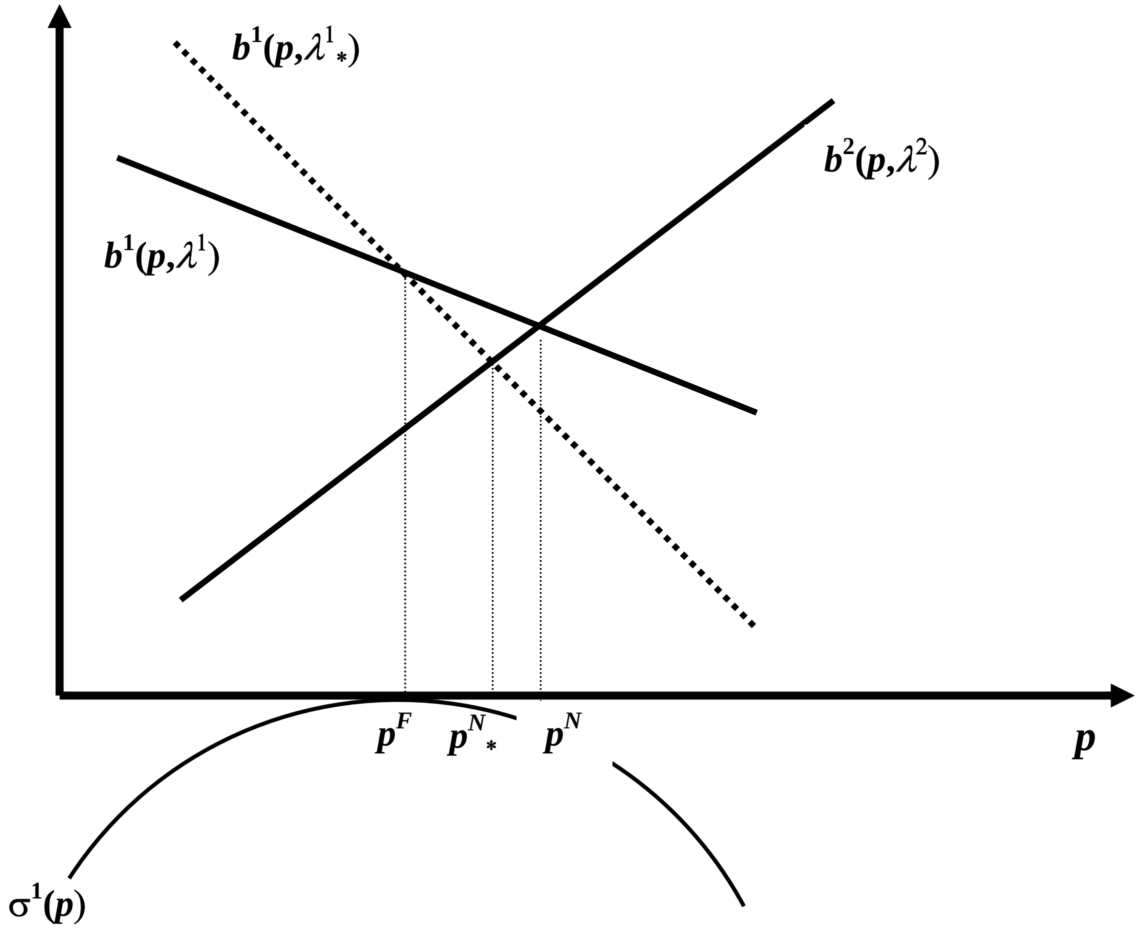

Figure 1: An increase of λ^1 to λ^1_* rotates the graph of b^1 around p^F and shifts the solution from p^N to p^N_*

and sufficient conditions are

$$\frac{n_j \frac{\mathrm{d}h^j}{\mathrm{d}p_j}\left(\frac{p_j}{n_j}\right)}{h^j\left(\frac{p_j}{n_j}\right)} - \mu = 0, \quad j = 1, 2, 3 \tag{11}$$

and

$$1 - \sum_{j=1}^{3} p_j = 0 \tag{12}$$

The comparative statics effects of an increase in λ^1 at $\lambda^j = 0$, $j = 1, 2, 3$ are obtained from the solution of the following system of equations:

$$\begin{pmatrix} W_1 & 0 & 0 & -1 \\ 0 & W_2 & 0 & -1 \\ 0 & 0 & W_3 & -1 \\ -1 & -1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}p_1}{\mathrm{d}\lambda^1} \\ \frac{\mathrm{d}p_2}{\mathrm{d}\lambda^1} \\ \frac{\mathrm{d}p_3}{\mathrm{d}\lambda^1} \\ \frac{\mathrm{d}\mu}{\mathrm{d}\lambda^1} \end{pmatrix} = \begin{pmatrix} -X \\ -Y \\ -Z \\ 0 \end{pmatrix} \tag{13}$$

where

$$W_j = n_j \frac{h^j\left(\frac{p_j}{n_j}\right) \frac{\mathrm{d}^2}{\mathrm{d}p_j^2} h^j\left(\frac{p_j}{n_j}\right) - \left(\frac{\mathrm{d}}{\mathrm{d}p_j} h^j\left(\frac{p_j}{n_j}\right)\right)^2}{\left(h^j\left(\frac{p_j}{n_j}\right)\right)^2}, \ j = 1, 2, 3 \tag{14}$$

and

$$X = n_1 \frac{h^1\left(\frac{p_1}{n_1}\right) \frac{\mathrm{d}}{\mathrm{d}p_1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_1} h^1\left(\frac{p_1}{n_1}\right)\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{\left(h^1\left(\frac{p_1}{n_1}\right)\right)^2}, Y = \frac{n_1 \frac{\mathrm{d}}{\mathrm{d}p_2}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{h^1\left(\frac{p_1}{n_1}\right)}, Z = \frac{n_1 \frac{\mathrm{d}}{\mathrm{d}p_3}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{h^1\left(\frac{p_1}{n_1}\right)} \tag{15}$$

Let D be the determinant of the bordered Hessian matrix in (13). Then solving equations (13) we obtain:

$$\frac{\mathrm{d}p_1}{\mathrm{d}\lambda^1} = \frac{1}{D}[X(W_2 + W_3) - YW_3 - ZW_2] \tag{16}$$

$$\frac{\mathrm{d}p_2}{\mathrm{d}\lambda^1} = \frac{1}{D}[-XW_3 + Y(W_1 + W_3) - ZW_1] \tag{17}$$

and

$$\frac{\mathrm{d}p_3}{\mathrm{d}\lambda^1} = \frac{1}{D}[-XW_2 - YW_1 + Z(W_1 + W_2)] \tag{18}$$

Thus

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\lambda^1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) = \sum_{j=1}^{3} \frac{\mathrm{d}}{\mathrm{d}p_j}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)\frac{\mathrm{d}p_j}{\mathrm{d}\lambda^1} \qquad (19)\\
&= \frac{1}{D}\left\{ X\left[W_2\left(\frac{\mathrm{d}}{\mathrm{d}p_1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_3}\sigma^1\left(\mathbf{p}\right)\right) + W_3\left(\frac{\mathrm{d}}{\mathrm{d}p_1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_2}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)\right)\right]\right.\\
&\quad +Y\left[W_1\left(\frac{\mathrm{d}}{\mathrm{d}p_2}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_3}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)\right) + W_3\left(\frac{\mathrm{d}}{\mathrm{d}p_2}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)\right)\right]\\
&\left. + Z\left[W_1\left(\frac{\mathrm{d}}{\mathrm{d}p_3}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_2}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)\right) + W_2\left(\frac{\mathrm{d}}{\mathrm{d}p_3}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) - \frac{\mathrm{d}}{\mathrm{d}p_1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)\right)\right]\right\}
\end{aligned}
$$

Let $\sigma^1\left(\delta^{-1}(\mathbf{p})\right) = -f\left(d\left(\mathbf{p}\right)\right)$ where $d\left(\mathbf{p}\right) = \sqrt{\sum_{j=1}^{3}\left(p_j - \frac{1}{3}\right)^2}$ is the Euclidean distance between $\mathbf{p}$ and $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and f is a monotonic increasing and strictly concave function satisfying $f\left(0\right) = 0$, $f\left(d\left(\frac{15}{30}, \frac{11}{30}, \frac{4}{30}\right)\right) = 50$ and $f'\left(d\left(\frac{15}{30}, \frac{11}{30}, \frac{4}{30}\right)\right) = 1$. Assume that $h^j\left(\frac{p_j}{n_j}\right) = p_j$ and $\lambda^j = 0$. Then, for $j, k = 1, 2, 3$,

$$
\frac{\mathrm{d}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{\mathrm{d}p_j} = f'\left(d\left(\mathbf{p}\right)\right)\frac{p_j - \frac{1}{3}}{-d\left(\mathbf{p}\right)},\quad \frac{\mathrm{d}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{\mathrm{d}p_j} - \frac{\mathrm{d}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{\mathrm{d}p_k} = f'\left(d\left(\mathbf{p}\right)\right)\frac{p_j - p_k}{-d\left(\mathbf{p}\right)},\quad W_j = -\frac{n_j}{p_j^2} \qquad (20)
$$

Moreover, since by (11) $n_j/p_j = \mu$, then (15) becomes

$$
X = \mu\left(f'\left(d\left(\mathbf{p}\right)\right)\frac{p_1 - \frac{1}{3}}{-d\left(\mathbf{p}\right)} - \frac{\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{p_1}\right),\ Y = \mu f'\left(d\left(\mathbf{p}\right)\right)\frac{p_2 - \frac{1}{3}}{-d\left(\mathbf{p}\right)},\ Z = \mu f'\left(d\left(\mathbf{p}\right)\right)\frac{p_3 - \frac{1}{3}}{-d\left(\mathbf{p}\right)} \qquad (21)
$$

Hence

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda^1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right) \quad = \quad & \frac{\mu^2 f'\left(d\left(\mathbf{p}\right)\right)}{d\left(\mathbf{p}\right)D}\{\left\{\left(f'\left(d\left(\mathbf{p}\right)\right)\frac{\left(p_1 - \frac{1}{3}\right)}{-d\left(\mathbf{p}\right)} - \frac{\sigma^1\left(\delta^{-1}(\mathbf{p})\right)}{p_1}\right)\right.\left(\frac{p_1 - p_3}{p_2} + \frac{p_1 - p_2}{p_3}\right) \qquad (22)\\
& + f'\left(d\left(\mathbf{p}\right)\right)\frac{p_2 - \frac{1}{3}}{-d\left(\mathbf{p}\right)}\left(\frac{p_2 - p_3}{p_1} + \frac{p_2 - p_1}{p_3}\right)\\
& \left. + f'\left(d\left(\mathbf{p}\right)\right)\frac{p_3 - \frac{1}{3}}{-d\left(\mathbf{p}\right)}\left(\frac{p_3 - p_2}{p_1} + \frac{p_3 - p_1}{p_2}\right)\right\}
\end{aligned}
$$

Suppose that $n_1 = 150, n_2 = 110$ and $n_3 = 40$ and consider $\mathbf{p} = \left(\frac{15}{30}, \frac{11}{30}, \frac{4}{30}\right)$ that satisfies equations (11) with $\mu = 300$. Then $d\left(\frac{15}{30}, \frac{11}{30}, \frac{4}{30}\right) = \frac{7.87}{30}$ and $\mathrm{sign}\mu^2 f'\left(d\left(\mathbf{p}\right)\right)/d\left(\mathbf{p}\right)D =$

$\mathrm{sign} D = (-1)^3 < 0$. Thus the sign of $\frac{\mathrm{d}}{\mathrm{d}\lambda^1}\sigma^1(\mathbf{p})$ is opposite to that of the expression in the curly brackets in equation (22). This expression is equal to

$$\left(-\frac{5}{7.87}+\frac{1500}{15}\right)(1+1)-\frac{1}{7.87}\left(\frac{7}{15}-1\right)+\frac{6}{7.87}\left(-\frac{7}{15}-1\right)=197.67$$

Hence

$$\frac{\mathrm{d}}{\mathrm{d}\lambda^1}\sigma^1\left(\delta^{-1}(\mathbf{p})\right)<0 \tag{23}$$

and the level of fairness according to group 1 decreases as their intensity of the sense of fairness increases. ■

It is easy to verify, using equation (16), that $\frac{\mathrm{d}}{\mathrm{d}\lambda^1}p_1 > 0$. In other words, the winning probability of group 1 under the Nash bargaining solution increases when the sense of fairness of group 1 increases.

4 Discussion

4.1 Different notions of fairness

The idea that in order to form moral value judgments individuals must conceive themselves as having to choose among policies or institutions from behind a veil of ignorance is philosophically compelling. However, its application requires that individuals be capable of detaching themselves from their own individual circumstances, including their personal histories and preferences, when contemplating choices among policies or institutions. This requirement is, in general, difficult if not impossible to meet. It seems reasonable to suppose that the idea of fairness itself varies among groups, reflecting the group's experience and sensitivities. For example, it would not be surprising if, in the United States, the concept of fairness held by African-Americans is distinct from that held by whites even if members of both races try to set aside their immediate interests. Our model is designed to accommodate situations in which different social groups entertain distinct notions of fairness. To grasp this, consider again the bilateral case and suppose that group 1 adheres to the concept of fairness embodied in the merit system and group 2 regards the proximity to proportional representation as the appropriated measure of fairness. Formally, let $\sigma^1(\mathbf{q}) = \ell(d(\mathbf{q}, \mathbf{q}(s^m)))$ and $\sigma^2(\mathbf{q}) = \ell(d(\mathbf{q}, \mathbf{q}(s^{pr})))$ where $\ell(\cdot)$ is a non-positive, concave function and $\ell(0) = 0$. Following the analysis of the preceding section it is clear that an increase in the sense of fairness of members of any group results in a shift in the Nash bargaining solution closer to what the group regards as the fairest policy. Obviously, the effect of an increase in the intensity of the sense of fairness of all groups is ambiguous.

4.2 On the significance of the disagreement point

Thus far we assumed that the disagreement point corresponds to the fairest treatment of the different groups. If the disagreement point does not correspond to the fairest point and is not Pareto optimal then $\sigma^j(\mathbf{d}) < 0$ for some group j. In this case, $\frac{\mathrm{d}}{\mathrm{d}\lambda^j} U^j(\mathbf{d}) \neq 0$ and the values of the functions at $\mathbf{d}$ must be taken into account. To grasp the significance of this change, consider the bilateral bargaining case. Let $(d^1, 1-d^1)$ be the point in $\{(p, (1-p)) \mid p \in [0,1]\}$ that is indifferent, from the point of view of group 1, to the disagreement point $\mathbf{d}$ and assume that $\sigma^1\left(\delta^{-1}(p^F)\right) - \sigma^1\left(\delta^{-1}(d^1)\right) > 0$. The boldness function of group 1 is given by:

$$b^1\left(p, \lambda^1\right) = \frac{n_1[\frac{\mathrm{d}}{\mathrm{d}p} h^1\left(\frac{p}{n_1}\right) + \lambda^1 \frac{\mathrm{d}}{\mathrm{d}p} \sigma^1\left(\delta^{-1}(p)\right)]}{h^1\left(\frac{p}{n_1}\right) + \lambda^1\left[\sigma^1\left(\delta^{-1}(p)\right) - \sigma^1\left(\delta^{-1}(d^1)\right)\right]} \tag{24}$$

where we normalized h^1 to satisfy $h^1\left(\frac{d^1}{n_1}\right) = 0$. The boldness function of the other group, b^2, is similarly defined. As before, our assumptions imply that b^1 decreases and b^2 increases, with respect to p.

Next consider the derivative of b^1 with respect to λ^1

$$\frac{\mathrm{d}}{\mathrm{d}\lambda^1} b^1\left(p, \lambda^1\right) = \frac{h^1\left(\frac{p}{n_1}\right) \frac{\mathrm{d}}{\mathrm{d}p}\sigma^1\left(\delta^{-1}(p)\right) - \frac{\mathrm{d}}{\mathrm{d}p} h^1\left(\frac{p}{n_1}\right)\left[\sigma^1\left(\delta^{-1}(p)\right) - \sigma^1\left(\delta^{-1}(d^1)\right)\right]}{\left\{h^1\left(\frac{p}{n_1}\right) + \lambda^1\left[\sigma^1\left(\delta^{-1}(p)\right) - \sigma^1\left(\delta^{-1}(d^1)\right)\right]\right\}^2} \tag{25}$$

This derivative is negative for $p \geq p^F$ and is positive for the point $\bar{p}_1 < p^F$ satisfying $\sigma^1\left(\delta^{-1}\left(\bar{p}_1\right)\right) - \sigma^1\left(\delta^{-1}\left(d^1\right)\right) = 0$. In other words, as it becomes more intensely concerned about fairness, group 1 becomes less bold when it is assigned a probability equal or larger that what its members perceive to be the fairest solution. At p^F this is due to the fact that the disagreement point is less fair than p^F coupled with the fact that, as they become more concerned about fairness, the members of group 1 become more reluctant to take a chance of a breakdown in the negotiation. When $p > p^F$ there are opposing forces at work. On the one hand the fact that the solution becomes less fair means that increased concern about fairness tends to make the group take a bolder stance in the negotiation. However, the marginal gain from an increase in the probability p is perceived to be unfair, hence the group tend to be less bold. When p is larger but close enough to p^F the second effect dominates the first and the overall stance becomes less bold.

Compared to the point $\bar{p}_1$ disagreement is not less fair. On the other hand, a marginal increase in p elevates the level of fairness. Not surprisingly, therefore, a greater concern for fairness makes members of group 1 more inclined to risk a breakdown in the negotiation to attain a solution that is at the same time fairer and preferred from their selfish point of view. They take a bolder stance.

By continuity, there is a point $p_1^* \in \left(\bar{p}_1, p^F\right)$ at which the countervailing forces describe above cancel each other out. At this point, defined by $db^1\left(p_1^*, \lambda^1\right)/d\lambda^1 = 0$, an increased intensity of the sense of fairness does not change the boldness of the negotiation stance of group 1. (This is first coordinate of the point around which the graph of the boldness function, b^1, pivots in response to changes in the intensity of the sense of fairness).

Thus far the discussion focused on the change in the boldness of group 1. However, the outcome of the negotiation (that is, the Nash bargaining solution) depends on the change in the boldness of group 1 and the level of boldness of group 2. In particular, if initially group 1 received less than its fair share, that is, $p_0^N < p_1^* \leq p^F$, an increase in group 1's concern for fairness induces it to take a bolder negotiating stance and, as a result, the solution p_1^N tends to be fairer from the perspective of this group. However, if $p_1^* < p_0^N < p^F$ then an increase in the intensity of the sense of fairness of members of group 1 makes them take a less bold position. As a result, they lose and the new bargaining solution is both less fair and less preferred from a selfish point of view (see Figure 2).

If initially group 1 received more than its fair share, that is, $p_0^N > p^F$, an increase in this group's concern for fairness makes it less bold and, consequently, the new bargaining solution p_1^N is such that $p_1^N < p_0^N$. In general, by becoming less bold, group 1 stands to lose from a selfish point of view, but it may gain from the point of view of fairness. If p_0^N is larger than p^F but is sufficiently close to it, the decline in the boldness of group 1 makes it lose on both counts. The new Nash bargaining solution is both, less fair and less preferred from group 1 selfish point of view.

To summarize this discussion we note that there exists $\varepsilon > 0$ such that if p_0^N, the Nash bargaining solution corresponding to λ^1 satisfies $p_0^N \in (p_1^*, p^F + \varepsilon)$ then the Nash bargaining solution , p_1^N, corresponding to $\hat{\lambda}^1 > \lambda^1$, is smaller and less fair (see Figure 2). In this case, an increase in group 1's intensity of the sense of fairness, by making it less bold, leads to a social policy that is less fair according to this group's notion of fairness and less preferred by this group selfish interests. If $p_0^N \in [0, p_1^*] \cup [p^F + \varepsilon, 1]$ then the new Nash bargaining solution, p_1^N tends to be closer to p^F. In this case, an increase in group 1's intensity of the sense of fairness leads to a social policy that is fairer according to this group's notion of fairness.

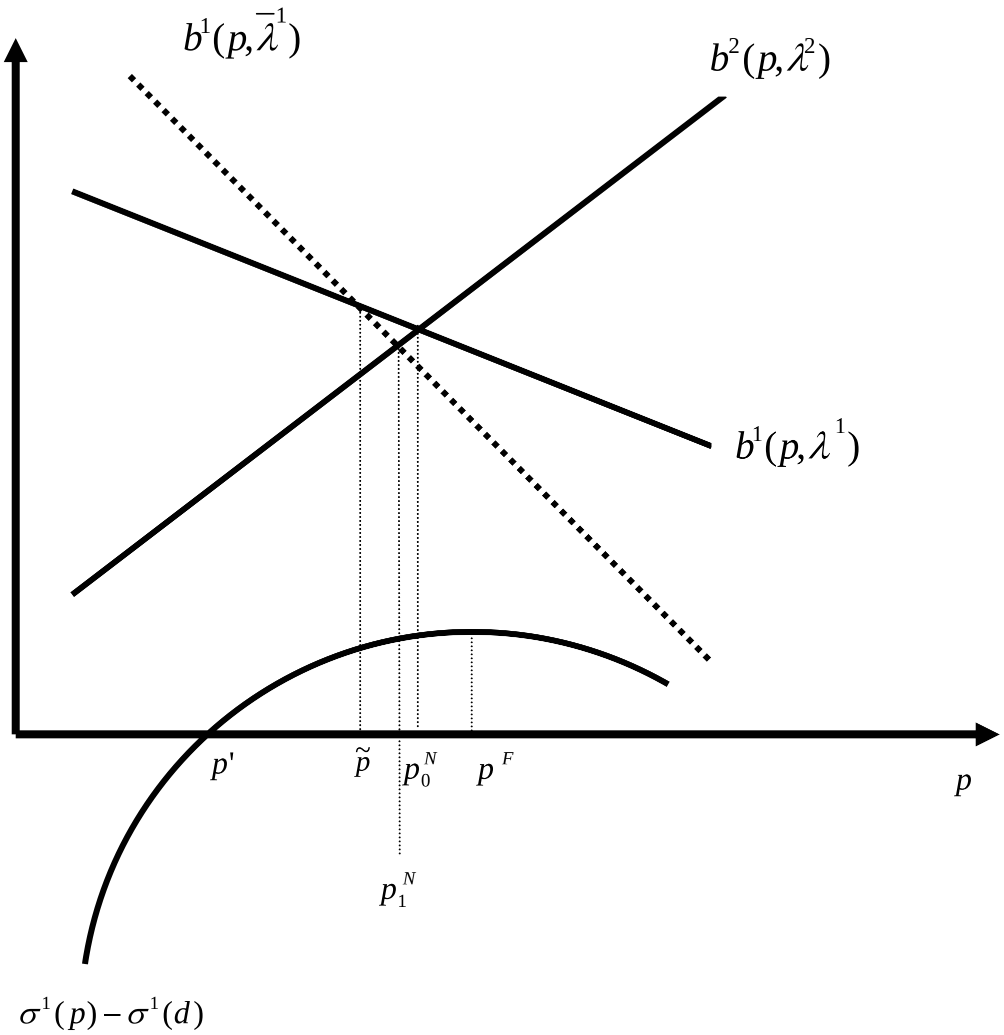

Figure 2: An increase in group 1 intensity may shift the solution further away from p^F

4.3 Increases in fairness vs. increases in risk aversion.

A well known comparative statics result in bargaining theory is that, in the bilateral case, an increase in the level of risk aversion of one of the bargainers improves the outcome from the point of view of the second (see Kihlstrom, Roth and Schmeidler [12]). In our model, increasing the intensity of the sense of fairness has an effect on the utility function, which seems to be similar to that of an increasing risk aversion (that is, both increase the concavity of the utility functions). Yet, unlike in the case of risk aversion, the effect of an increase of the intensity of sense of fairness of one of the parties on the welfare of the other is ambiguous. The explanation for this difference between the conclusions has to do with the role of the disagreement point. Specifically, in the case of increasing risk aversion the ranking of the disagreement point relative to the other outcomes is unchanged whereas in our model an increase in the intensity of the sense of fairness makes some outcomes less desirable than the disagreement point.

5 Conclusions

Building upon Hume's [6] idea that human actions are governed, in part, by a moral sense and are guided by the particular pleasure and pain associated with virtue and vice, Karni and Safra [10] developed an axiomatic model of individual choice over allocation procedures that are ex-ante fair to different degrees. In that model, both the notion of fairness and the intensity of individual sense of fairness are subjective.

In this paper, we propose a social choice theory based on the premise that social policies reflect the ideas of fairness held by the constituent members and their relative power. In particular, we bring our model of individual behavior to bear upon the analysis of the fairness of social policies that, in order to be implemented, must be agreed upon by the constituent members holding distinct notions of fairness and conflicting interests. Accordingly, we model the resolution of social conflicts using the Nash bargaining solution. Thus in our model a social conflict is resolved once a policy is found such that there exist no other policy and a bargaining party that may claim that a change to that policy would yield this party a proportional utility gain exceeding the proportional utility loss to the other parties. In this context we conducted comparative statics analysis of the effect of an increase in the sense of fairness of one group on the fairness of the social policy. We show that the results depend on the fairness of the outcome in case the negotiation break down and no agreement is achieved.

It is natural to think that if the intensity of the individual sense of fairness of the members of a particular group increases, that is, if the group stands more ready to reject agreements that its members consider to be unfair, the resulting social policy will be fairer according to this group's notion of fairness. We show that this is true in the limit, when the sense of fairness is infinitely strong. We also show that this conclusion is necessarily true in bilateral bargaining situations, provided the disagreement point is as fair as the fairer solution. However, in bilateral bargaining situations when disagreement results in an a policy that is not the fairest, and in multilateral bargaining situations it may be the case that, locally, an increase in the sense of fairness of a group results in a social policy that is less fair according to this group idea of fairness.

References

[1] Aumann RJ, Kurz M (1977) Power and Taxes. Econometrica 45: 1137-1161

[2] Camerer CF (1997) Progress in Behavioral Game Theory. J Econ Perspect 11: 167-188

[3] Harsanyi JC (1953) Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking. J Polit Econ 61: 343-435

[4] Harsanyi JC (1955) Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility. J Polit Econ 63: 309-321

[5] Harsanyi JC (1977) Rational Behavior and Bargaining Equilibrium in Games and Social Situations. Cambridge University Press, Cambridge

[6] Hume D (1740) Treatise of Human Nature. Dent JM and Sons, London, 1939

[7] Jasso G (1989) Self-Interest, Distributive Justice, and the Income Distribution: A Theoretical Fragment Based on St. Anselm's Postulate. Soc Just Res 3: 251-276

[8] Kalai E, Smorodinsky M (1975) Other Solutions to Nash's Bargaining Problems. Econometrica 43: 513-518

[9] Karni E (1996) Social Welfare Functions and Fairness. Soc Choice Welf 13: 487-496

[10] Karni E, Safra Z (2002a) Individual Sense of Justice: A Utility Representation. Econometrica 70: 263-284

[11] Karni E, Safra Z (2002b) The Intensity of the Sense of Fairness: Measurement and Behavioral Characterization. J Econ Theory 105: 318-337

[12] Kihlstrom RE, Roth AE, Schmeidler D (1981) Risk Aversion and Solutions to Nash's Bargaining Problem. In Moeschlin O, Pallaschke D (eds) Game Theory and Mathematical Economics. North Holland, Amsterdam, pp. 65-71

[13] Machina M (2001) Payoff Kinks in Preferences over Lotteries,. J Risk Uncer 23: 207-260

[14] Myerson RB (1991) Game Theory. Harvard University Press, Cambridge

[15] Nash JF (1950) The Bargaining Problem. Econometrica 18: 155-162

[16] Rubinstein A, Safra Z, Thomson W (1992) On the Interpretation of the Nash Bargaining Solution and its Extension to Non-Expected Utility Preferences. Econometrica 60: 1171 - 1186

[17] Young PH (1994) Equity In Theory and Practice. Princeton University Press, Princeton New Jersey.