

MORAL SENTIMENTS AND SOCIAL CHOICE:

FAIRNESS CONSIDERATIONS IN UNIVERSITY ADMISSIONS

Edi Karni and Zvi Safra*

Johns Hopkins University and Tel Aviv University

December 2, 2002

Abstract

We examine the implications of individuals having an intrinsic sense of fairness with regard to social choice. 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 university admission policies. We assume that these policies are determined by bargaining over test scores to be used as cut-off points for admission of members of diverse social groups. We show when a more intense sense of fairness of the members of a group leads to admission policy that is more compatible with the idea of fairness held by members of that group. Consequently, a society whose members have a common notion of fairness may implement fairer admission policies even if these policies are ultimately determined by the bargaining power of the different groups.

*While working on the research that is summarized in this paper the authors visited several institutions. Edi Karni visited the Bergals School of Economics, Tel Aviv University and GRID, at Ecole Normale Supérieure 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.

1 Introduction

Social policies and institutions are shaped by the power of the constituent members of the society to influence these policies and institutions. A well recognized source of power is the conviction of individuals that the policies that they support are just. Since, in general, different individuals may hold different views as to what constitute fairness, the ultimate shape of policies and institutions 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 concern for fairness in shaping admission policies at selective colleges and universities. 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 justice and are willing to act upon it. 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.

Our analysis highlights several aspects of the issue: the effect on the admission 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

¹This idea has a long history that goes back to St. Anselm (see discussion and references in Jasso, 1989). Karni and Safra (2002a) provides additional arguments and further references.

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 (2002a) 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 admission policies may be modeled using a similar analytical framework.

Even if there is agreement on the fairness of ranking of alternative admission policies, there may be individual differences in 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 emotional attachment to fairness. In Karni and Safra (2002b) we developed measures of the intensity of individual sense of fairness. Here we apply these measures to the analysis of college admission policies.

There are many processes by which social policies may be decided; *ceteris paribus* the outcome may depend on the particular process employed in a given situation. We assume that the only restriction on the adoption of an admission policy is that it is agreed upon by all interested parties and 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 equitable if no change is justified.³ We do not claim that this is how admission policies are decided in practice or that this is how they should be decided. Rather we use this model to illustrate how the procedure affects the manner in which the moral sentiments of the constituent members are expressed in the formulation of the policy. Specifically, we show that in bilateral bargaining situations, other things being equal, an

²See Hume (1740).

³For a discussion of bargaining as a social choice process, see Young (1994).

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 multilateral bargaining . In general, it is no longer true that an increase in the intensity of the sense of fairness of members of one group implies that the Nash bargaining solution tends to be fairer according to the notion of fairness held by members of this group. We show an example in which more intense concern for justice results in a social policy that is less just. We also give necessary conditions for such result to occur.

In Karni and Safra (2002a) we raised the possibility that, even though self-interest is the dominant motive governing individual behavior compared to which moral sentiments are a weak force, social policies may still be shaped largely by the individuals' moral value judgment. In this paper we return to this theme and discuss how this idea may work in the context of college admission policies.

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.

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 (1996). 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 (2002a) developed an axiomatic model of self-interest seeking moral individuals which is applied here to the analysis of social choice.

Let $N = \{1, \dots, n\}$, $2 < n < \infty$, 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 \succsim^i , representing individual i 's actual choice behavior and the fairness relation \succsim_F^i , representing individual i 's moral value judgment. The relation \succsim^i has the usual interpretation, namely, for any pair of allocation procedures \mathbf{p} and \mathbf{q} in P , $\mathbf{p} \succsim^i \mathbf{q}$ means that, if he were to choose between \mathbf{p} and \mathbf{q} , individual i would choose \mathbf{p} or would be indifferent between the two. The fairness relation, \succsim_F^i , has the interpretation of 'being fairer than' and $\mathbf{p} \succsim_F^i \mathbf{q}$

means that, according to individual i 's moral value judgment, the allocation procedure \mathbf{p} 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 (2002a) we used the juxtaposition of the preference relation and the fairness relation to derive a new binary relation \succsim_S^i on P representing the self-interest motive implicit in the individual choice behavior. Broadly speaking, an allocation procedure \mathbf{p} 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{p} 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 \succsim^i is represented by a utility function $V^i : \kappa^i(P) \times \sigma^i(P) \rightarrow \mathbb{R}$. Thus for all $\mathbf{p}, \mathbf{q} \in P$, $\mathbf{p} \succsim^i \mathbf{q} \Leftrightarrow V^i((\kappa^i \cdot \mathbf{p}, \sigma^i(\mathbf{p}))) \geq V^i((\kappa^i \cdot \mathbf{q}, \sigma^i(\mathbf{q})))$. In addition we characterize the case in which function V^i is additively separable in the self-interest and fairness components. Formally,

$$V^i((\kappa^i \cdot \mathbf{p}, \sigma^i(\mathbf{p}))) = h^i(\kappa^i \cdot \mathbf{p}) + \sigma^i(\mathbf{p}), \quad (1)$$

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

In Karni and Safra (2002b) we developed measures that make it possible to compare the intensity of the sense of justice of different individuals. In other words, we defined and characterized on the set N 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 (\succsim, \succsim_F) and $(\hat{\succsim}, \hat{\succsim}_F)$ are

comparable if they incorporate the same idea of fairness and induce the same self-interest relation (that is, if $\succsim_F = \hat{\succsim}_F$ and $\succsim_S = \hat{\succsim}_S$). For comparable preference-fairness relations with corresponding functional representations (h, κ, σ) and $(\hat{h}, \kappa, \hat{\sigma})$, $(\hat{\succsim}, \hat{\succsim}_F)$ possesses a stronger sense of fairness than (\succsim, \succsim_F) if and only if there exist real-valued functions f, g satisfying $\hat{h} = f \circ h$, $\hat{\sigma} = g \circ \sigma$ and $f'(h(\kappa \cdot \mathbf{q})) \leq g'(\sigma(\mathbf{q}))$ for all $\mathbf{q} \in P$. In what follows we take f to be the identity function and $g \circ \sigma = \lambda \sigma$, $\lambda \geq 0$. Hence the preference-fairness relation pair $(\hat{\succsim}, \hat{\succsim}_F)$ displays a more intense sense of fairness than the preference-fairness relation pair (\succsim, \succsim_F) if they are comparable and $\lambda \geq 1$.

2.2 Admission policies as random allocation procedures

Let $N = \{1, \dots, n\}$ $n > 2$ be a society consisting of a finite number of individuals. Let $\{N_j\}_{j=1}^m$ be a partition of N and suppose that each N_j represents a distinct social group. Denote by n_j the number of individuals belonging to N_j , and let $\alpha_j = n_j/n$ be the proportion of group N_j in the population. Consider an institution, for instance a college, that has a limited number b ($b < n$) of openings for new students and denote $\beta = b/n$. Assume that there exists a test whose score is positively correlated with the students' college performance. Each 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 are different and that, for all j , F_j first-order stochastically dominates F_{j+1} . A *feasible admission policy* is an m -tuple $s = (s_1, \dots, s_m) \in [0, 1]^m$, where s_j is the cutoff score for admission of members of group N_j , that satisfies the feasibility constraint $\sum_{j=1}^m \alpha_j (1 - F_j(s_j)) \leq \beta$. The set of all feasible admission policies is denoted by $S(\beta)$. 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.

The following two steps show that the set of feasible admission policies may be identified with the $m - 1$ dimensional simplex Δ . Let s be a given admission policy. First, associate

with it a vector of *normalized personal admission probabilities* $\mathbf{q}(s) \in \mathbb{R}^n$: for $i \in N_j$, let $\gamma^i(s) = (1 - F_j(s_j))$ and define $q^i(s) = \gamma^i(s) / \sum_{k \in N} \gamma^k(s)$. In this way $S(\beta)$ is embedded in P , the $n - 1$ dimensional simplex of \mathbb{R}^n . Then, note that since personal admission probabilities are identical within groups, the set $\mathbf{q}(S(\beta))$ can naturally be identified with the $m - 1$ dimensional simplex Δ : for $j \in \{1, \dots, m\}$, let $\pi_j(s) = (1 - F_j(s_j))$ and define $p_j(s) = \pi_j(s) / \sum_{k \in \{1, \dots, m\}} \pi_k(s)$. The first step of this construction is required for using Karni and Safra's (2002a) representation (defined over P). The second step helps in simplifying the forthcoming analysis. This issue is elaborated upon after the example.

Example: Affirmative action and the merit system.

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 figures 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 applied to the design of college admission policy. Foramlly, 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 probabilities $p_j(s^m) = (1 - F_j(r)) / \sum_{k \in \{1, \dots, m\}} (1 - F_k(r))$. The *proportional representation policy* s^{pr} is the admission policy satisfying $1 - F_j(s_j^{pr}) = \beta$, for all j . A fairness relation may take into consideration both the efficiency and 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 address 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:

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

where d is the Euclidean metric, $\mathbf{p}^m = \mathbf{p}(s^m)$, $\mathbf{p}^{pr} = \mathbf{p}(s^{pr})$, and ℓ is a monotonic decreasing and concave real-valued function. Note, however, that our approach is general and can accommodate different concepts of justice as they pertain to college admission policies. ■

Individual i in N has a preference relation \succsim^i and a moral value judgement \succsim_F^i that are defined on P . Clearly, these preferences correspond to the way the individual ranks feasible admission policies s in $S(\beta)$. We assume that these preferences satisfy the axioms of Karni and Safra (2002a) and hence \succsim^i can be represented by a separable utility function as in (1). The restrictions of these preferences to Δ are naturally defined and the same representations hold. Since the domain is restricted to Δ , we assume that all members of the same social group share the same preferences. In other words, for all $j = 1, \dots, m$ and for all $i, k \in N_j$, $\succsim^i = \succsim^k$ and $\succsim_F^i = \succsim_F^k$. Note that if the moral value judgment is commonly shared by all society members, 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 (1955) concept of social preference relation and to his concept of preference relation of an impartial observer (Harsanyi, 1953 and 1977). Even if the moral value judgment is commonly shared by all groups, the intensity of the sense of fairness may still vary among them.

Taking into account the assumptions made at the end of Section 2.1 and capturing the idea that each group's selfish concern is of a pure self-interest, the utility function in (1) gets the form

$$U^j(\mathbf{p}) = h^j(p_j) + \lambda^j \sigma^j(\mathbf{p}) \tag{2}$$

Note that β is omitted from the above expression. This can be done since its value is fixed. For convenience reasons, we henceforth assume that σ^j is non-negative and that its maximal value is zero.

3 Social Choice as Bargaining Solution

3.1 The bargaining model

The analysis of college admission policies is based on the premise that the 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 main difficulty in applying the Nash bargaining model stems from the fact that Nash's solution is based on the assumption that bargainers preference relations over the set of lotteries on agreements are linear in the probabilities. In our framework the agreements themselves, namely, the admission policies, are identified with lotteries that represent random allocation procedures of the available slots. In this context if, in order to conform with Nash's model, we were to introduce lotteries on random allocation procedures, we would have to assume that reduction of compound lotteries does not apply. This does not seem to make sense. We choose instead a different, more natural, approach.

Let Δ be the set of all possible agreements and denote by d the disagreement point. 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 the individuals 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 admission policy are rejected. In view of this consideration one interpretation of the disagreement point 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 assume that for every j there exists some random allocation procedure in the interior of 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 $(\alpha; \mathbf{p}) \in [0, 1] \times \Delta := B$, where α denotes the probability that the bargaining process will end with an agreement \mathbf{p} and the probability $(1 - \alpha)$ that 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 (1993).

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

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

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

where $U^j(\mathbf{p})$ is as in (2). Following the previous discussion regarding the disagreement point, it is assumed here that $U^j(d)$ is independent of λ^j . This, together with the assumption $\sigma^j \leq 0$, implies that the set of individually rational agreements shrinks as λ^j increases.

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, 1997). 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.

In order to guarantee that our bargaining problems are ‘well behaved’, that is, that the image of B in the utilities space is convex, we assume that for each j both h^j and σ^j are strictly concave. This concavity assumption plays in our model a role analogous to that of risk-aversion in the original Nash bargaining model.

Under these assumptions our model conforms to a Nash bargaining structure and the Nash bargaining solution, $N(B, d)$, is defined by

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

Our assumptions imply that the solution is unique. Clearly, it is achieved for $\alpha = 1$.

3.2 Admission policies and the power of moral conviction

3.2.1 The bilateral case

We examine the implications of variations in the intensity of the sense of fairness for college admission 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 heightened sense of fairness for the admission policies depend on the particular nature of the social decision-making procedure, in our case the Nash bargaining solution. To analyze

these effects it is best to start by considering a situation of bilateral bargaining where the influences are most transparent. In this case the set of allocation procedures Δ is the one-dimensional simplex. To simplify the exposition we denote the agreement point $(p, 1 - p)$ by p and write $h^i(p)$ for short. Note that this change of notation means that h^1 increases with p while h^2 decreases with p .⁴ Using the remaining degree of freedom in the utility representation, we normalize the utility functions so that $U^j(d) = 0$, for all j . The Nash bargaining solution is therefore the solution of

$$\arg \max [h^1(p) + \lambda^1 \sigma^1(p)]^{n_1} [h^2(p) + \lambda^2 \sigma^2(p)]^{n_2} \quad (5)$$

and the necessary and sufficient condition for maximum is:

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

Following Aumann and Kurz (1977) we define $b^j : \Delta \times \mathbb{R}_+ \rightarrow \mathbb{R}$, $j = 1, 2$ by

$$b^j(p, \lambda^j) = \frac{n_j \left[\frac{d}{dp} h^j(p) + \lambda^j \frac{d}{dp} \sigma^j(p) \right]}{h^j(p) + \lambda^j \sigma^j(p)}$$

and refer to it as the *boldness function*. The numerator of the 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. Clearly, the larger is the group the bolder it is and, consequently, the more favorably it is treated in the bargaining solution. Note that our assumptions imply that 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.

⁴More generally, elements of P may be normalized such that $p_n = 1 - \sum_{i=1}^{n-1} p_i$ and all derivatives are taken with respect to the first $n - 1$ variables. Under this normalization, the functions σ^i and $h^i \circ \kappa^i$ are defined over the projection of P over \mathbb{R}^{n-1} . For an elaborate discussion of issues involve in defining ‘probability derivatives’ see Machina (2001). Note, however, that choose a different approach for the multivariate case.

Theorem 1 *An increase in the intensity of the sense of fairness of members of either group implies that the admission policy under the Nash bargaining solution is fairer according to the notion of fairness held by members of that group.*

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

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

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

Clearly, the sign of $\partial b(p, \lambda^1) / \partial \lambda^1$ is determined by the sign of the numerator. We show next that

$$h^1(p) \frac{d}{dp} \sigma^1(p) - \sigma^1(p) \frac{d}{dp} h^1(p) \geq 0 \text{ if and only if } p \leq p^F \quad (8)$$

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

Since $\sigma^1(p) \leq 0$ and is concave and $\frac{d}{dp} \sigma^1(p^F) = \sigma^1(p^F) = 0$, $p \leq p^F$ implies $\frac{d}{dp} \sigma^1(p) \geq 0$. This, in turn, implies $h^1(p) \frac{d}{dp} \sigma^1(p) - \sigma^1(p) \frac{d}{dp} h^1(p) \geq 0$ for $p \leq p^F$.

If $p > p^F$ then $\frac{d}{dp} \sigma^1(p) < 0$ and expression (7) seems ambiguous. To show that it is negative note first that, for every given p , for λ^1 sufficiently close to 0, $h^1(p) + \lambda^1 \sigma^1(p) > 0$ and for λ^1 sufficiently large $h^1(p) + \lambda^1 \sigma^1(p) < 0$. Moreover, for every finite λ^1 ,

$$h^1(p^F) + \lambda^1 \sigma^1(p^F) > 0 \text{ and } \frac{d [h^1(p) + \lambda^1 \sigma^1(p)]}{dp} \Big|_{p=p^F} > 0. \quad (9)$$

Define $\bar{\lambda}^1$ by the equation $h^1(p) + \bar{\lambda}^1 \sigma^1(p) = 0$. Note that $\frac{d}{dp} [h^1(p) + \bar{\lambda}^1 \sigma^1(p)] < 0$ at p . Hence

$$\frac{h^1(p)}{\sigma^1(p)} = -\bar{\lambda}^1 < \frac{\frac{d}{dp} h^1(p)}{\frac{d}{dp} \sigma^1(p)}. \quad (10)$$

Therefore $h^1(p) \frac{d}{dp} \sigma^1(p) - \sigma^1(p) \frac{d}{dp} h^1(p) < 0$.

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

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

If the initial position is an admission policy in which members of group 1 are treated favorably relative to the fair admission procedure, that is $p^F < p^N$ then the new Nash solution is at a smaller value of p , hence closer to p^F (this is the situation depicted in Figure 1). ■

3.2.2 The multilateral case

Consider next the implications of an increase in the intensity of the sense of fairness of members of one of the groups in a multilateral bargaining situation. The example below shows that, unlike the bilateral bargaining case, it is not true that an increase in the intensity of the sense of fairness of members of any group implies that the admission policy under the Nash bargaining solution is fairer according to the notion of fairness held by members of that group. In this example 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.

Example: An increase in the sense of fairness that causes a decrease in fairness.

Let $m = 3$ and consider the log of the Nash maximizing problem:

$$\max_{\mathbf{p}} \sum_{j=1}^3 n_j [\log h^j(p^j) + \lambda^j \sigma^j(\mathbf{p})] \quad \text{subject to} \quad 1 - \sum_{i=1}^3 p_i = 0$$

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

and sufficient conditions are

$$\frac{n_j \left[\frac{dh^j}{dp_j} (p^j) \right]}{h^j (p^j)} - \mu = 0, \quad j = 1, 2, 3 \quad (11)$$

and

$$1 - \sum_{j=1}^3 p_j = 0 \quad (12)$$

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:

$$\begin{pmatrix} W_1 & 0 & 0 & -1 \\ 0 & W_2 & 0 & -1 \\ 0 & 0 & W_3 & -1 \\ -1 & -1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{dp_1}{d\lambda^1} \\ \frac{dp_2}{d\lambda^1} \\ \frac{dp_3}{d\lambda^1} \\ \frac{d\mu}{d\lambda^1} \end{pmatrix} = \begin{pmatrix} -X \\ -Y \\ -Z \\ 0 \end{pmatrix} \quad (13)$$

where

$$W_j = n_j \frac{h^j (p^j) \frac{d^2 h^j}{dp_j^2} (p^j) - \left(\frac{dh^j}{dp_j} (p^j) \right)^2}{(h^j (p^j))^2}, \quad j = 1, 2, 3 \quad (14)$$

and

$$X = n_1 \frac{h^1 (p^j) \frac{\partial \sigma^1}{\partial p_1} (\mathbf{p}) - \frac{dh^1}{dp_1} (p^j) \sigma^1 (\mathbf{p})}{(h^1 (p^j))^2}, \quad Y = \frac{n_1 \frac{\partial \sigma^1}{\partial p_2} (\mathbf{p})}{h^1 (p^j)}, \quad Z = \frac{n_1 \frac{\partial \sigma^1}{\partial p_3} (\mathbf{p})}{h^1 (p^j)} \quad (15)$$

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

$$\frac{dp_1}{d\lambda^1} = \frac{1}{D} [X (W_2 + W_3) - YW_3 - ZW_2] \quad (16)$$

$$\frac{dp_2}{d\lambda^1} = \frac{1}{D} [-XW_3 + Y (W_1 + W_3) - ZW_1] \quad (17)$$

and

$$\frac{dp_3}{d\lambda^1} = \frac{1}{D}[-XW_2 - YW_1 + Z(W_1 + W_2)] \quad (18)$$

Thus

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

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

$$\frac{\partial\sigma^1(\mathbf{p})}{\partial p_j} = f'(d(\mathbf{p})) \frac{p_j - \frac{1}{3}}{-d(\mathbf{p})}, \quad \frac{\partial\sigma^1(\mathbf{p})}{\partial p_j} - \frac{\partial\sigma^1(\mathbf{p})}{\partial p_k} = f'(d(\mathbf{p})) \frac{p_j - p_k}{-d(\mathbf{p})} \text{ and } W_j = -\frac{n_j}{p_j^2} \quad (20)$$

Moreover, since by (11), $n_i/p_i = \mu$ then, by (15),

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

Hence,

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

Suppose that $n_1 = 150, n_2 = 110$ and $n_3 = 40$ and consider $\mathbf{p} = (\frac{15}{30}, \frac{11}{30}, \frac{4}{30})$ that satisfies equations (11) with $\mu = 300$. Then $d(\frac{15}{30}, \frac{11}{30}, \frac{4}{30}) = \frac{7.87}{30}$ and $\text{sign} \mu^2 f'(d(\mathbf{p})) / d(\mathbf{p}) D = \text{sign} D = (-1)^3 < 0$. Thus the sign of $d\sigma^1(\mathbf{p}) / d\lambda^1$ 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

$$\frac{d\sigma^1(\mathbf{p})}{d\lambda^1} < 0. \quad (23)$$

■

It is easy to verify, using equation (16), that $dp_1/d\lambda_1 > 0$. In other words, the admission probability of group 1 under the Nash bargaining solution increases when the sense of fairness of group 1 increases. The following theorem shows that, under the Nash bargaining solution, an increase in its own probability is a necessary condition for the admission policy to be less fair according to the notion of fairness held by members of the group whose sense of fairness become more intense.

Theorem 2 *Assume that, under the Nash bargaining model, an increase in the intensity of the sense of fairness of members of a group leads to a decrease in admission probability of members of that group. Then an increase in the intensity of the sense of fairness of members of that group implies that the admission policy is fairer according to the notion of fairness held by members of that group.*

Proof. Without loss of generality we prove the theorem for an increase in the sense of fairness of members of group 1. Let $g(\mathbf{p}) = \prod_{j=2}^m [(h^j(p_1) + \lambda^j \sigma^j(\mathbf{p}))^{n_j}]$. For every given $\lambda^1 \in [0, \infty)$ let $\mathbf{p}^N(\lambda^1)$ be the corresponding Nash bargaining solution. That is

$$\mathbf{p}^N(\lambda^1) = \arg \max_{\mathbf{p} \in \mathbf{P}} (h^1(p_1) + \lambda^1 \sigma^1(\mathbf{p}))^{n_1} g(\mathbf{p}) \quad (24)$$

Let $\mathbb{P} = \{\mathbf{p}^N(\lambda^1) \in \Delta \mid \lambda^1 \in [0, \infty)\}$. Note that $\mathbf{p}^N(\lambda^1) \rightarrow \mathbf{p}^F$ as $\lambda^1 \rightarrow \infty$, where \mathbf{p}^F is the most fair outcome according to group 1. We need to show that $d\sigma^1(\mathbf{p}^N(\lambda^1))/d\lambda^1 > 0$ for all $\lambda^1 \in [0, \infty)$.

For every given $\bar{\lambda}^1 \in [0, \infty)$, define the boldness function, b^1 , of group 1 to be the derivative of $\log[h^1(p^N(\lambda^1)_1) + \bar{\lambda}^1 \sigma^1(\mathbf{p}^N(\lambda^1))]^{n_1}$ with respect to λ^1 . Thus

$$b^1(\mathbf{p}^N(\lambda^1), \bar{\lambda}^1) = n_1 \frac{\frac{d}{d\lambda^1}[h^1(p^N(\lambda^1)_1) + \bar{\lambda}^1 \sigma^1(\mathbf{p}^N(\lambda^1))]}{h^1(p^N(\lambda^1)_1) + \bar{\lambda}^1 \sigma^1(\mathbf{p}^N(\lambda^1))} \quad (25)$$

The corresponding boldness function of g is defined by $b^g(\mathbf{p}^N(\lambda^1)) = d \log g(\mathbf{p}^N(\lambda^1)) / d\lambda^1$.

The Nash bargaining solution implies that $b^1(\mathbf{p}^N(\lambda^1), \lambda^1) = -b^g(\mathbf{p}^N(\lambda^1))$ for every $\lambda^1 \in [0, \infty)$. Because the last equality must hold as an identity for all $\lambda^1 \in [0, \infty)$ we have,

$$\frac{db^1(\mathbf{p}^N(\lambda^1), \lambda^1)}{d\lambda^1} = \frac{-db^g(\mathbf{p}^N(\lambda^1))}{d\lambda^1} \quad (26)$$

But

$$\frac{db^1(\mathbf{p}^N(\lambda^1), \lambda^1)}{d\lambda^1} = \frac{\partial b^1(\mathbf{p}^N(\lambda^1), \bar{\lambda}^1)}{\partial \lambda^1} \Big|_{\lambda^1 = \bar{\lambda}^1} + \frac{\partial b^1(\mathbf{p}^N(\bar{\lambda}^1), \lambda^1)}{\partial \lambda^1} \Big|_{\lambda^1 = \bar{\lambda}^1} \quad (27)$$

Hence

$$\frac{\partial b^1(\mathbf{p}^N(\bar{\lambda}^1), \lambda^1)}{\partial \lambda^1} \Big|_{\lambda^1 = \bar{\lambda}^1} = - \left[\frac{\partial b^1(\mathbf{p}^N(\lambda^1), \bar{\lambda}^1)}{\partial \lambda^1} + \frac{db^g(\mathbf{p}^N(\lambda^1))}{d\lambda^1} \right] \Big|_{\lambda^1 = \bar{\lambda}^1} > 0 \quad (28)$$

where the inequality is implied by the second order condition. Differentiating $b^1(\mathbf{p}^N(\bar{\lambda}^1), \lambda^1)$ with respect to λ^1 we obtain:

$$\begin{aligned} & \frac{\partial b^1(\mathbf{p}^N(\bar{\lambda}^1), \lambda^1)}{\partial \lambda^1} \Big|_{\lambda^1 = \bar{\lambda}^1} = \\ & n_1 \frac{h^1(p^N(\bar{\lambda}^1)_1) \frac{d}{d\lambda^1} \sigma^1(\mathbf{p}^N(\lambda^1)) \Big|_{\lambda^1 = \bar{\lambda}^1} - \sigma^1(\mathbf{p}^N(\bar{\lambda}^1)) \frac{d}{d\lambda^1} h^1(p^N(\lambda^1)_1) \Big|_{\lambda^1 = \bar{\lambda}^1}}{\left[h^1(p^N(\bar{\lambda}^1)_1) + \bar{\lambda}^1 \sigma^1(\mathbf{p}^N(\bar{\lambda}^1)) \right]^2} \quad (29) \end{aligned}$$

Suppose that

$$\frac{d\sigma^1(\mathbf{p}^N(\lambda^1))}{d\lambda^1} \Big|_{\lambda^1=\bar{\lambda}^1} \leq 0 \quad (30)$$

Then, by (29),

$$\frac{\partial b^1(\mathbf{p}^N(\bar{\lambda}^1), \lambda^1)}{\partial \lambda^1} \Big|_{\lambda^1=\bar{\lambda}^1} \leq n_1 \frac{-\sigma^1(\mathbf{p}^N(\bar{\lambda}^1)) \frac{d}{d\lambda^1} h^1(p^N(\lambda^1)_1) \Big|_{\lambda^1=\bar{\lambda}^1}}{\left[h^1(p^N(\bar{\lambda}^1)_1) + \bar{\lambda}^1 \sigma^1(\mathbf{p}^N(\bar{\lambda}^1)) \right]^2} \quad (31)$$

If $d\mathbf{p}_1^N(\lambda^1)/d\lambda^1 \Big|_{\lambda^1=\bar{\lambda}^1} \leq 0$ then $dh^1(p^N(\lambda^1)_1)/d\lambda^1 \Big|_{\lambda^1=\bar{\lambda}^1} \leq 0$, and the expression on the right hand side of inequality (31) is negative. Thus

$$\frac{\partial b^1(\mathbf{p}^N(\lambda^1), \bar{\lambda}^1)}{\partial \bar{\lambda}^1} \Big|_{\lambda^1=\bar{\lambda}^1} < 0 < -\left[\frac{\partial b^1(\mathbf{p}^N(\bar{\lambda}^1), \lambda^1)}{\partial \bar{\lambda}^1} + \frac{db^g(\mathbf{p}^N(\lambda^1))}{d\lambda^1} \right] \Big|_{\lambda^1=\bar{\lambda}^1} \quad (32)$$

But (32) contradicts (28). Hence $d\sigma^1(\mathbf{p}^N(\lambda^1))/d\lambda^1 > 0$. ■

To grasp the intuition of the counter example and theorem 4 consider the situation depicted in Figure 2. The initial solution is at the point p_0^N in panel (c) and its projections are indicated in panels (a) and (b) where the axes depict the probabilities of group 1 and group $j = 1, 2$. In each case p_1 is increasing towards the right. In panel (a) p_2 increases towards the left and p_3 is constant and in panel (b) p_3 increases towards the left and p_2 is constant. (This directions are also indicated by the solid arrows in panel (c).) Note that the graph of the fairness function is strictly below the horizontal axes. This is apparent from panel (c) which shows that $p^F = (1/3, 1/3, 1/3)$ is on neither of these axes. Consequently, it is easy to verify that an increase in the sense of fairness of group 1 rotates the graphs of boldness function b^1 around points *to the right* of the points \bar{p}_2 and \bar{p}_3 that correspond to the maximal levels of σ . More specifically, differentiating the boldness function of group 1 with respect to λ^1 and we get:

$$\frac{\partial b^1(\mathbf{p}, \lambda^1)}{\partial \lambda^1} = n_1 \frac{h^1(p_1) \frac{d}{d\lambda^1} \sigma^1(\mathbf{p}) - \sigma^1(\mathbf{p}) \frac{d}{d\lambda^1} h^1(p_1)}{\left[h^1(p_1) + \bar{\lambda}^1 \sigma^1(\mathbf{p}) \right]^2}$$

Now if the points $(p_1, \hat{p}_2, \bar{p}_3)$ and $(p_1, \bar{p}_2, \hat{p}_3)$ are in the neighborhood of \mathbf{p}_0^N (see Figure 2 panel (c)) then $d\sigma^1(\mathbf{p}_0^N)/d\lambda^1$ is close to zero and $\partial b^1(\mathbf{p}, \lambda^1)/\partial \lambda^1 > 0$. But in this case the effect of the increase boldness of group 1 is to shift the solution in a direction that entails increase in the selfish component of the utility namely, an increase in p_1 . (The effect of fairness in the neighborhood of \mathbf{p} is negligible.) If the graphs of the function b^g are in the positions depicted in Figure 2 then the equilibrium solution shifts to the right and to lower level of fairness (this explains the counterexample). This cannot happen if the intersections of b^1 and b^g are to the left of the points \bar{p}_1 and \bar{p}_2 because, in this case, an increase in p_1 corresponds to increase in σ (this explains theorem 4). Clearly if the intersections of b^1 and b^g are to the right of the points around which b^1 rotates then the shift of the solution is to the left and the level of fairness increases. Other possibilities are present but not discussed.

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 institution. This requirement is, in general, difficult if not impossible to meet. It seems reasonable to suppose that the idea of justice 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 $\sigma^1(p) = \ell(d(\mathbf{p}, \mathbf{p}^m))$ and $\sigma^2(p) = \ell(d(\mathbf{p}, \mathbf{p}^{pr}))$, where $\ell(\cdot)$ is a nonpositive, concave function and $\ell(0) = 0$. (The situation is depicted in Figure 1.) 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 if no agreement is reached then the college will suffer a cut-off of fund and will be forced to close down. The disagreement point corresponds to the fairest treatment of the different groups in these circumstances. In less extreme situations disagreement may imply the college has to limit admissions (lower β) thus reducing the overall probability of admission. Thus one may think of the disagreement point as a point $\mathbf{d} \in \mathbb{R}^m$ such that $\sum_{j=1}^m d_j = 1$ and $\sum_{j=1}^m \alpha_j d_j < \beta$. In this interpretation d_j is the admission probability of individuals belonging to group j if no agreement is reached. Now, if $\sigma^j(\mathbf{d}) < \sigma^j(\mathbf{p}^F)$ for some group j , which is bound to be the case if the idea of fairness is not shared by all the social groups, then, in general, $u^j(\mathbf{d}) \neq 0$. To grasp the significance of this change of interpretation consider the bilateral bargaining case and let $\mathbf{d} = (d, 1 - d)$. The boldness function of group $j = 1, 2$ is given by:

$$b^j(p, \lambda^j) = \frac{n_j \left[\frac{d}{dp} h^j(p) + \lambda^j \frac{d}{dp} \sigma^j(p) \right]}{[h^j(p) - h^j(d)] + \lambda^j [\sigma^j(p) - \sigma^j(d)]} \quad (33)$$

We may choose $h^j(d) = 0, j = 1, 2$. As before, our assumptions imply that b^1 decreases and b^2 increases, with respect to p . Now, if $\sigma^j(p^F) - \sigma^j(d) > 0$ then there is a neighborhood of p^F in which $\sigma^j(p^F) - \sigma^j(d)$ is strictly positive (the situation is depicted in Figure 4). Denote by p' and p'' the solutions of $\sigma^1(p^F) - \sigma^1(d) = 0$ such that $p' < p''$.

Consider next an increase in the intensity of the sense of fairness of members of group 1 from λ^1 to $\bar{\lambda}^1$ (see Figure 4). To verify that the boldness function of group 1 to shifts and $b^1(\tilde{p}, \lambda^1) = b^1(\tilde{p}, \bar{\lambda}^1)$ at some point $\tilde{p} \in (p', p^F)$. Note that

$$\frac{db^1(p, \lambda^1)}{d\lambda^1} = \frac{h^1(p) \frac{d}{dp}\sigma^1(p) - \frac{d}{dp}h^1(p) [\sigma^1(p) - \sigma^1(d)]}{\{h^1(p) + \lambda^1 [\sigma^1(p) - \sigma^1(d)]\}^2} \quad (34)$$

Thus, for $p < p'$, $[\sigma^1(p) - \sigma^1(d)] < 0$ hence the expression in (34) is positive. For $p \geq [p^F, p'']$ the expression in (34) is negative.

Let p_0^N and p_1^N be the Nash bargaining solutions corresponding to λ^1 and $\bar{\lambda}^1$, respectively. Then using Figure 4 it is easy to verify that: (a) If $p_0^N \leq \tilde{p}$ or $p^F \leq p_1^N$, an increase in the sense of fairness of group 1 implies that the admission policy under the Nash bargaining solution is fairer according to the notion of fairness held by members of that group. (b) If $p_0^N > \tilde{p}$ and $p^F > p_1^N$ then an increase in the intensity of the sense of fairness of members of group 1 implies that the admission policy under the Nash bargaining solution becomes less fair according to the notion of fairness held by members of that group.

4.3 On power, justice, and social policies

Discussing the potential implications of our approach for social choice theory in Karni and Safra (2002a) we raised the possibility that, even if self-interest is the dominant motive governing individual behavior, social policies may nevertheless be largely shaped by a commonly held moral value judgment. The argument is that when a moral value judgment is shared by many individuals, the aggregation of individual preference in the social decision-making process accumulates the intensity of their moral sentiments. No such cumulative effect is present when the selfish component of the individual preferences is aggregated. To illustrate this idea consider again the bilateral bargaining model of the preceding section and the situation depicted in Figure 3 below. Let p_0^N denote the Nash bargaining solution corresponding to the case in which individuals have no concern for fairness, i.e., $\lambda^1 = \lambda^2 = 0$. Introducing

some concern for fairness so that $\lambda^1 = \lambda^2 = \lambda > 0$ shifts the solution to p_1^N which is closer to the fair solution p^F . This shift is due to the cumulative effect of the increase in the intensity of the sense of fairness of both groups. The fairness motive makes the different groups ready to accept a shift of the Nash solution in the same direction, name, in the direction of a fairer outcome.

References

- [1] Aumann, Robert J. and Mordecai Kurz (1977): “Power and Taxes,” *Econometrica*, 45, 1137-1161.
- [2] Camerer, Colin F. (1997): “Progress in Behavioral Game Theory,” *The Journal of Economic Perspectives*, 11, 167-188.
- [3] Harsanyi, J. C. (1953), “Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking,” *Journal of Political Economy*, 61, 343-435.
- [4] Harsanyi, John (1955): “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility,” *Journal of Political Economy*, 63, 309-321.
- [5] Harsanyi, John (1977): *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
- [6] Hume, David (1740): *Treatise of Human Nature*.” London: J. M. Dent and Sons, 1939.
- [7] Jasso, Guillermina (1989) “Self-Interest, Distributive Justice, and the Income Distribution: A Theoretical Fragment Based on St. Anselm’s Postulate,” *Social Justice Research*, 3, 251-276.
- [8] Karni, Edi (1996) “Social Welfare Functions and Fairness,” *Social Choice and Welfare*, 13, 487-496.
- [9] Karni, Edi and Zvi Safra (2002a) “Individual Sense of Justice: A Utility Representation,” *Econometrica*, 70, 263-284.
- [10] Karni, Edi and Zvi Safra (2002b) “The Intensity of the Sense of Fairness: ” *Journal of Economic Theory* 105, 318-337.

- [11] Machina, Mark (2001) "Payoff Kinks in Preferences over Lotteries," *Journal of Risk and Uncertainty* 23, 207-260.
- [12] Nash, John F. (1950) "The Bargaining Problem," *Econometrica*, 18, 155-162.
- [13] Rubinstein, Ariel, Safra, Zvi and William Thomson (1992) "On the Interpretation of the Nash Bargaining Solution and its Extension to Non-Expected Utility Preferences," *Econometrica*, 60, 1171 - 1186.
- [14] Young, Peyton H. (1994) *Equity In Theory and Practice*. Princeton University Press, Princeton, New Jersey.