Interpersonal Comparison in Egalitarian Societies

Ken Binmore
Philosophy Department
Bristol University
Bristol, BS8 1TN
United Kingdom

k.binmore@ucl.ac.uk

Abstract: When judging what is fair, how do we decide how much weight to assign to the conflicting interests of different classes of people? This subject has received some attention in a utilitarian context, but has been largely neglected in the case of egalitarian societies of the kind studied by John Rawls. My *Game Theory and the Social Contract* considers the problem for a toy society with only two citizens. This paper examines the theoretical difficulties in extending the discussion to societies with more than two citizens.

Interpersonal Comparison in Egalitarian Societies[1]

by Ken Binmore

1   Introduction

My naturalistic reinterpretation of Rawls' [13] theory of justice treats fairness as a type of social norm that evolved for the purpose of selecting an efficient Nash equilibrium in coordination games that have many Nash equilibria. Only the bare bones of the basic theory are reviewed in this paper, which is devoted to a discussion of the difficulties in extending the theory to societies with more than two citizens. For versions of the two-person theory with progressively more detail, see Binmore [4, 3, 2]. For the relevant bargaining theory, see Binmore [5, 1].

Neoclassical or behavioral?   My theory is neither neoclassical nor behavioral. It differs from the standard behavioral approach in not treating fairness as an argument in a person's utility function. For example, Fehr and Schmidt's [8] theory of inequity aversion assumes that Adam suffers a direct loss of utility if he is assigned more of a surplus than Eve.

People certainly do have social preferences to a greater or lesser degree. Why else do most of us give money to charity? However, I do not think this is the right way to model fairness norms. In my theory, citizens may or may not have social preferences, but even if they were all entirely selfish, they would still be stupid to ignore the fairness norms that operate in their society.[2]

I differ from the tradition in neoclassical economics in recognizing that almost any realistic game has many Nash equilibria. The folk theorem of repeated game theory shows that the equilibrium selection problem is endemic. It tells us, for example, that even perfectly competitive markets have multiple Nash equilibria if they are repeated every day. It then ceases to be true that there is a necessary trade-off between equity and efficiency. Neoclassical arguments to this effect rely on models that have only one equilibrium. I think that it is only because traditional neoclassical economics largely confined its attention to models with a unique equilibrium that it managed to get so far without paying any serious attention to the relevance of fairness or other social norms to human behavior.

[2]The behavioral literature sometimes blurs the distinction between a social preference and a social norm, but it is important in my work to make a sharp distinction between a payoff function and an equilibrium selection device.

Modern equity theory. A small school of psychologists who work on "modern equity theory" comes closest to my own findings.[3] They find experimental support for Aristotle's ancient contention that "what is fair is what is proportional". More precisely, they argue that an outcome is regarded as fair when each person's gain over the status quo is proportional to that person's "social index". This conclusion is consistent, for example, with the widespread concern about preserving differentials between the wages paid to different trades (such as electricians or carpenters) when there is a general wage increase.

What determines a person's social index? The psychological literature suggests that social indices vary with the context. They also seem to vary with the culture within which a particular context arises (Elster [7], Young [23]). My theory therefore treats social indices as the product of social or culural evolution. In my admittedly simplistic models, the directions in which Adam and Eve's social indices vary with contextual changes in need, effort, ability, and social status are the same in all cultures, but the importance attached to such factors may well be markedly different when one moves from one culture to another.

One reason that my models are simplistic is that they consider only idealized societies with just two players. This paper discusses the problems involved in generalizing the approach to larger societies. The conclusion is that further pursuit of my theoretical approach to the evolution of standards of interpersonal comparison is unlikely to be profitable. If appeals are to be made to to our natural sense of fair play when seeking support for reforms in matters of public policy, it will be necessary to use empirical methods to determine the social indices that people regard as relevant in the context under discussion.

2 The Original Position

Rawls' [13] original position is a hypothetical standpoint from which to make judgements about how a just society should be organized. Members of a society are asked to envisage the social contract to which they would agree if their own social role were concealed behind a "veil of ignorance". Behind this veil of ignorance, the distribution of advantage in the planned society would seem as though determined by a lottery. Agents thinking of advocating an unfair social contract in the original position therefore have to face the prospect that they themselves might end up as one of the disadvantaged citizens if their proposal were implemented.

Rawls defends the original position as a version of the Kantian categorical imperative, but I think the idea hits the spot with most people because they intuitively recognize the deep structure of the fairness norms that they actually use every day in solving the equilibrium selection problem in the myriads of small coordination games of which daily life largely consists.

---

[3] For example, Deutsch [6], Kayser *et al* [10], Lerner [11, 12], Reis [14], Sampson [17], Schwartz [18], Wagstaff [20], Walster *et al* [21, 22].

I am at one with Rawls in hoping that we shall one day learn to organize our societies on fair principles, but I think it a mistake to look to the way nation states are currently organized for inspiration on how the fairness norms that evolved with our species are structured. Still less is inspiration to be found in the rhetoric of philosophers of the metaphysical persuasion. We need to lower our sights to the kind of coordination games that we commonly solve without thought or discussion—usually so smoothly and effortlessly that we don't even notice that there is a coordination problem to be solved. Who goes through that door first? How long does Adam get to speak before it is Eve's turn? Who gives way to whom when cars are maneuvering in heavy traffic? Who gets that parking space? Whose turn is it to wash the dishes tonight? These are picayune problems, but if conflict arose every time they needed to be solved, our societies would fall apart.

Evolution. Anthropologists commonly attribute the origin of social behavior in the human species to the gain in fitness we enjoy as a consequences of insuring each other against starvation by sharing food. For this purpose, evolution must have provided us with the mental equipment to operate what economists will recognize as an implicit insurance contract. I think this mental equipment survives as the deep structure of the fairness norms we use in everyday life.

My theory offers the basic similarity in structure between a food-sharing insurance contract and Rawls' original position as the reason that our intuitions find the device of the original position so satisfying. However, just as the deep structure of language is universal in the human species but the languages spoken in different countries are different, so the deep structure of fairness norms can be universal without the outcomes that are regarded as fair always being the same. What I think differs between different contexts or cultures is the standard of interpersonal comparison—without which the device of the original position cannot be operated.

Interpersonal comparison. If one asks what would be the fair way to split the proceeds when Adam and Eve sell the flower shop they have been running together, people want to know all sorts of things about the context. Are Adam and Eve married? Who invested how much when the business was set up? How many hours did each partner work per day? Who has children to support? Who brought what skills to the enterprise? Who attended the rotary luncheons? Who made unwelcome advances to the customers? It hardly needs to be said that such questions are unlikely to be regarded with the same importance in different cultures.

My approach appeals to Harsanyi's [9] work on what I call empathetic preferences to reduce the kaleidoscope of possible considerations that might arise when making interpersonal comparisons within a given context to a pair of social indices to be assigned to Adam and Eve. With some heroic simplifying assumptions, one is then led to outcomes that can be described using ideas from game theory.
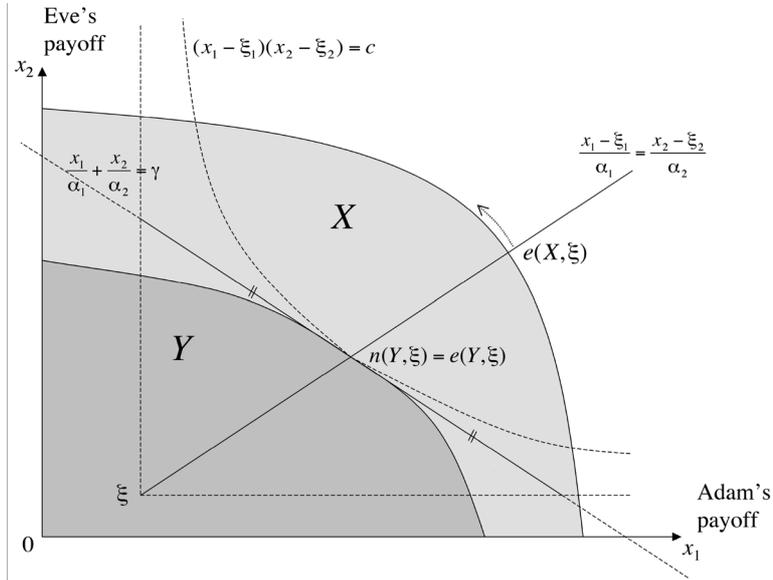
Figure 1: The set Y summarizes cultural history. In the medium run, cultural evolution adapts the social indices α_1 and α_2 until the egalitarian bargaining solution $e(Y, \xi)$ coincides with the Nash bargaining solution $n(Y, \xi)$. When a new feasible set X arises in the short-run future, fairness issues are settled using the egalitarian bargaining solution $e(X, \xi)$ with the old social indices α_1 and α_2. If X typifies the medium-run future, these social indices will gradually adapt as the bargaining outcome drifts towards $n(X, \xi)$.

Bargaining. Two-person coordination problems can often be regarded as Nash bargaining problems, which are described by a convex, compact set X of feasible payoff pairs and a deadlock point ξ. The Nash bargaining solution $n(X, \xi)$ attempts to predict the outcome if Adam and Eve negotiate on equal terms, each seeking to maximize their individual payoffs without concern for fairness issues. In the absence of outside options, $n(X, \xi)$ is the point x in X at which the Nash product

$$(x_1 - \xi_1)(x_2 - \xi_2),$$

is maximized subject to the constraint $x \geq \xi$.

With n players, the Nash product has n factors. Without a symmetry assumption, each factor $(x_i - \xi_i)$ is raised to a non-negative power p_i, which is said to be the bargaining power of player i. One application of the asymmetric version of the Nash bargaining solution arises when a large collection of players who treat each other on equal terms can be split into a small number of types. If players of a given type are identical, one finds the payoff x_i that each player of type i receives by locating the point x that maximizes the asymmetric Nash product in which the bargaining power p_i is the number of players of type i.

4

To obtain a bargaining solution that takes account of fairness, it is necessary to allow for some interpersonal comparison of utility, which Nash explicitly excludes. Various axiomatic schemes intended to capture our intuitions about fairness agree on characterizing the fair outcome of a Nash bargaining problem as the egalitarian (or proportional) bargaining solution $e(X, \xi)$ when full interpersonal comparison is required.[4] The egalitarian bargaining solution awards each player the maximum feasible amount consistent with their gains over the deadlock point being proportional to parameters that I call their social indices (Figure 1). It therefore enjoys some empirical support from psychological equity theory.

My theory uses the (symmetric) Nash bargaining solution to predict the outcome of bargaining in the original position. In the absence of any external enforcement, the theory then shows that Adam and Eve will emerge from behind the veil of ignorance to find themselves dividing the surplus according to the egalitarian bargaining solution.[5]

Determining social indices. I assume that social indices are fixed in the short run, so that a new coordination problem will be solved using the egalitarian bargaining solution with whatever standard of interpersonal comparison is current in the relevant context. However, the current standard of interpersonal comparison is seen as the product of cultural evolution, which I regard as operating in the medium run. Over this time span, it is assumed that the standard of interpersonal comparison adapts to whatever coordination problems were faced in this period. Cutting a long story short, the players' social indices α_1 and α_2 adjust until the egalitarian bargaining solution $e(Y, \xi)$ is identical to the Nash bargaining solution $n(Y, \xi)$, where Y is the average of the feasible sets encountered regularly in the past.[6]

Figure 1 illustrates how one determines the social indices α_1 and α_2 given that cultural history is known to be summarized by the set Y. First compute $n(Y, \xi)$. Draw a straight line from the point ξ to $n(Y, \xi)$. The slope of this line is α_2/α_1. In the short run, fair interactions that arise in the same context but with a new feasible set X will be determined by the egalitarian bargaining solution $e(X, \xi)$.

3 Outside Options

Even in the two-person case, Nash's classic formulation of a bargaining problem is seldom adequate to describe all of the relevant issues (Binmore [5]). In using

---

[4] Both the zeros and the units on the players' Von Neumann and Morgenstern scales are then compared. The chief rival to the egalitarian bargaining solution is the utilitarian bargaining solution, which ignores the location of zero points.

[5] With external enforcement, one is led to Harsanyi's [9] conclusion that the surplus will be divided according to the utilitarian bargaining solution.

[6] It happens that the utilitarian bargaining solution $u(Y, \xi)$ with social indices α_1 and α_2 is then located at the same point, which perhaps explains why utilitarians and egalitarians seldom seem to differ much on matters of practical policy.

the methodology described in the previous section to estimate social indices, other considerations may have to be taken into account. For example, are the payoffs stocks or flows? How do sunk investments influence the deadlock point? To what extent can binding commitments be made? Do the players have outside options?

In the next section, it is important that the payoffs are taken to be income flows measured in utils per unit time. The deadlock point is then the income pair the players enjoy while negotiating in the process of reaching an agreement. Investment is not considered at all here (Binmore [3]). Nor are binding commitments allowed. However, we cannot neglect outside options when seeking to expand the number of players.

A player's outside option in a bargaining problem is the largest income the player can enjoy if the negotiations irrevocably break down because he looks elsewhere for a deal. The breakdown point η consisting of the players' outside options need not be equal to the deadlock point. It will normally be a Pareto improvement on ξ.

The Rubinstein [16] bargaining model with a vanishingly small interval between successive offers predicts that the bargaining outcome with outside options is not $n(Y, \eta)$ as commonly assumed. The bargaining outcome is $n(Z, \xi)$, where Z is the part of the feasible set Y in which both players get at least their outside options (Figure 2). If η lies outside Y so that Z is empty, no rational agreement is available.
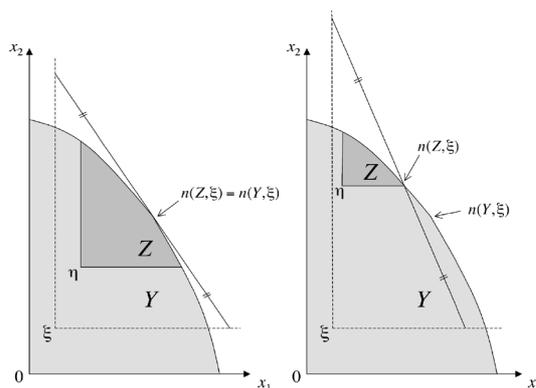
Figure 2: In the left diagram, $n(Y, \xi) > \eta$. Introducing outside options then leaves the Nash bargaining solution unchanged. In the right diagram, $n_2(Y, \xi) < \eta_2$. Eve then gets no more than her outside option. If η lies outside Y, rational agreement is impossible.

Three players and only three cakes. The preceding discussion of outside options makes it possible to analyze the case in which only two out of three players can productively reach an agreement. Binmore [1] shows that only the conclusions illustrated in Figure 3 can be defended as subgame-perfect equilibrium outcomes in a suitably generalized Rubinstein bargaining model.

6

The left diagram of Figure 3 is a generalization of the symmetric case studied by Von Neumann and Morgenstern [19] for three-person, zero-sum games. One of the three outcomes in the Von Neumann and Morgenstern triple $\{v_{12}, v_{23}, v_{31}\}$ will be observed. If the coalition $\{ij\}$ forms to implement v_{ij}, the relevant outside option for player i is the payoff he would receive if v_{ik} were implemented as a consequence of his abandoning player j for player k. Similarly, the relevant outside option for player j is the payoff she would receive if v_{jk} were implemented as a consequence of her abandoning player j for player k. In equilibrium, all these outside options are active and the Nash bargaining solution ends up being applied to feasible sets Z that are reduced to a single point.
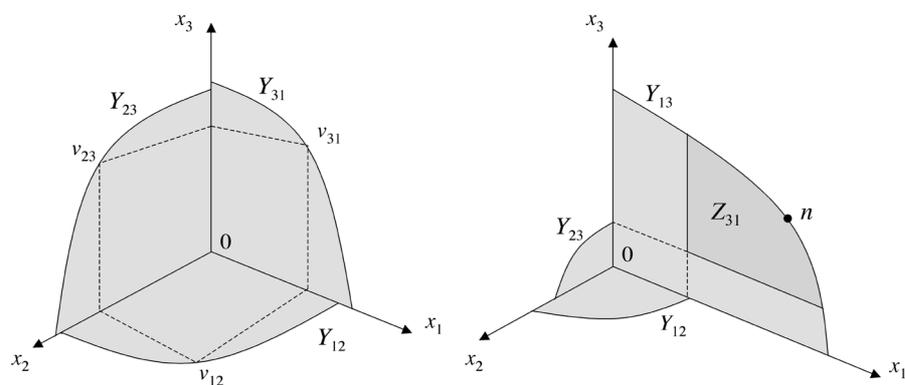

Figure 3: The figure shows a case in which the coalition of all three players cannot form. For simplicity, the deadlock point is placed at the origin, and individual players have no external outside options. However, outside options arise endogenously because when two players make a deal, they always have the alternative of proposing a deal to the third player who wil otherwise be left with nothing. The feasible set Y_{ij} is what is available to the coalition $\{ij\}$. The left diagram shows a case in which a Von Neumann and Morgenstern triple exists. When such a triple $\{v_{12}, v_{23}, v_{31}\}$ exists, the final outcome is one of its three elements. The right diagram shows a case in which no Von Neumann and Morgenstern triple exists. Only one coalition can then form. In the figure, this coalition is $\{31\}$, and the final outcome is the Nash bargaining solution $n = n(Z_{31}, 0)$.

The right diagram of Figure 3 shows a case in which no Von Neumann and Morgenstern triple exists. Only one outcome bargaining outcome n is possible, which is the Nash bargaining solution of the problem $(Z_{31}, 0)$. The relevant outside options are found by locating the most that the excluded player can offer the two included players.

An example of the second case arises in a market with two risk-neutral buyers (players 1 and 2) and one risk-neutral seller (player 3) of a single indivisible object. In this case, $Y_{12} = \{0\}$. Player 3 sells the object to player 1 if he values it more than player 2. If player 2's valuation is sufficiently high, then the relevant outside

option is active and the object is sold to player 1 at player 2's valuation (as in an English auction). Otherwise, the selling price is determined by the Nash bargaining solution of the problem with player 2 excluded.

4   Communities with Three Players

The case of three players and four cakes—the fourth being what is available to all three players acting together—is not amenable to analysis using a generalized Rubinstein bargaining model, because multiple subgame-perfect equilibria emerge unless one restricts the possibilities by admitting only Markov equilibria. I can therefore no longer appeal to an orthodox noncooperative analysis in defending my continued use of the Nash bargaining solution. Nor is it clear how the potential behavior of coalitions of only two players affects breakdown and deadlock issues when all three players bargain together. One approach to the breakdown question is to restrict attention to outcomes that lie in the core of the game.

Core?   The game shown on the left of Figure 3 has an empty core. The game on the right has a non-empty core, namely the Pareto frontier of the set Z. Even when it is non-empty, the core is commonly criticized as a cooperative solution concept on the grounds that the players would need to be myopic to confine their attention to outcomes inside the core. It may be that outcome y can be blocked by coalition S because S can enforce z, and everybody in S prefers z to y, but why should S actually block y if z itself will be blocked by another coalition?

This question seems fatal to the concept of the core when the players are modeled as perfectly rational. However, it ceases to be so damaging when the mechanism that shapes the behavior of the players is evolutionary in character. Quotes like the following from Mark Ridley [15] are commonplace in evolutionary texts:

> Evolution, in Darwin's theory, is short-termist; it works opportunistically, with piecemeal changes and immediate advantages, not with long-term strategic plans.

Since my theory of interpersonal comparison assumes that the social indices relevant in a given context are a product of the cultural history of a society, I cannot therefore reject core arguments simply on the grounds that they assume the players to be myopic. The question is rather: what degree of myopia is it reasonable to assume?

I shall defend my answer only in the three-player case. The defense is based on the bargaining analysis of the previous section. Cultural evolution is assumed to take account of all the bargaining considerations that would be relevant if the players in a coalition were consciously to seek to negotiate a deviation from the current status quo y to a new outcome z, but does not take account of the new negotiation possibilities that would become available after z is established as a new status quo. Even if the players were aware of such future possibilities for negotiation, they might largely discount them on the grounds that cultural evolution operates too slowly to be relevant to their present concerns

8

To be more precise, suppose that the set-up illustrated in Figure 3 is augmented by including a feasible set Y_{123} that is available to the grand coalition $\{1, 2, 3\}$. Imagine that an outcome y in Y_{123} has become established so that each player i is currently receiving the income y_i. It may now be that a pair of players (possibly myopically) see some advantage in excluding the third player. For example, if $(y_1, 0, y_3)$ lies in the interior of Y_{31}, players 1 and 3 might see themselves as facing a bargaining problem with deadlock point (y_1, y_3). To assume that they will then automatically disrupt the established outcome y neglects the influence that player 2 will have on the negotiations between players 1 and 3 as their source of outside options. If we take these outside options into account, as we surely should, stability considerations lead to the set Z of Figure 4 (which includes the core as a subset).
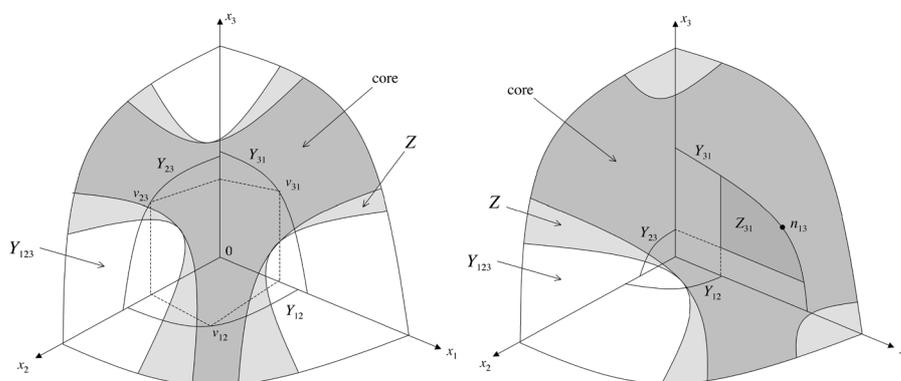
Figure 4: The core is heavily shaded in the two diagrams.

When the set Z is empty, the theory predicts that the three-player coalition will not form at all. Some players will then be seen as being unworthy of fair treatment. Human beings held as slaves in times gone by are an example.

No coalitional history. Both Harsanyi [9] and Rawls [13] regard the possible formation of coalitions other than the whole society as irrelevant to their conceptions of a just society. Citizens in the original position are therefore envisaged as negotiating as free-standing individuals.

The corresponding situation in my evolutionary theory arises in contexts when coordination within groups of players smaller than the grand coalition of all players can be ruled out. In prehistoric times, for example, all the grown men in a hunter-gatherer society presumably acted together when hunting a large animal like a mammoth. If sufficient men were unavailable, they presumably sought smaller prey.

In such a case, a generalization of the two-player analysis to the multi-player case seems relatively easy. For the examples shown in Figure 4, one simply applies the three-person Nash bargaining solution to the feasible set Z, taking the deadlock point to be the origin. In a society with many players divided into three types, one

would operate in a similar way (with the three axes now recording the payoff to each type), but now the set Z would be more complicated and it would be necessary to use an asymmetric Nash bargaining solution with bargaining powers set equal to the numbers of players of each type.

Coalitional history. I think Harsanyi and Rawls underestimate the problems that one has to face in putting together the beginnings of a realistic theory of inter-personal comparison. In particular, one can seldom hope to give an acount of the cultural evolution of some aspect of a society that ignores the history of conflict and cooperation between different factions or subsocieties within that society. Consider, for example, the decision that slaves should count as twothirds of a full citizen when determining constituencies for the original House of Representatives.

Figure 5 illustrates a case that might arise when the grand coalition is formed by accretion, with player 2 joining a previously existing coalition of the other players (perhaps as a result of the set Y_{123} having expanded in size). However, matters become more complicated when the question of which coalition of two players forms first is determined by factors outside the domain of the model (as in the left diagram of Figure 3). When we move to the multiplayer case, such historical accidents will be dominant in most cases, leaving little that game theory can usefully contribute.
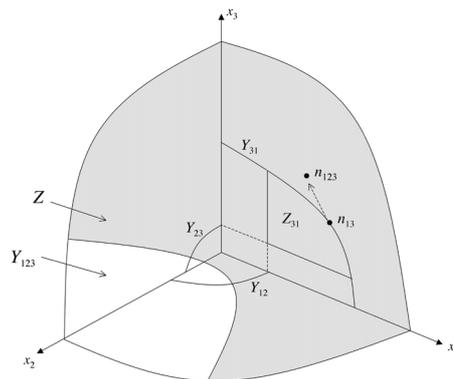
Figure 5: The outcome n_{123} is the Nash bargaining solution of the bargaining problem in which the feasible set is Z and the deadlock point is n_{31}. A cultural history that led to social indices being computed using n_{123} might begin with a society in which player 2 was treated as an outsider, and so excluded from fairness considerations. It may be that the set Y_{123} then expanded, with the eventual result that player 2 began to be treated as an insider to whom fairness norms apply. However, the players' social indices continue to reflect the history of social exclusion suffered by player 2 .

5   Conclusion

Our sense of fairness is simply a social tool that evolution has washed up on the human beach along with all the other flotsam and jetsam that make our lives possible. I share the aspirations of those who would like to see fairness applied much more widely than at present, but we shall not be successful in extending its range if we view the way it works at present through the rosy-colored spectacles of orthodox moral philosophy.

For example, everybody now understands that it is impossible to meet all the demand for health care. Public health systems must necessarily refuse some of the demands made upon them. For such refusals to be politically acceptable, they need to be perceived as fair by the general public. A government might seek to meet this aim by employing philosophers to instruct the public in the social indices they ought to use when employing the device of the original position, but if I am right that the social indices we actually use as inputs when making small-scale fairness judgments are culturally determined, then nobody is going to consent to a proposed rationing scheme on fairness grounds if it seems unfair to them according to established habit and custom, no matter what may be preached from the pulpit.

This paper reviews the difficulties of seeking to estimate social indices indirectly by looking at the cultural history of a society. It is possible to carry through some useful exercises in comparative statics in the case of only two players, but matters quickly get out of hand when more than two players are considered. The conclusion would seem to be that there is no alternative but to embark on empirical studies, both in the laboratory and in the field. However, even within social psychology, researchers seem to have lost sight of the importance of interpersonal comparison to problems of justice and equity. It would be particularly valuable if experimental economists were to bring their specialized expertise to bear in making modern equity theory into a useable applied tool.

References

[1] K. Binmore. Nash bargaining theory I. In K. Binmore and P. Dasgupta, editors, *Economics of Bargaining*. Cambridge University Press, Cambridge, 1987.

[2] K. Binmore. *Just Playing: Game Theory and the Social Contract II*. MIT Press, Cambridge, MA, 1998.

[3] K. Binmore. *Natural Justice*. Oxford University Press, New York, 2005.

[4] K. Binmore. The origins of fair play. *Proceedings of the British Academy*, 151:151–193, 2007.

[5] K. Binmore. *Playing for Real*. Oxford University Press, New York, 2007.

[6] M. Deutsch. *Distributive Justice: A Social Psychological Perspective.* Yale University Press, Newhaven, 1985.

[7] J. Elster. *Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens.* Russell Sage Foundation, New York, 1992.

[8] E. Fehr and K. Schmidt. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114:817–868, 1999.

[9] J. Harsanyi. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations.* Cambridge University Press, Cambridge, 1977.

[10] E. Kayser, T. Schwinger, and R. Cohen. Layperson's conceptions of social relationships: A test of contract theory. *Journal of Social and Personal Relationships*, 1:433–548, 1984.

[11] M. Lerner. The justice motive in human relations: Some thoughts about what we need to know about justice. In M. Lerner and S. Lerner, editors, *The Justice Motive In Social Behavior*. Plenum, New York, 1981.

[12] M. Lerner. Integrating societal and psychological rules of entitlement: The basic task of each social actor and a fundamental problem for the social sciences. In R. Vermunt and H. Steensa, editors, *Social Justice in Human Relations I: Societal and Psychological Origins of Justice*. Plenum, New York, 1991.

[13] J. Rawls. *A Theory of Justice.* Oxford University Press, Oxford, 1972.

[14] H. Reis. The mutidimensionality of justice. In R. Folger, editor, *The Sense of Injustice: Social Psychological Perspectives*. Plenum, New York, 1984.

[15] M. Ridley. *Mendel's Demon: Gene Justice and the Complexity of Life.* Weidenfeld and Nicolson, London, 2000.

[16] A. Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50:97–109, 1982.

[17] E. Sampson. On justice as equality. *Journal of Social Issues*, 31:54–64, 1975.

[18] S. Schwartz. The justice of need and the activation of humanitarian norms. *Journal of Social Issues*, 31:11–136, 1975.

[19] J. Von Neumann and O. Morgenstern. *The Theory of Games and Economic Behavior.* Princeton University Press, Princeton, 1944.

[20] G. Wagstaff. *An Integrated Psychological and Philosophical Approach to Justice.* Edwin Mellen Press, Lampeter, Wales, 2001.

[21] E. Walster, E. Berscheid, and G. Walster. New directions in equity research. *Journal of Personality and Social Psychology*, 25:151–176, 1973.

[22] E. Walster and G. Walster. Equity and social justice. *Journal of Social Issues*, 31:21–43, 1975.

[23] P. Young. *Equity.* Princeton University Press, Princeton, 1994.