Analyse & Kritik 28/2006 (© Lucius & Lucius, Stuttgart) p. 13–31

Anthony de Jasay

Fairness as Justice

Abstract: The paper questions Binmore's identification of justice with fairness and his corresponding focus on bargains to the neglect of conventions, notably of ownership. Section 1 deals mainly with the role ascribed to man's earliest genetic heritage in shaping fairness norms and the putative effect of such norms on bargaining solutions. Section 2 argues that the scope of fairness as opposed to justice in determining the social order is quite narrow, It sketches a theory of fairness distinct from justice, derived from the principle of treating like cases alike.

0. Introduction

In the matter of moral theory, we have had, and are all too slowly getting over, a hefty dose of 'justice as fairness' and a number of lesser works that presented justice as something else than itself. There is now an original, fascinating and irritating construction by Ken Binmore that he calls *Natural Justice* and that treats (natural) justice as synonymous with fairness. The adjective refers to his passionate belief that Kant is an emperor clothed only in the obscurity of his language and that "to pretend that the good and the right are anything other than the products of human evolution is to abandon all hope of making sense of human morality" (93).[1] Natural justice is a set of efficient equilibria selected by maximising man guided in their selection by fairness norms, the deepest of which he has inherited from pre-history where it proved best for the survival of his genes. Each equilibrium is conceived as a bargaining solution, and each is called a "social contract"—a somewhat unusual usage of the term.

Bargaining problems can have many solutions. Binmore insists that what 'fairness norms' do is to help the players select one out of the many available. Fairness norms are thus instrumental both in finding a solution and in ensuring that the solution will be a fair, that is to say 'naturally' just one. However, if bargaining selects the fair solution, and fair is what bargaining selects, we are mired in a tautology. Binmore's central thesis is meant to pull us out of it. The thesis identifies fairness norms independently of the fact that they are being chosen. It affirms that every fairness norm actually used in bargains harks back to the ancestral, and egalitarian "Deep Structure" (18) that we are genetically "hard-wired" to follow. This thesis, it seems to me, is ridden throughout the book much harder than it can bear, and is the source of some legitimate doubt

[1] Page numbers without further reference generally refer to Binmore 2005. [Editors' remark: This citation practice will remain the same throughout this issue.]

about the kind of theory we are being offered. As a descriptive theory, it negates the Deep Structure, for it concedes that the really important equilibria of the overall social order and its justice are not such as his thesis would predict. (Even the residual claim made on behalf of the relatively unimportant small-scale 'social contracts' is contestable; cf. *'After You'* below). To have the Deep Structure prevail, the theory veers sharply to the prescriptive, the counterfactual, the imaginary *als ob*.

One is led to question whether Binmore's large one-way bet on the Deep Structure and consequently on the albeit implicit primacy of biological over cultural evolution has been judiciously placed. Some reasons for holding that the bet is not a winning one are marshalled in Section 1. Section 2 describes an alternative, and more modest, bet with shorter odds speaking for it.

There are two great absents from this book on 'Natural Justice'. Considering its object and its vast scope one should have expected both to play a noticeable role. One is convention in general. Equilibria tend to appear in Binmore's book as mostly bilateral bargaining solutions, 'social contracts' and not as products of unilateral adhesion to what others happen to be doing, though that may be no more than an only just discerned, incipient pattern of conduct. Yet one wonders whether the convention is not more important in shaping the social order, as well as human morality, than the bargain. Note that conventions are equilibria explicable in terms of ordinary payoff-maximising and as far as one can see, do not require fairness as an equilibrium selection device.

The other great absent is ownership, itself a special convention or rather a group of conventions. The omission seems odd in view of Binmore's veneration for Hume that one is happy to share. Was it not Hume who clearly named the 'stability of possession' and 'its transference by consent' as forming, jointly with 'the keeping of promises', the starting line of society, antecedent to the state and any sort of 'original contract'? I find no ready explanation for leaving ownership and property out of 'natural' social theory, nor for their virtual absence from most of game theory in general. One can get weary when shown distribution games cleverly plotting the division of a game sum, a 'cake to be sliced' that has just fallen from heaven for no reason at all, (when it is not a research grant to fund game theory experiments). This is not dealing with the facts of life as they are (as Binmore would have us do). Cakes have been baked by somebody and are owned by somebody. The owner is typically forgotten, reduced to let others slice his cake. Should he not be made to play, too, if the distribution is to mimic the 'game of life'?

1. The Deep Structure

Hunter-gatherer bands that can still be found today in inhospitable corners of the Earth where human beings have no business to dwell, appear to be living by a share-and-share-alike rule where all hunt-and-gather according to their abilities and mostly according to their luck, and all share the food according to need, though with a bias in favour of the hunter's nearer kin so as to promote the

reproduction of his own genes. There is a plausible conjecture that those of our hominoid and prehistoric human ancestors whose progeny has survived have practiced much the same mode of distribution. One might add the unfalsifiable hypothesis that those who did not share the carcase of the occasional mammuth have failed to survive, so that we survivors only have genes that predispose us to share good things, and bear bad ones, more or less equally. Though equal sharing as a form of insurance is no longer the best, let alone the only survival strategy, our genes still keep telling us that it is. We may believe that it is our moral intuition that dictates egalitarian leanings, but it is in fact simply genetics that does it.

Binmore calls this egalitarian leaning the "deep structure" of fairness, the original norm that is the root of all fairness norms (18) both when they spring from biological and from cultural evolution. (Why the latter would be conditioned by 'communist' genes in the Deep Structure is not really evident, but in any event Binmore seems to hand cultural evolution the second fiddle).

Ascribing to the genetic memories embodied in the Deep Structure the egalitarian propensity so manifest in most societies living by some democratic decision rule is an interesting idea. However, it seems to this reader that it calls quite audibly for Occam's razor. A less conjectural and far simpler explanation is at hand. Goods are defined as objects we would rather have more of than less. Most people in the nine lower deciles of a distribution would be quite content to have the top of the wealth of the tenth decile sliced off and redistributed to them. Once this was done, people in the eight lower deciles would be content to see the same thing done to the tenth and ninth deciles and so on to the tenth, ninth and eighth deciles until finally all the 'tall poppies' have been chopped off and wealth is evenly spread. Without listening to their genetic memories of between 100.000 and 10.000 B.C. and the Deep Structure, most of the beneficiaries would probably think that this move toward equality was the fair thing to do.

Sharing as an essential survival strategy is now obsolete, whatever other merit it may have. However, in Binmore's theory the Deep Structure continues to play the key role in what he disarmingly calls a 'stylized' manner but that is in effect a bold leap from the factual to the counterfactual.

In the Deep Structure as it may have existed, it was incumbent on every hunter-gatherer to share his success when he was successful. He did not know (or so we suppose for the sake of the theory) whether he was more likely to be successful than unsuccessful tomorrow or thereafter. He did not know, in other words, whether the duty to share success would work to his and his family's disadvantage (if he had to share his own success with others) or advantage (if others shared their success with him). He could not expect to do better unilaterally by deviating from the Deep Structure than he was expecting to do by remaining committed to it. Therefore the latter was an evolutionary equilibrium, and stable as long as hunting-gathering remained the best way to make a living.

For Deep Structure, time-bound and subject to obsolescence, write Original Position, timeless and imagined to last. Its properties have been made familiar by all the Tibetan prayer mills in academe. Suffice it to say that instead of

the realities of the bush or the ice bank, it is an imaginary 'veil of ignorance' that conceals from you whether you are more likely to be a successful than an unsuccessful hunter in what Binmore calls the 'game of life'.

1.1 The Original Position On Trial

The device of the Original Position (though not the name) was first employed by William Vickrey (1945) and later by John Harsanyi (1977) as a place from which an individual should best start when trying to make welfare judgments. Vickrey's most elegant use of the device was for long very unjustly ignored. Admittedly, he aimed it merely at locating the maximum of someone's social welfare function—a limited aim. It is not evident that the problem admits a more ambitious one. Harsanyi has erected a grander edifice of utilitarian theory, making incomplete information serve a moral purpose. Both these exceptional minds were perfectly clear in presenting the device as a tool of moral theory, serving to evaluate states of affairs in a way that abstracts the evaluation from the person who is performing it. Neither pretended to be describing or explaining real coordination or real conflict reflecting individuals' maximising strategy under 'natural' conditions.

Acting as the spiritual heir of the factual Deep Structure, the counterfactual Original Position in Binmore's natural scheme of things is meant to explain how fairness helps to select bargaining equilibria in real-life situations. In this incomparably more exacting role the device must be seriously tried to see whether it can play it without provoking derision, and without being decisively upstaged by other, less counterfactual means of achieving equilibrium from plausible strategies.

Fairness norms are said to constitute natural justice that does not hang by metaphysical suppositions, understandably dismissed by Binmore as "skyhooks" (148). If that is the case, why should the able, strong, clever, tenacious, handsome, sexy, rich and well-connected choose to 'enter the Original Position'? To plain readers, the Original Position is about the last place where such well-endowed persons would choose to go to strike their distributive bargains. No doubt there are high-minded individuals who listen to 'the whisper in us' and lean over backwards not to profit from their natural advantages But why should all the well-endowed go along with what the high-minded among them would consider a fair equilibrium?

We cannot say how much of an advantage their superior endowment would gain them when bargaining outside the Original Position. Surely, however, it could not possibly work to their disadvantage compared to how they would fare in the bargain struck inside it. If so, for the genetically or culturally better endowed, staying outside would weakly dominate going inside. The voluntary population of the Original Position would then be mostly recruited from the ranks of the ill-endowed and the high-minded. They would derive little benefit from agreeing on a distribution among themselves, and the Original Position would come close to looking pointless or farcical.

Suppose, however counter-intuitively, that all the well-endowed did enter it and bargained over distribution with the ill-endowed, with everybody's endowments being shrouded by the self-imposed 'veil of ignorance'. Suppose also that this bargain was 'fair' in Binmore's sense, either or both because the very fact of its being selected revealed it as being such, and because of its resemblance to the Deep Structure that our genetic memory deems to be fair. Suppose last, as it seems reasonable to do, that when the parties throw away the 'veil of ignorance' and recover awareness of their proper identities, the well-endowed find that under the bargain they are worse off than their endowments would entitle them to expect. Would keeping the bargain be compatible with the assumption that the parties are payoff-maximisers? (Remember that this foundational social contract is negotiated once and for all, so the folk theorem is of no help in producing compliance). Binmore looks like being in two minds about this. He says that post-contractually a utilitarian distribution would require external enforcement, but an egalitarian one would not (174). It is hard to comprehend why this should be the case. In both types of distribution, some individuals must be forgoing some part of their potential payoffs and could unilaterally capture them. (If they had no wish to do so, what was the point of having them bargain in the Original Position rather than outside it?) What would stop these 'suckered' people, freed of the 'veil', from repudiating the bargain concluded while they were blindfolded? To imply that they would not do so is to assume that passage through the Original Position has 'hard-wired' into them a commitment to honour bad bargains concluded there. It is not evident that Binmore is making this assumption. His affirmation that a utilitarian bargain would require enforcement suggests that he is not. Nor would it sit well with what the plain man would understand by 'natural justice'.

Binmore assures us that emerging from behind the 'veil' and finding themselves in an egalitarian state of affairs, people would be motivated not by their personal, but by their 'empathetic' preferences and would not wish the 'phantom coin' to be tossed again. The reason why this should be so is also hard to comprehend. Why do people not seek to maximise their expected personal utility?

The plain reader is left to wonder why and how the social contract emerging from the Original Position is an equilibrium, or indeed what the Original Position can be expected to achieve.[2]

---

[2] In Rawls's theory, people are driven into the Original Position by the moral force of an idea of fairness under which differential endowments that are not positively deserved are undeserved. *Tertium non datur.* Once inside the position, they do not maximise expected payoffs, but play maximin. Thus an egalitarian bargain is produced. The theory, however, provides a short cut to the same result. The preference function for 'primary goods' suffices to generate it. Everyone wants the same minimum bundle of primary goods, neither more nor less; he "cares very little, if anything for gains above the minimum" but "can hardly accept" anything less (Rawls 1972, 154). If this is what all want, there is no reason why they should settle on anything else, whether they bargain in the Original Position or out of it. They would, as a practical matter, have the 'difference principle' beavering away to excuse or iron out the unwanted inequalities. The rigmarole about fairness and the veil of ignorance may all be thrown away as redundant.

1.2 'After You'

A group of people wishes to go through a door where only one person can pass at a time. In the group there are men and women, young and old, there is a dowager duchess, a lavatory attendant and all ranks in between. Any random pair among them is in a two-person game where two strategies can be chosen, Before You and After You. Arriving at the door, if both play Before You, they get stuck in the doorway. If they both play After You, they never even try to pass through it. The game sum is maximised if one plays Before You when the other plays After You. However, its division between the Before You and the After You player is not markedly different, for it is only slightly better to pass first than second. The game is mainly one of coordination and only to a minor extent distribution. The main thing is to achieve a protocol of at least rough-and-ready precedence, so that people will have some idea with whom they should play Before You and with whom After You.

Binmore sets us a puzzle when he affirms that in this and similarly 'picayune', small-scale problems to be solved by a 'social contract', people in effect use the device of the Original Position (21, 52). This means that the protocol of precedence is established as if everybody had an equal chance of being anybody (including herself) in the group. If so, the dowager duchess would not be particularly interested in having a protocol that gave precedence to dowager duchesses, and the lavatory attendant would not much mind if lavatory assistants (or at least the youngish ones) were assigned the last place. Being in the Original Position would leave everyone quite clueless about what protocol to go for, since every feasible one is as good as every other to a person who has an equal chance to fare well and to fare ill under each protocol. Far from being the equilibrium selection device that Binmore makes it out to be, the Original Position (respectively the Deep Structure) offers no guidance whatever to the players as to which equilibrium to select and which one would be fair. It could be argued that once they become indifferent among all feasible protocols, it becomes easy to accept one. However, the selection problem remains unsolved, and possibly more insoluble than if the players had preferences to bargain with. Whatever reducing alternatives to indifference nay accomplish, it does not select the fair protocol.

One could speculate how the protocol did get established, for there is ample evidence that it exists, albeit in slightly different variants. The answer probably is: like every other convention.

While asserting that small-scale 'social contracts' are in actual fact concluded as if the Deep Structure were shaping them, Binmore expresses the wish that large-scale 'social contracts' were also shaped by it. Apparently, they are not, and he omits to tell why they do not.

1.3 Agreement and Adhesion

It may be helpful in understanding the conceptual distinction between contract and convention to observe that a coordination equilibrium is reached by one of two routes. The first has, as it were, three distinct phases. It starts with the parties deciding to enter into negotiation. The second phase, if there is one, brings the agreement. If the agreement provides for reciprocal performances ("consideration" being the legal term), it is a contract. The third phase leads to execution, which may be simultaneous or provide for one party performing first, the other second. Execution may also be once-for-all or repeated finitely or indefinitely. Most games follow this model. Binmore's theory of fairness as 'natural' justice follows it exclusively; every equilibrium on which people coordinate is a 'social contract' issuing forth from a bargaining solution. But as the example of the After You game discussed above shows, at least some of these bargaining solutions are implausible and could not produce the equilibrium that empirical evidence shows to exist.

The other, and distinct, road to equilibrium is the emergence of a convention—emergence that could logically be instantaneous but is most likely to be so gradual that it is initially hard to perceive.

While bargaining solutions presuppose an intent to agree, conventions are adhered to without anybody agreeing with anybody else. Nobody intends to initiate them. They may be imagined to start from some random bunching of behaviour into a patterned subset within a patternless set of behaviour of the population. Adhering to thee pattern yields a higher payoff even if it is not efficient, and a fortiori if it is. Thus thee pattern attracts adherents and the payoff of adhesion is likely to increase as the patterned behaviour gains over the patternless one, until all or nearly all the population adheres and the convention is fully established and rises to the status of a rule. However, whether this simple-minded sketch is a plausible account of the emergence of conventions or not is of little importance. For unlike bargains that are too often mere inferences from given incentives and may or may not take place, established conventions manifestly do exist and do their work whichever way they have originated.

The work they do takes care of the lion's share of arranging and maintaining the social order. Conventions exist against torts in the widest sense. They forbid killing, maiming and sequestration except as punishments in their defence, There is a set of conventions against trespass, theft, robbery, wilful damage, fraud and usurpation of title to property, and against breaches of promise. We may conveniently call this set the convention of ownership, or of title. Of lesser gravity are the countless conventions of civility, of which the After You is a perhaps 'picayune' but typical example. Conventions between them create a rule system that may not be seamless yet goes quite far toward rendering human coexistence at least feasible. With a little luck, it may also render it 'commodious'.

The striking thing seems to be that all these conventions are equilibria (though the self-enforcing character of some of them is not obvious to the naked eye) in which fairness plays little or no part. It is not instrumental in selecting the convention. Nor is fairness a result, a fruit of the behavioural rule the con-

vention represents. As far as one can judge, the bulk of conventions is neither fair nor unfair.

The reason is presumably that they are mostly solutions of coordination problems with no or only scant distributive effects. We shall revert to this aspect in Secton 2, but we may already discern a tentative answer to Binmore's question of why the Deep Structure, (or should we say the Original Position?) is employed to generate 'social contracts' in small-scale and 'picayune', but not in large-scale and vital 'games of life'. Where conventions are the solutions that have emerged and are in force, there is no vacant room for bargaining solutions to fill and the question of their fairness can barely arise.

1.4 Goats on the Commons

Finding an efficient 'social contract' where some of the most essential conventions that together secure title to property, are implicitly treated as if they were not available 'strategy' options, may be the perfect testing ground for the capacity of fairness to play the role Binmore's theory of fairness as justice assigns to it. The contract is for the management of the commons. All efficient solutions would put exactly 1,000 goats to graze on it. One hundred families enjoy free access to the commons for their goats to graze on costlessly. One of the efficient solutions, the fair one, "shrieks for our attention" (14): Subject only to their particular circumstances, each family should be allowed to graze 10 goats on the commons.

Binmore is perfectly aware that an agreement among the families to do this is not an equilibrium, for he mentions that it would need an external enforcer to police it. But perhaps understandably, he believes that if the game of managing the commons were (indefinitely) repeated, the fair solution would turn out to be self-enforcing. The reason is found in the folk theorem.

However, the inconvenient empirical facts tell us that though the 'game' is repeated every spring when the families put out their herds (including the annual crop of new kids) to pasture, the commons remains chronically overgrazed and no efficient solution, fair or unfair, is found for it. An inefficient equilibrium may be reached when the pasture is so degraded that the annual increment of the herds is just balanced by the goats that starve to death. Historically, the 'solution' used to be enclosure, i.e. the exclusion of the villagers with or without adequate compensation. This was solving the problem efficiently by abolishing it.

The point to note is that if each family keeps no more than 10 goats on the pasture and does never or only rarely exceed this number except by agreement with another family, by which one family trades its grazing quota to the other, then the commons has ceased to be common. It has become a joint tenancy, where each joint tenant has a 1 per cent equity in both the grazing land and the herd. It can trade his equity up or down with another joint tenant. The 'game strategy' of respect by every tenant for the equities of every other is the convention of title or ownership. It may be coupled with a contingent strategy that 'kicks in' to respond to deviation and punishes deviants. However, owner-

ship is in absolute contradiction to the principle of the commons: it is exclusion, not free access. Yet neither the Deep Structure nor any other fairness norm one could think of could serve as a substitute for it. The empirical evidence testifies that none has done so.

1.5 Interpersonal Arithmetic

Looking at the Nash Demand Game, Binmore reports with his characteristic honesty that extensive experiments with this game have produced a large variety of solutions which were all claimed to be fair by the subjects on debriefing, but which showed no tendency to bunch anywhere near any such fairness norm as reference to the Original Position might conjure up (74). What is one to make of such evidence?

Binmore's answer is that the fairness of a bargain turns not on how it distributes dollar sums, but how it distributes utilities (27). For all he knows, the wide variety of solutions reported by the respondents as fair in dollar terms may in fact converge closely to recognized fairness norms, such as the egalitarian or the utilitarian distribution, when translated into utilities.

Dollars are visible and tangible payoffs. Utilities are invisible entities that we impute to people's state of mind to explain their choices and if the explanation is complete, choice reveals its own contribution to utility. The relation must be rescued from being a redundant tautology before it can be of any use, and for the present I shall take it for granted that this has been done. Obviously, however, in order to assess bargains in terms of utility payoffs, more is required.

Binmore finds it laughable that some people, notably 'diehard neo-classical economists'[3] declare interpersonal comparisons of utility to be impossible. If this were what they declared, they would be wrong. Anything can be compared to anything else by reference to a trait or traits they have in common. Apples can be compared to pears by referring to their shape, goats to pigs by their smell, a Shakespeare sonnet to the Chrysler Building in New York by referring to their respective structures. However, while all things can be compared, not all are commensurate. The utilities accruing to a person from various choices or states of affairs are commensurate both on ordinal and on cardinal scales. The person's choices of prospects can reveal both that one utility is greater than another and by how much greater. The same inferences cannot be drawn interpersonally. Each person's choices testify only about the utility of that person and not about the utilities of other persons.

Binmore hides this impossibility by an astonishing sleight of hand. His start is perfectly acceptable. In his Meeting Game, he posits one cardinal scale of Adam's personal utilities and one of Eve's personal utilities attached to each

[3] One must suppose that Binmore confines the term "neo-classical" to (welfare) economics after Lionel Robbins. Much neo-classical economics is, of course, pre-Robbins and indulges freely in 'interpersonal comparisons of utility'. The arch-neoclasical A. C. Pigou believed in the 'diminishing marginal utility of money' and deduced from it that taking it from the rich and giving it to the poor increased 'total utility'. If "die-hard neo-classical" meant post-Robbins, the present writer would be pleased to be counted as such a one.

outcome of their bargaining over the choice of a meeting place. So far, all is well. Then, with a blinding manoeuvre that leaves the reader rubbing his eyes, he denominates both the 'Adam-utils' and the 'Eve-utils' simply as 'utils' *tout court*' as if they were units of the same entity, like temperature or distance. The two personal utilities have thus been rendered homogenous and he can perform upon them arithmetic operations of various kinds without manifesting any of the unease that he would no doubt display if he were subtracting three goats from four pigs. (For instance, he multiplies alternative pairs of the two kinds of utils to find the highest product and thus locate the Nash bargain.)

Once he has put the problem of interpersonal arithmetic behind him in such short order, Bimore proceeds to deflate Adam's and Eve's 'utils' by different 'social indices'. Technically, this is as unobjectionable as converting Adam's temperature measured on a Fahrenheit scale, and Eve's temperature measured on a Reaumur scale, both into Celsius degrees. The intended effect is to make the arithmetic come out with the desired fairness norm, e.g. the egalitarian one, where Adam and Eve have the same indexed utility. Here, Binmore is achieving with his utility indices what he later accomplishes by adopting Harsanyi's empathetic preferences. The latter look less arbitrary than the indices, springing as they are said to do from man's ability to put himself in other men's shoes. In the process, Binmore enlists not only Adam and Eve, but also John (von Neumann), Oskar (Morgenstern) and Pandora, displays his renowned talent for clever diagrams, and never heeds Alfred Marshall's warning against 'long chains of abstract reasoning'.

There can be no quarrel with Harsanyi for setting out a rule-utilitarian moral theory and relying on the assumption that there is only one rational welfare judgment, He is well within his rights as a normative theorist to postulate that the judgment be made by introspection permitting the observer to imagine herself in everybody else's shoes Nowhere does he pretend that it is genetically 'hard-wired' in the observer's brain to do this and it is in some obscure sense in his continuing interest to be so 'programmed'—although it is aeons ago that the programme ceased to be a useful, let alone an evolutionarily stable survival strategy.

By transcribing the prescriptive theory of Harsanyi into the descriptive terms in which he conceives 'natural justice, Binmore tries to get the best of both worlds. But the two worlds are peopled by players with different motivations. In both worlds, they are supposed to pass through the Original Position. Emerging from it, however, they would go their separate ways. Keeping them together is rather like riding two spirited horses with one bottom, a feat that is perhaps too much even for a horseman of Binmore's class.

1.6 Enforced Natural Justice?

Prima facie, enforcement should play no part in the theory of natural justice that is the whole set of 'social contracts' or self-enforcing equilibria, where no one can unilaterally improve his own payoff in the repeated game. Binmore might

have left it at that, but fortunately he did not. Fortunately, for much that he says on enforcement is most valuable, though I believe not all of it is.

He distinguishes between internal and external enforcement. In internal enforcement, all parties to a 'social contract' play the same (almost) subgame-perfect strategy. If one player deviates, one other player's strategy will tell him to punish her if only because if he did not, an nth player would punish him; the 'line bends back upon itself' and if the number of players is finite, it forms a closed circuit. The guardians guard themselves; the question 'who guards the guardians' is not answered by an infinite regress as Kant 'naively' thought that it should be (85).

What of external enforcement? Binmore clearly holds it to be relevant to his theory and in various places refers to an 'external enforcement agency' being needed to police, for example, a utilitarian distribution. He makes the excellent point that a society as a whole cannot have external enforcement, because nothing is external to it.[4] However, it can have a sub-society whose 'social contract' is enforced from within the rest of society, or *vice versa*. (141, 148). One sub-society may also have an internally enforced rule directing it to exercise external enforcement over the rest of society. It is fundamental for the social order and the principal-agent problem it generates whether such a sub-society enforces rules upon the rest of society with the object of guarding the latter's safety (like a tame police force) or in order to enslave and exploit it (like the Praetorian Guard or the Mameluks Hume cites in discussing power dependent on opinion). In any event, it makes good sense to speak of external enforcement.

Is it still the case that the guardians running and manning the external agency will guard themselves, and Kant has once more been nave when he thought that an infinite regress of layer upon layer of guardians was needed? The guardians do not have the same incentives as the players they guard, though some of their interests may overlap. Their relation is a characteristic principal-agent problem, and we know that agency problems cannot be solved, though they can be attenuated. There is no line 'bending back upon itself' as when everybody is playing the same strategy of punishing deviation. An infinite regress of guardians guarding the guardians below them would still be needed to force the guardians not to play their own maximising strategy but conform to that of the rest of society.

Though it is not clear to what end he is doing so, Binmore seems to suggest that the external enforcement agency is not so formidable as all that. The secret police is divided against itself, its members spy upon each other (196). They very likely do, but that will only keep them all the more on their toes to be model secret policemen and defend their position above society More weightily and solemnly, Binmore cuts the external enforcement agency really down to size by making the Lockean point that "the power we lend to the mighty actually

[4] There is an amusing parallel between this idea and Hayeks spontaneous order. If nothing is outside society, a state having the monopoly of power and imposing its designs is also part of society and its impact is no less reflected in the order that results than that of any other social force, Every order produced by the interaction of all the forces that make up society is a spontaneous order. One that is non-spontaneously imposed from the outside is that decreed by a foreign army of occupation. Another may be ordained by God.

remains in our collective hands" (86). Coming from someone who brands so many philosophical positions so severely as nave, such reversion to one that is more nave than most is difficult to accept. It does nothing to bolster faith in Binmore's political credo in reforming the social contract and planning a perfect commonwealth that concludes his book.

He sees himself as a Whig, but what Whig has ignored ownership and regarded wealth as lying there waiting to be redistributed according to an idea of fairness incorporated in some fancy counterfactual 'original position'? It is a strange Whig who is avowedly yearning for a 'return to something closer to the social contracts of the hunter-gathering folk for whose way of life our genes are predisposed (197). Closing the book, one is still kept waiting to learn why, if our genes are predisposed for it, there has never been and there still is no sign of such a return. Could it be that those genes and the particular kind of fairness they supposedly instil in us, matter less than this interesting but unpersuasive book would have us believe?

2. Treating Like Cases Alike

One evening the shepherd told the shepherd boy to count the flock as they come back to the fold through the stiles. The boy (he was Irish) reported that he had counted them all except one lamb that was gambolling about and would not let itself be counted.

The word "fair" behaves rather like that gambolling lamb that would not be counted. It is hopping between a multitude of meanings more or less as it pleases. The consequences for political philosophy are not propitious. To do better, one should try to stop the all too free gambolling of fairness ideas and firmly put what one can into the fold to see how it fits.

My first move will be to round up and set to one side as irrelevant to the present purpose the fairness words of ordinary language that do not really impact the social order and its justice.

In a second move, I shall show that while fairness judgments made about them might sway our moral sentiments, they are irrelevant to most of the important equilibria that regulate society. These equilibria are conventions. They are ascertainable facts without any apparent role being played in their establishment by putative fairness norms or devices, such as the Original Position. Where they operate, conventional equilibria pre-empt any effect such norms may have.

The third and final move of this section and of the paper as a whole is to explore the potential of a particular principle of fairness, of which Aristotlean equality is a special case, for providing some discipline in the use of fairness as a workable concept. The concept must lend itself to the shaping of just arrangements in that limited part of the social order that is not already pre-empted by the great body of conventional rules.

2.1 Fair Words That Must Be Set Aside

Among the great variety of sometimes widely divergent uses to which the word "fair" is put in ordinary language, two types have no or almost no relevance to what kind of coordination and distribution equilibria emerge to govern society They need to be, as it were, inventoried and set aside mainly in order to keep them firmly out of the discussion to follow.

One type of such uses is simply to express pleasure or approval without acts necessarily following the words. 'A fair woman' is either pretty or blonde or both. 'A fair man' does not allow his interest or prejudice to distract him from seeking to discover and tell the truth or give another person her due. In a 'fair trial', the accused gets his day in court. "Fair play" may mean no more than play according to the rules of the game, but it may also mean that no player enjoys an advantage or suffers from a handicap the other players do not have. A 'fair deal' is one where neither party has cause to grumble and force or fraud were not used. A 'fair prospect' has a reasonably good chance of coming off and would be welcome if it did. A 'fair statement' takes sufficient account of the relevant facts. Sometimes it means that it 'splits the difference' between opposing views. In that case, it is an intimation that 'the truth is in the middle', and if so, it falls within the other major type of usage that we seek to sweep to one side to clear the ground.

The other type comprises fairness statements that soothe, appease, promote compromises and offer platforms for milquetoast consensus. The archetype of these expressions is "fair enough", neither quite good nor quite bad but something to be content with. The 'fair settlement' of a dispute is one where both sides have given some ground. "Faq" (fair average quality) is the description of Australian export wheat which, in the egalitarian spirit of the country, is not graded. 'A fair try' is neither success nor failure. When you get 'fair shakes' you must take your chance but the odds are not against you. In 'fair criticism', the critic has a valid point, but not a devastating one; the author can live with it. In all these uses, "fair" is a term of reconciliation and not of self-vindication or protest.

2.2 Where Justice Preempts Fairness

Within a society's feasible set, there is a subset of acts that are interdicted and made liable to sanctions under an existing rule system. The rules I propose to take as rules of justice are conventions in the game theoretic sense, i.e. equilibria which, if enforcement-dependent, have built-in conditional strategies for punishing deviation. They arise spontaneously without resort to a rule for rule-making that would confer authority to rule-makers Rules of justice also serve as rules of freedom if we identify freedom, as I think we should do, with the aggregate of acts not liable to sanction under the rules of justice.

Adhering to a convention is a payoff-improving strategy. It is not the outcome of a bargain and involves no forward commitment; deviation is one of the strategy

options. Adhering to the convention is strategic in that each individual player chooses his strategy in expectation of the responses of other players, but the choice is not made so as to achieve a distribution of the payoff sum among the players in any particular way. No fairness norm is employed to select the equilibrium. Unlike a bargain to share a perfectly divisible pie that has infinitely many equilibria, a convention may have only two and the choice between them is unproblematic: 'first-come-first served' that protects first possession of unowned property as well as queuing, has no real alternative in originating property and in deciding who gets on the next No. 23 bus. Fairness, as far as one can see, is irrelevant to its selection. Justice in compliance with spontaneously emerging self-enforcing rules supersedes unenforced considerations of fairness; it does all the work in its sphere and leaves none over for fairnesss. Acts in breach of the rules of justice are unjust, and that settles the matter: they are wrongs.

In saying that justice supersedes fairness, we are not claiming that it suppresses awareness of fairness. A price that clears the market may be influenced by some sellers withholding their goods because they strongly hold some idea of a fair price. However, any number of other factors may also enter into the sellers' motivation. We are probably unable to discover all of them and assess their relative importance, including the importance of fairness among them. But there is no need and no sensible call for doing so, for all these influences are subsumed in the supply price over which the seller is free, in justice, to make any decision he pleases.

Acts admitted by the rules belong to one of two classes. They are either liberties one may, or obligations one must do or forbear from doing. A liberty is a relation between a person and his act, whether potential or actual. The prototypical liberty is the pursuit of one's peaceful purpose. It is not a game and involves no strategy. Interference with a liberty by an act that is itself an interdicted wrong is a violation of liberty, but it is not such violation that makes it a wrong. A wrong is wrong in that it is a breach of a rule of justice and not because, as is often and confusedly claimed, it violates a 'right'.

One person's liberty may be interfered with by another's pursuit of a peaceful purpose that conflicts with his own. Two men both seeking to take the same young woman out to dinner is a case in point. The two men are exercising their liberties and what they are doing is just. However, it may or may not be fair. A sense of its unfairness may influence one or both of the competitors according to the complexities of the case, but it need not do so; it is not unjust that they compete head on, and the eventual unfairness of the competition does not make it any less just.

Obligations are created by surrendering freedoms and conceding to others the rights to require the formerly free acts to be performed, respectively to be forborne. Thus, a right/obligation is a two-person one-act relation. It is fundamentally a bargaining solution that implies agreement on the terms the obligor concedes to the rightholder. This is evident in one-of-a-kind transactions where many solutions look equally plausible. The terms actually agreed upon may be agreed because the parties thought they would be fair, or for a host of other reasons. In market transactions, the bargain character of the terms is concealed,

and only one solution (one price, one set of delivery or warranty conditions) can clear the market, the solution being determined by the preferences of the marginal buyer and seller. Fairness considerations may affect the solution if they enter into the preferences of the marginal traders. Once again, however, as in the case of liberties that may be freely exercised according to the rules of justice whether or not their exercise is felt to be fair, bargaining decisions belong to the prospective obligors and rightholders. Justice assures them decisiveness over what is theirs, money or goods. The free and thus just bargain may be felt unfair by a party or by the impartial observer, but that will not stop it from taking place.

Discussing these matters in terms of freedoms, obligations and rights is pushing abstraction and generality beyond the comfort of ordinary speech. It may suffice to say, more simply, that the decisive factor in determining what happens within the rules is what belongs to the actor and what to others. We do not go far wrong if we take it that the centre of the space preempted for justice, where fairness is trumped by it or is irrelevant, is occupied by property.[5]

An advantage of surveying the space preempted for justice is that doing so fills one of the two basic maxims of justice, 'to each, his own' with defined content. This maxim (it may be better known by its Latin name *suum cuique*) is widely regarded as empty because no matter what is claimed to be 'his own', the maxim is not violated as long as the claimed thing is rendered to the claimant. Obviously, 'his own' must be determined separately and 'fed into' the maxim.

2.3 Sharing the Cake

'How to share the cake',—or whether to share it at all—is a question ubiquitous in political thought and is never far from discourse about both justice and fairness. The way it straddles both is instructive.

When the cake is owned by someone, sharing it with others or keeping it for the owner alone is a question of justice. The owner has discretion to dispose of it. He may keep it or, motivated by love or duty, divide it within his family. He may share it with strangers, and there is an endless sea serpent of psychological attributes that may motivate him to do so, the most trite being that conforming to a putative behavioural norm of sharing is an argument of his utility function.

[5] I define 'property' as a set of liberties of use and disposal of resources enjoyed by an owner. It would make for greater clarity of our concepts to distinguish between property and property rights. Being a liberty, the former relates one person to one act (more precisely, a set of related acts). Being a right, the latter relates two persons to an act (or set of acts). A property right arises when an owner surrenders his freedom to exercise some prerogative of his ownership, and undertakes an obligation to let another do so. Leasing, lending, easement and options are examples of property rights/obligations. They are typically temporary, and are granted for a consideration. Property may exist without property rights being created. Property rights, being bargains, may have a wide range of terms; the terms actually agreed upon may be subject to fairness judgments. By contrast, property, being a matter of rule, protecting (as Hume put it) 'stability of possession' and 'transference by consent' is a binary concept. Title to it is either good or not good, and judging good title as fair or unfair seems to me a misuse of "fair", even if it were relevant to the 'game of life'.

In any case, the owner does what he does, he has justice on his side and his action may or may not be judged fair. Such judgment would be an expression of a moral sentiment rather than the application of some identifiable principle.

Consider by contrast thee situation where the cake is unowned; it has fallen from heaven unaccompanied by the divine donor's instruction on how to use it. This is a situation greatly favoures by game theory and is the typical setting of distribution games. The Ultimatum Game is a good illustration. Here, two players try to share the cake. One, the proposer, offers some part of the cake to the other, the respondent. If the respondent accepts the offer, the cake is shared accordingly; if he rejects it, the cake is snatched away and neither player gets any part of it. The proposer should 'rationally' offer the smallest part that is yet large enough to persuade the respondent to take it rather than punish the proposer for his greed, reject the offer and cause the whole cake to be snatched back to heaven.

Numerous experiments, many with non-negligible game sums, have shown that most proposers offer an even split, the median offer being just under 50 per cent. This typical solution is at least consistent with the principle of treating like cases alike, for neither party has a better claim to the cake than the other. The case of one for cake is just like the case of the other and it is fair to treat them the same way. Ownership having been assumed away, justice has no say in the matter; it falls within the scope of fairness.

2.4 Where Justice Lets Fairness Prevail

What is left for us to do is readily signposted by the second maxim of justice, 'to each, according to ...' (which, with due regard to algebraic sign, can also read as 'from each, according to ...'). This maxim is to be understood to say that where some recurrent benefit or burden is not distributed by the operation of the first maxim, it shall not be distributed in a random fashion, but 'according to ...' some trait that is common to each of these recurrent situations. These distributions are neither conventions individuals adhere to, nor bargains parties agree on and though they do involve the interaction of two or more persons, their behaviour is non-strategic and does not constitute a Nash equilibrium. If the distribution is non-random, it must have some regularity or pattern. The sentencing of criminals, the awarding of honours, the promotion of employees, the praise or blame meted out to children all follow some more or less discernible and also somewhat erratic pattern, but a pattern nonetheless.

It is this pattern that may or may not be judged fair. Such judgments will be made by all who are either participants in the distribution or witnesses to it—including of course our old friend, the empathetic 'impartial spectator'. But his presence is not essential. The intrinsic looseness of ideas of fairness may well cause several contradictory fairness judgments to be made of a given distribution. Nevertheless, in any given culture one should expect no more than a small number of conflicting judgments to be made of a given distribution. Each judgment will incorporate criteria which have made the distribution fair

(because they have been applied) or would have made it fair (if they had been applied).

Each of the rival fairness judgments that postulates a different 'according to ...' criterion does so by applying the 'treat like cases alike' principle. I will argue that it is to this principle that we must look to find reasoned answers to such question as "what is fair?" and to sort out the confusion that mistakes justice for fairness or fairness for justice.

2.5 Aristotlean Equality

As the slaves are rowing the galley, the overseer flails only the slackers and does not flail the rest who pull hard enough. He is treating like cases alike and unlike ones differently, and this is fair. If some of the slaves were caught in the Balkans and others in black Africa and the overseer, himself a Serb, flogged the black ones harder, he would not be fair because he was supposed to distribute lashes among the slaves according to how they rowed and not according to whether they were white or black. However, this presupposes prior understanding that rowing and not skin colour was to be the relevant trait on which thee distribution of lashes must be based.

If the pasha's galley slaves have to row while his domestic slaves cultivate his extensive rose gardens, the galley slaves are treated unfairly, for they are all equally slaves but those pruning roses have an easier time than those who row. Finally, it is unfair that some people are free and others are slaves, for their cases are alike in all being human beings, and should in fairness be all treated alike. Yet it would be unfair to treat them alike, for they are not all alike, some being more sensitive, fragile or more attached to their native village than others. Fairness demands that the latter should be treated less harshly.

Obviously, such sophisms can be spun endlessly round and round. The point in doing any such spinning at all is to show that the 'Treat Like Cases Alike' Principle is an empty shell that can hold any content, for every case is like every other in some trait and unlike any other in some other trait. The fair content is one where the cases that are treated alike have traits that are judged relevant. The conclusion is inescapable that fairness being entirely a matter of judging which trait of a case is relevant, an objective fairness concept is a chimera. A good deal of subjectivity must be tolerated, though a frivolous degree of it need not be.

The basic form of the theory of fairness, Aristotlean Equality, operates with a single relevant trait common to each case under consideration. More complex forms of the theory may include any number of common traits as relevant to some degree. The more general and comprehensive the theory, the emptier it becomes, for every treatment of separate cases can be shown to be fair if differences in treatment can be imputed to traits that some cases possess and others do not, or some possess to a greater degree than others. By the same token, every treatment of separate cases can also be shown to be unfair. To restrain the theory from becoming a tautology, the relevance of traits must itself be restrained by the

discipline of common sense. A fairness judgment that held, on commonsense grounds, that equal work merited equal pay and more work merited more pay in proportion to how much more it was, would become unworkable if it had to make room for such refinements as the state of each worker's health, their resistance to stress, the time it takes each to come to work, the intensity of the effort each needs to get the same work done, and so on. The line separating the relevant from the irrelevant cannot be generated by the theory itself. It must be borrowed from the cultural environment, of which common sense is no doubt a part.

A perhaps naive story may be helpful in illustrating the basic form of the theory, namely Aristotlean or proportionate equality. An army is fighting a long war and soldiers periodically get leave from the front. A student doing research into fairness examines the service records of every front line soldier and finds countless common traits among them. One trait is that all get at least a certain minimum number of days' leave, but some get variable numbers of additional days as well. He tries to explain the variation by regression analysis with respect to alternative traits of these cases. He finds little correlation between age and extra leave, family status and extra leave, disciplined conduct and extra leave, but discovers a close correlation between number of enemies killed and extra leave. He concludes that the fairness norm applied by the army command uses 'enemies killed' as the sole trait relevant in distributing the benefit of marginal leave. The distribution y is a linear function of the relevant variable trait x of the form $Y = mx + c$, which is the basic form of Aristotlean equality.

An obvious line of development of this basic form is to allow more than one trait of a case to serve as relevant ground for the distribution and for each to have its separate coefficient.

It might be thought that the distributor (say, the army command) could include any number of traits as grounds for distributing the benefit and attach to each whatever coefficient he deemed suitable. He does indeed have this freedom in that he distributes at his discretion; the space justice leaves open for fairness has no rules. However, even a discretionary choice must in substance meet with a rough-and-ready measure of acceptance by those among whom the distributor distributes if he does not want to fail in his purpose. Seeking to take account of more traits of the cases for ever more refined and detailed assessment of the merits of each, a complex fairness norm runs the danger of lacking transparency and not being recognized as fair by those among whom the distributor distributes.

2.6 Non-Linear Equal Treatment

There is no logical reason implied in the 'treat like cases alike' principle of fairness, nor does it inspire an ethical argument that the benefit or burden distributed to a set of cases should vary throughout in the same proportion Convicted criminals are 'like cases'. They must be 'treated alike' in that they must all be sentenced. But the fairness principle does not require that the murderer who killed two persons should be sentenced to twice as long a prison

term than the other who killed only one. Fairness demands that he should be given a longer sentence, but 'longer' may mean more or less than proportionately longer; the function relating y to x may well be non-linear. All the principle requires is that the relation between the distribution and the ground for it should display a pattern sufficiently regular and visible for the observer to recognize.[6]

Relaxing strict proportionality between a benefit (or burden) and the ground or grounds for it, and admitting a non-linear relation between them, takes us from Aristotlean to general equality as the fairness norm. General equality admits not only discretion in the choice of the traits that give rise to a distribution, but also the extent to which each may do so.

General equality includes the special and limiting case of absolute equality. Here, all possible grounds for a distribution except one are explicitly purged from the fairness function. (Formally, they figure to the power of zero.) The exception is one common trait all cases possess to the same extent, namely that all are human beings. Whether they exert great efforts at work, kill many enemy soldiers and rear many children or just laze away the day in serene contemplation, they all get the same income because absolute equality is act-irrelevant.

The special case of absolute equality has always been a magnetic pole of attraction for political thought, though it has seldom been presented in its stark naked form; more often than not it has been embellished, camouflaged or disguised under some pseudonym. It merits reflection that in Ken Binmore's treatise on fairness as justice, a nostalgia for fairness as equality is never far from the surface.

Bibliography

Binmore, K. (2005), *Natural Justice*, Oxford-New York

Davidson, D. (1974), Psychology as Philosophy, in: D. Davidson, *Essays on Actions and Events*, Oxford 1980, 229–244

Harsanyi, J. (1977), *Rational Behaviour And Bargaining Equilibrium In Games And Social Situations*, Cambridge

Rawls, J. (1972), *A Theory of Justice*, Cambridge/MA

Vickrey, W. S. (1945), Measuring Marginal Utility By Reactions To Risk, in: *Econometrica 13*, 319–333

[6] The reader will appreciate the risk of the theory moving progressively toward emptiness as it is generalised. The more intricate the pattern that relates a distribution to the ground for it, and the greater the number of traits that are included as relevant grounds for the distribution, the more difficult it is to discern the regularity of the latter. Consequently, it is becoming progressively less possible to demonstrate that a given distribution is not fair. The tendency of the concept of fairness to disintegrate as it is generalised could be reasonably regarded as a symptom of its intrinsic pliability and lack of a hard core.